EXPANDING COMPUTATIONAL RESOURCES FOR THE DISCOVERY OF MOLECULAR BIOMARKERS OF EXPOSURE

A THESIS

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY:

KEVIN J. MURRAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ADVISORS: DR. SILVIA BALBO AND DR. JERRY COHEN

FEBRUARY 2024

Acknowledgments

My adventure to complete my doctoral dissertation is different from the journey I expected. This process has been full of trials and tribulations, unplanned detours, and stark realizations. This journey has had its ups and downs, victories, and defeats. Looking back, I see many places where I should have known better. Things I could have done better. However, if it were not for those mistakes and struggles, I would not be where I am now. When I set off on this journey to obtain my Ph.D. I had a clear vision of what this process would look like. As this degree concluded, I realized this journey was nothing like what I expected but precisely what I needed it to be. I would not change a thing.

I thank Dr. Silvia Balbo for the immense opportunity to expand and develop my analytical chemistry and bioinformatic skills. Our introduction was borne from a misunderstanding. While I am disappointed there is no consortium of scientists studying the formation of adduct ion artefacts, the opportunity to learn more about DNA adducts and systems toxicology has been much greater than anything I could have hoped. I am grateful for your constant support and mentorship. Under your guidance, I finally achieved my goal of publishing an original research article. I am inspired by the positive working environment fostered under your leadership. In the Balbo lab, I felt free to explore my ideas and was supported by everyone to see my projects completed. I thank you for this wonderful opportunity and look forward to future collaborations.

I thank my advisor, Dr. Jerry Cohen, for his constant guidance. He helped me develop my skills as a scientist and supported me when I needed it the most. I am grateful for his enduring wisdom and advice. I am appreciative of the time he took to support my research efforts and listen to my concerns. Having Dr. Cohen as my advisor and mentor throughout this process has been an honor.

I want to thank the remainder of the thesis committee: Dr. Susan Van Riper and Dr. Scott Walmsley. I would not have pursued a degree in Bioinformatics and Computational Biology if not for Sue's mentorship. I learned so much from you. I am grateful you took the time to meet with me regularly and pushed me to complete my degree. I want to thank Scott for his mentorship. There are few computational mass spectrometry experts at this university, and I am grateful you agreed to provide your guidance during this process. I hope to be able to support the research community with my computational skills like you do at the Masonic Cancer Center.

I would like to express my gratitude to all past and present members of the Balbo Lab. It is rare to see such a supportive and friendly group working toward a common goal. I would like to offer my thanks to Dr. Alessia Stornetta, Dr. Romel Dator, and Dr. Valeria Guidolin for their guidance as I learned about the world of DNA adductomics. I want to thank Foster Jacobs, Andrew Floeder, Michael Sausen, and Dr. Matthew Ludtke for the opportunity to work with you and for your assistance in navigating this domain of science. I want to give special thanks to Dylan McKeon for all his assistance and support in completing my thesis work. I was able to complete this work thanks to your assistance. I thank Chiara Lecchi for her constant support and friendship during this process. Finally, I would like to thank the Director of the Analytical Biochemistry Shared Resource, Dr. Peter Villalta. I am grateful for the opportunity to learn from you and your feedback developing my thesis project.

My ability to complete my thesis work was only possible with the support and assistance of the Center for Metabolomics and Proteomics research staff. The opportunity to work at CMSP has been one of the proudest achievements of my life. I am so grateful to work and learn from every member of this facility. I want to thank Dr. Timothy Griffin for his assistance over the years. Tim supported my education as an undergraduate and has

been a strong supporter of the decision to work full-time while I finish my degree. It has been a joy to work alongside Dr. Candace Guerrero. I have learned essential analytical skills from her and critical business acumen to help support our facility and research efforts. I thank Dr. Andrew Rajczewski for his guidance in the thesis writing process. I want to thank Todd Markowski for his constant support. Finally, I would like to offer my profound gratitude to Dr. LeeAnn Higgins. I would not be where I am today without LeeAnn's support and guidance. She has taught me almost everything I know about mass spectrometry sciences and is the first person I seek out to help me understand a new concept. I am grateful for her assistance during this process, and I look forward to continuing to work with her for more years to come.

Next, I would like to thank my family and friends. My mother, Ute Jantz, is an endless source of support and encouragement. The accomplishments of my older sister, Vanessa Murray, have served as an inspiration as I selected my career path. I know she is always there for me. I am grateful that my little sister, Jacqueline "Sassy" Murray, is always available if I need to talk and decompress. I thank my best friend, Joe Agee, for all his support during this process. I thank Andrew and him for opening their home to me and helping me navigate life's many challenges. Our movie nights were a much-needed distraction during the most challenging parts of this journey.

Finally, I thank my partners, Sean Kennedy, and Alex Urquhart. Sean and Alex were my bedrock during this journey. Their love and constant support have helped me navigate the most challenging decisions I made during this process. I know they will always be there for me when I need it and help me keep track of what is most important. Thank you, Alex, for all your assistance in editing my manuscripts. Thank you, Sean, for being the source of much-needed fun when I was stressed out. I am grateful I could share this journey with Sean and Alex, and I look forward to what adventures still await us.

Dedication

To my partners, Alex Urquhart, and Sean Kennedy. Your love, support, and wisdom have served as the foundation for me to pursue my dreams. Words cannot express my gratitude. Thank you.

Abstract

Recurrent exposure to genotoxic chemicals and other harmful environmental agents is directly related to the development of adverse phenotypes, carcinogenesis, and developmental disorders through chemical-mediated toxicity. Systems toxicology aims to chemically profile and computationally model the toxicity pathways to gain mechanistic insights into the outcomes of environmental exposure and develop modern safety guidelines. Analytical methods are being developed to perform broad, unbiased characterization of molecular biomarkers of exposure to improve the understanding of the molecular consequences and physiological responses following the introduction of reactive electrophilic chemicals. There are few computational workflows available to support this rapid development of new analytical technology. This thesis introduces novel computational resources for the parameterized analysis of discovery-driven, systems toxicology results generated by liquid chromatography-coupled mass spectrometry.

Chapter 1 of this thesis presents a comprehensive review of existing analytical technology for the molecular profiling of exposure biomarkers. This review focuses on discovering toxicologically relevant compounds using constant ion monitoring, fragmentation filtering, in-source collision-induced dissociation, mass defect filtering, or isotope pattern filtering. The mechanism of discovery of each analytical method is examined and the strengths and limitations of each approach are discussed.

Chapter 2 of this thesis presents a novel computational workflow for discovering molecular biomarkers of exposure using fragmentation filtering. In this study, an original module named DFBuilder for the metabolomics data processing software MZmine is presented. The application of this software tool for the discovery of covalently modified

DNA nucleosides is evaluated. In this work, a novel colibactin-derived, E. coli associated DNA adduct product discovery is highlighted.

Chapter 3 of this thesis extends the application of the DFBuilder module to discover urinary metabolites produced from detoxifying reactive electrophilic chemicals in tobacco cigarette smoke. This work presents the first reported application of a high-resolution mass spectrometry method for profiling mercapturic acid conjugates in positive ion mode. The combination of this novel analytical method and computational workflow discovers numerous prospective mercapturic acid signatures never reported in human urine. Statistical evaluation of these results demonstrates that many of these products are associated with cigarette usage.

Chapter 4 of this thesis reports a novel mass spectral library of conjugated mercapturic acids. Using multiple fragmentation strategies, this library represents thousands of mass spectra collected in positive and negative ion polarity. This work serves as a foundation of resources necessary for verifying discovery results produced from the analytical methods presented in this thesis. Metadata insights from this library that will help future efforts to characterize mercapturic acid conjugates are discussed.

This thesis concludes with a summary and future perspective evaluating the remaining computational challenges in exposomics and analytical chemistry-based approaches for systems toxicology. The areas that most need support are highlighted, and the capacity of emerging computational solutions to improve experimental outcomes is discussed.

Table of Contents

List of Tables

List of Figures

Abbreviations

1HMPeMA	N-Acetyl-S-[1-(hydroxymethyl)-2-propen-1-yl]-L-cysteine
1PMA	N-Acetyl-S-(n-propyl)-L-cysteine
23HPMA	N-Acetyl-S-(2,3-dihydroxypropyl)-L-cysteine
24NPhMA	N-Acetyl-S-(2,4-dinitrophenyl)-L-cysteine
2CaEMA	N-Acetyl-S-(2-carbamoylethyl)-L-cysteine
2CoEMA	N- Acetyl-S-(2-carboxyethyl)-L-cysteine
2CoMEMA	N-Acetyl-S-(2-carboxy-1-methylethyl)-L-cysteine
2CoPMA	N-Acetyl-S-(2-carboxypropyl)-L-cysteine
2CyEMA	N-Acetyl-S-(2-cyanoethyl)-L-cysteine
2HBeMA	N-Acetyl-S-(2-hydroxy-3-buten-1-yl)-L-cysteine
2HEMA	N-Acetyl-S-(2-hydroxyethyl)-L-cysteine
2HPMA	N-Acetyl-S-(2-hydroxyethyl)-L-cysteine
2PeMA	N-Acetyl-S-(2-propen-1-yl)-L-cysteine
34HBMA	N-Acetyl-S-(3,4-dihydroxybutyl)-L-cysteine
3HMPMA	N-Acetyl-S-(3-hydroxy-1-methyl- propyl)-L-cysteine
3HPMA	N-Acetyl-S-(3-hydroxyethyl)- L-cysteine
4NPhMA	N-Acetyl-S-(4-nitrophenyl)-L-cysteine
5AcAHPhMA	N-Acetyl-S-[5-(acetylamino)-2-hydroxyphenyl]-L-cysteine
8-oxo-dG	8-oxo-7,8-dihydro-2´-deoxyguanosine

AGC	automatic gain control
BzMA	N-Acetyl-S-(benzyl)-L-cysteine
CEMA	cyanoethyl mercapturic acid
CGT	cysteinylglycine transpeptidase
CID	collision-induced dissociation
CNL	constant neutral loss
ctDNA	calf thymus DNA
D5-ethyl-dC	D5-ethyl-2´-deoxycytidine
DART	direct analysis in real time
DDA	data dependent acquisition
DIA	data independent acquisition
EIC	extracted ion chromatogram
EMA	N-Acetyl-S-(ethyl)-L-cysteine
GGT	γ-glutamyltransferase
GSH	glutathione
GST	glutathione s-transferases
HCD	higher-energy collision dissociation
HNE-dG	6-(1-hydroxyhexanyl)-8-hydroxy-1, N2-propano-2´-deoxyguansine
HPLC	High performance liquid chromatography

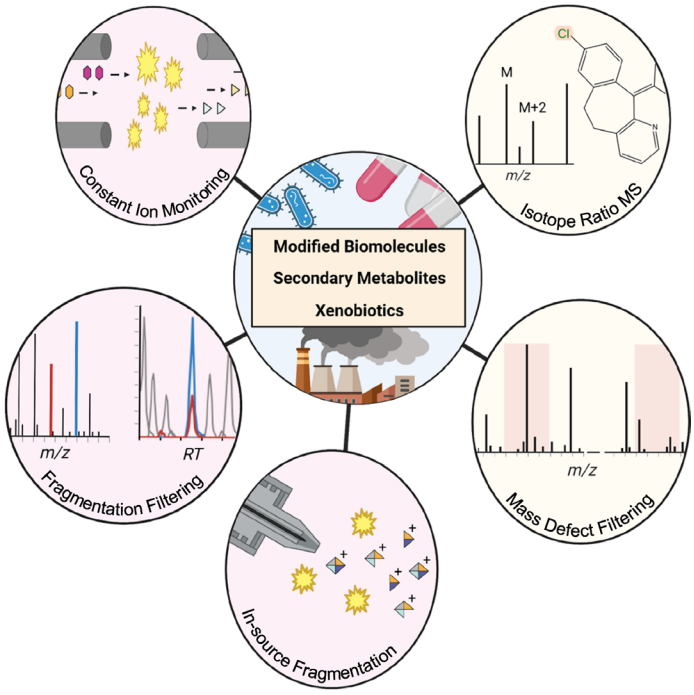
HRMS	high-resolution mass spectrometry
IPF	isotope pattern filtering
iRT	internal retention time
ISCID	in-source collision induced dissociation
LC-MS	liquid chromatography-coupled mass spectrometry
LC-MS/MS	liquid chromatography-coupled tandem mass spectrometry
M	monoisotopic molecular ions
m/z	mass-to-charge
MDF	mass defect filtering
MMA	N-Acetyl-S-(methyl)-L-cysteine
MRM	multiple reaction monitoring
MS	mass spectrometry
MS/MS	tandem mass spectrometry
N2-ethyl-dG	N2-ethyl-2´-deoxyguanosine
N6-Me-dA	N6-hydroxymethyldeoxyadenosine
NAT	n-acetyltransferase
NMR	nuclear magnetic resonance
NNAL	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone

O2-POB-dT	O2-[4-(3-pyridyl)-4-oxobut-1-yl]thymidine
O6-me-dG	O6-methyl-2´-deoxyguanosine
OH-PdG	(6R/S)-3-(2´-deoxyribos-1´-yl)- 5,6,7,8-tetrahydro-6-hydroxypyrimido[1,2-a]-purine-10(3H)one
OPLS-DA	orthogonal partial least squares discriminant analysis
PFAS	per- and polyfluoroalkyl substances
PhMA	N-Acetyl-S-(phenyl)-L-cysteine
PI	precursor ion
pks-	pBeloBAC bacterial artificial chromosome
pks+	BACpks island bacterial artificial chromosome
Q1	first quadrupole
Q2	collision cell
Q3	third quadrupole
RT	retention time
VOC	volatile organic compound
ε-dA	1, N6-etheno-2´-deoxyadenosine

Chapter 1: Systems Toxicology and Small Molecule Analytics

The following chapter provides background and context for the analytical approaches utilized in systems toxicology to discover molecular biomarkers of chemical exposure and its effects. The chapter was adapted with permission from the following review article published in *Chemical Research in Toxicology*. The original article was authored by Kevin Murray and edited with the assistance of all co-authors.

Murray, K. J.; Villalta, P. W.; Griffin, T. J.; Balbo, S. Discovery of Modified Metabolites, Secondary Metabolites, and Xenobiotics by Structure-Oriented LC–MS/MS. *Chem. Res. Toxicol.* **2023**. https://doi.org/10.1021/acs.chemrestox.3c00209.



1.1 Overview of Systems Toxicology

Toxicity assessment has been at the core of modern toxicology sciences for decades. Animal experimentation was the primary method to evaluate a substance's toxicological hazards and risks to human health. While this form of testing helps develop guidelines to protect humans and animals from acute threats, the role of chemical toxins in the pathogenesis of complex diseases and developmental disorders is gathering attention.[1] Exposure to harmful chemical agents can impact numerous biological processes directly, in addition to inducing secondary physiological responses, such as inflammation or oxidative stress, that confound characterization efforts. An intricate combination of genetic, lifestyle, and environmental factors influence the etiology of many clinical phenotypes and directly produce or modify the severity of clinical outcomes.[2–4] With advancements in analytical technology for toxicity testing and risk assessment, a prevailing paradigm shift emerged in the early 21st century that promotes a mechanistic, pathway-oriented approach to evaluate toxicity in the context of multiple biological systems.[5]

Systems toxicology is an extension of systems biology that aims to combine biological and toxicity data with advanced computational modeling techniques to obtain mechanistic insights into toxicological outcomes and understand the relative risks of harmful chemical exposure events.[1,6] Multiple streams of biological data are required to understand the pathway of toxicity, the sequence of molecular events leading to an adverse effect, and predict the toxicological outcomes of undescribed toxic materials. Physiological responses may occur at the cellular, tissue, organ, or organism level, reflecting molecular alterations in the genomic, transcriptomic, proteomic, or metabolomic systems.[1] Systems toxicology aims to utilize this information-rich, -omics data to elucidate

the pathway of toxicity of chemical agents, construct predictive models to improve chemical risk assessment, and evaluate potential therapeutic targets to enhance the success of clinical intervention and prevention strategies. The power of this approach has been demonstrated for the prediction and validation of selective drug targets that inhibit cell growth and proliferation for treating several forms of cancer.[7,8]

Application of systems toxicology for the comprehensive exposure assessment struggles with many challenges, notably in collecting large omics datasets representative of multiple biological systems. The advent of genomics-scale data collection and processing revolutionized the biological sciences at the start of the 21st century. In the following two decades, parallel efforts to develop and expand “big data” methods in analogous biological systems to enable the complete biomonitoring of an organism were a top priority. The development of high throughput methodology in these contemporary fields of transcriptomics, proteomics, and metabolomics offers unique opportunities to characterize chemical exposure events. As dynamic systems, perturbations in transcript, protein, or metabolite concentrations may be readily associated with variations in clinical phenotypes and interrogated to infer the potential pathway of toxicity.

The application of RNAseq and contemporary technologies offers quantitative profiling of tens of thousands of RNA transcripts that can be functionally annotated and associated with biological processes. For example, RNAseq-based toxicogenomics discovered that e-cigarette usage induced a transcriptional response associated with lung cancer, inflammation, and fibrosis genes but often with reduced expression relative to control cigarette smoke models.[9–11] Technological advances in mass spectrometry-based methodology facilitated the simultaneous detection and quantitation of thousands of proteins in tissue or cell extracts. Proteomics assessment provides mechanistic insights

into the physiological impact of exposure. Quantitative protein profiling enabled the development of toxicology models to describe the reproductive toxicity associated with mycotoxin exposure of *Fusarium* fungi contamination in cereal crops.[12–14] Similar to advances in proteomics technology, modern metabolic profiling via mass spectrometry-based platforms provides insights into the chemical characteristics of thousands of metabolites and small molecules in biological matrices. Toxicometabolomic evaluation of industrial contamination by per- and polyfluoroalkyl substances (PFAS) in municipal water supplies suggests these “forever chemicals” impact numerous chemical pathways, including the pentose phosphate shunt pathway, endocrine signaling, and lipid metabolism.[15,16] The application of small molecule profiling analytical technologies contends with numerous obstacles preventing the widespread collection of information-rich chemical data from biological matrices. Despite the limitations of current metabolomic technologies, applications of small molecule profiling in toxicology are proliferating, with increasing demand for improved analytical workflows.

The metabolic response to a chemical challenge is often considered the most closely associated response with phenotypic outcomes. Metabolic fluctuations occur rapidly and may modulate biochemical pathways and physiological responses in a manner not related to transcriptional or translational responses from the host organism. Small molecule profiling directly characterizes chemical exposure through xenobiotic screening, detoxification profiling, and genotoxicity testing. In contrast to well-defined primary metabolites, research into the non-homeostatic aspects of metabolism, physiology, and toxicity is in its infancy.[17] Few reference materials are available for purchase, and chemical database entries are sparse as most of these analytes represent novel and undescribed compounds. These modified biomolecules, secondary metabolites, and xenobiotics are

crucial for describing the unique characteristics of an organism and understanding its interaction with the environment.[17]

1.2 Molecular Biomarkers of Exposure and its Effects

Many factors influence the direct biomonitoring of xenobiotics and harmful chemical agents. The persistence of chemical agents in the body depends on the target substances' unique molecular properties, dosage strength, and route of exposure. Fat-soluble or protein-binding toxins may accumulate in cells and tissues to be slowly metabolized and excreted days or months after an initial exposure event. However, many harmful chemical agents encountered daily represent volatile and reactive compounds. These toxic agents exhibit a short half-life and rapidly metabolize or transform to impose detrimental effects on the surrounding cells and tissues. The brief persistence of these toxicants impedes their direct biomonitoring. Molecular biomarkers of exposure represent metabolic by-products or chemical outcomes generated by interaction with reactive chemical agents that are more readily measured in biofluids or tissue biopsies. In contrast to their reactive progenitor compounds, the biomarkers of exposure often persist longer in the body and can be readily detected days or weeks after the initial exposure. Measuring these molecular targets can provide insights into the type and strength of chemical exposure events and offer valuable information on the bodily response to harmful toxins.

A multitude of metabolic by-products and reaction adducts are produced by exposure to toxic chemical materials. Exposure to 4-(methylnitrosamino)-1-(3-pyridyl)- 1-butanone (NNK), a tobacco-specific carcinogen, is known to produce numerous unique products following metabolic transformation, such as 4-(Methylnitrosamino)-1- (3-pyridyl)-1-butanol (NNAL), in addition to untold numbers of covalently modified DNA, RNA, and protein species (**Figure 1.1**).[18–20] Many environmental and household exposure events

involve coexposure to numerous chemical agents simultaneously, increasing the complexity and distribution of metabolic by-products and reaction adducts throughout the body.[21] The combined effects of coexposure are not strictly additive. Increasing reports of synergistic and antagonistic effects of coexposure have been reported.[22] Developing toxicological models for potential chemical interactions represents a substantial challenge in modern systems toxicology.

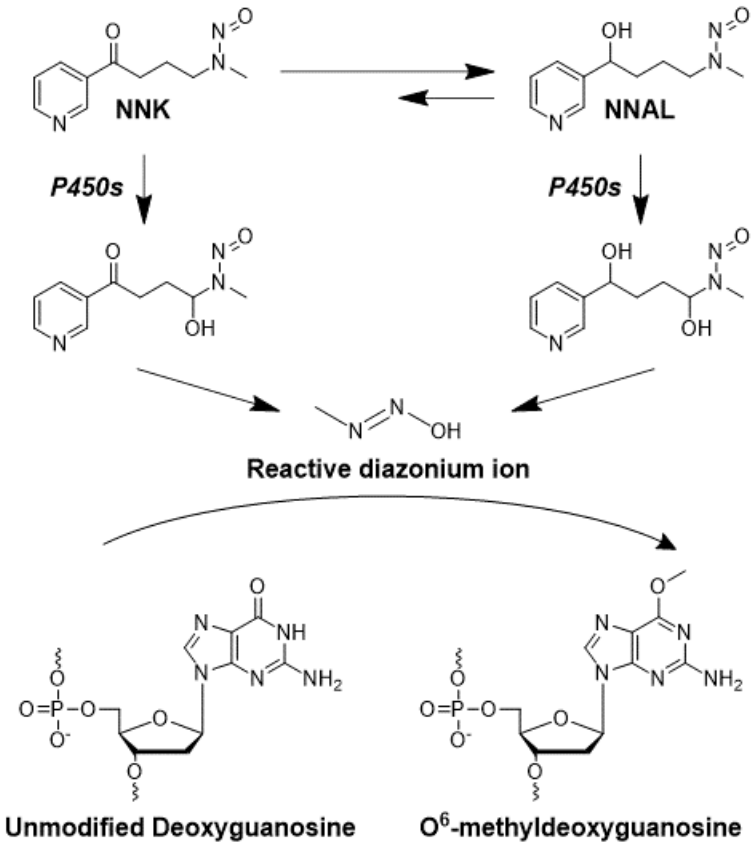


Figure 1.1 - Formation of DNA adduct O6-methyldeoxyguanosine by metabolic transformation of tobacco-specific carcinogen NNK.

The simultaneous characterization and quantification of all molecular outcomes induced by chemical exposure is beyond the capacity of current metabolomics

technologies. To overcome this obstacle, the development of systematic analytical workflows focuses characterization efforts on specific classes of metabolic by-products and reaction adducts. Adductomics technologies aim to characterize all chemical adduction products generated by the reaction of chemical toxins with DNA,[23–25] RNA,[26,27] or protein[28,29] macromolecules, respectively. The systematic study of metabolic transformation and detoxification products has classified multiple sub-metabolomes, the complete collection of all metabolites of a particular class, such as the mercapturome[30–32] and sulfatome[33,34], the collection of mercapturic acid and sulfate conjugates, respectively. These emerging methods represent new frontiers in systems biology, providing valuable insights into the biological impacts and responses to exposure to harmful chemical agents. The capacity of reactive chemical agents to damage DNA is of primary concern in systems toxicology. In contrast to epigenetic modifications that regulate gene expression, covalent DNA modifications and the resulting DNA adducts can disrupt normal cellular function. If not restored by endogenous DNA repair mechanisms, DNA damage may result in genetic mutations that can translate into the development of cancer and other complex disorders. How xenobiotics and reactive chemical agents mutate DNA is an active research area. DNA adductomics is an emerging analytical approach that aims to identify the complete collection of DNA adducts in cells or tissue. Characterization of DNA adducts can provide valuable insights into the pathways of genotoxicity of a chemical agent. DNA adduct concentration and signatures of mutation may be used to infer the effective exposure dose and are increasingly used as biomarkers in cancer research and chemoprevention studies.

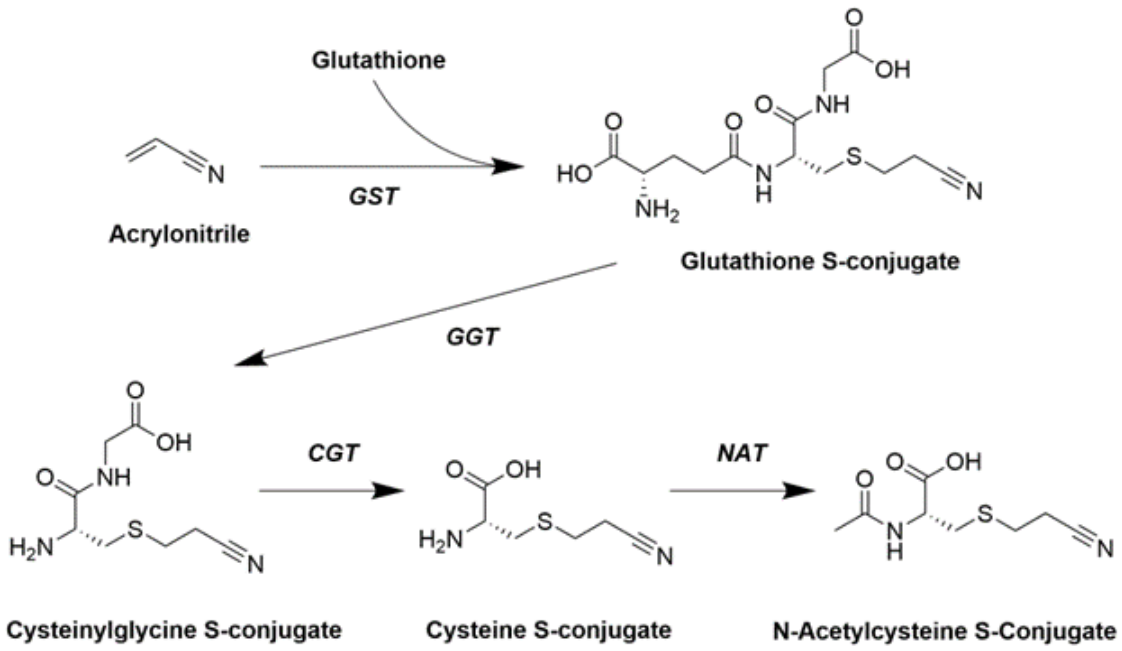


Figure 1.2 - Formation of cyanoethyl mercapturic acid (CEMA) from acrylonitrile transformation. Acrylonitrile is conjugated to glutathione by a glutathione s-transferases (GST) enzyme. Hydrolytic conversion by γ-glutamyltransferase (GGT) removes the glutamic acid functional group, and further modification removes the glycine by a cysteinyl glycine transpeptidase (CGT). Acetylation by n-acetyltransferase (NAT) forms the final mercapturic acid conjugate to be excreted in urine.

Neutralizing harmful chemical agents through metabolic transformation and detoxification pathways represents an organism's primary mechanism for eliminating toxic substances. Metabolic transformations are broadly categorized into Phase I and Phase II reactions. Many classes of xenobiotics, particularly lipophilic compounds, require biotransformations to expose or introduce polar functional groups. Phase I metabolism involves enzymatically-catalyzed reduction, oxidation, or hydrolysis of chemical toxins, adding or exposing primary amine (-NH_2), hydroxyl (-OH), or thiol (-SH) functional groups to increase polarity and water solubility. Phase II metabolism conjugates xenobiotics or Phase I metabolic products to polar compounds to further increase polarity for urinary secretion. Glucuronidation, sulfation, and acetylation reactions conjugate distinct polar

groups onto the molecule that help neutralize reactivity and further improve solubility in water. Phase II conjugation with glutathione, a tripeptide of glycine, cysteine, and glutamic acid, represents a common detoxification mechanism in biological organisms. Subsequent transformation of glutathione-conjugates leads to the formation of mercapturic acid conjugate (**Figure 1.2**), a readily observed urinary metabolite. Multiple methods have been developed in recent years for the unbiased detection of metabolic transformation products. Characterizing these intermediates and secreted products provides valuable information on the metabolic response to harmful chemical exposure.

1.3 Analytical Technology for Characterizing Molecular Outcomes of Exposure

Small molecule discovery analytics faces many obstacles arising from chemical heterogeneity, frequent occurrences of structural isomers, high dynamic range, and trace-level concentrations. Liquid chromatography-coupled mass spectrometry (LC-MS) represents a flexible approach for small molecule analytics capable of comprehensively characterizing complex chemical matrices and accurately quantitating molecular concentrations.[35–37] Mass spectrometry (MS) is a powerful tool that detects ionized molecules' mass-to-charge ratio (m/z) to obtain chemical composition and infer structural arrangements. LC-MS enables specific chemical analyses capable of temporal resolution of independent molecular signatures in complex matrices and offers potential chromatographic separation of some isomeric and isobaric compounds. Molecular mass alone is often not enough to provide confident identification of unknown molecules in the absence of standard reference material. Using tandem mass spectrometry (MS/MS), characteristic fragmentation patterns may be collected and used for identification purposes via spectral library matching or manual annotation. In contrast to contemporary analytical platforms, such as nuclear magnetic resonance (NMR), LC-MS represents a

highly sensitive analytical technology capable of characterizing molecular concentrations over multiple orders of magnitude, down to trace-level detection at nanomolar to picomolar concentrations.[38]

Applications of LC-MS for the characterization of modified biomolecules, secondary metabolites, and natural products span multiple decades.[24,37,39–44] In the health sciences, LC-MS approaches have been applied to describe chemical exposure and its impact at the molecular level. Conjugated glutathione (GSH) adducts, as a measure of antioxidant activity, can serve as valuable biomarkers for exposure and provide valuable insight into the toxicity of various human activities and occupational hazards.[45–47] Monitoring a specific subset of molecules in a complex biological matrix is challenging without *a priori* knowledge of the target class of interest. The common glutamylcysteinylglycine core structure of all GSH adducts enables an intuitive mechanism for LC-MS/MS-based screening approaches. Researchers observed that fragmentation of various GSH adducts produced repeating product ion patterns that could be monitored independently of the intact precursor mass.[40,45,48] This discovery became a driving force in the characterization of drug detoxification products and other xenobiotic clearance mechanisms in human and animal models[46,49] expanding our ability to understand the molecular mechanisms of chemical exposures and their interaction with physiological processes and pathways.

Structure-oriented monitoring methods are employed for the detection of multiple classes of toxicologically or pharmacologically relevant compounds. In contrast to unbiased shotgun methodologies, structure-oriented monitoring utilizes prior knowledge of the elemental composition, structural moiety, or functional group characteristics to develop an analytical pipeline honed for the detection of a subset of compounds.[50,51] Fragmentation-based LC-MS/MS methods are routinely used to characterize phase I and

phase II metabolites such as cysteinyl,[30,31,42,52,53] sulfate,[43,54,55] or glucuronide conjugated products. [44,56,57] Additionally, discovery methods have been developed to monitor for specific classes of clinically relevant types of molecules by diagnostic fragmentation patterns, including: steroid hormones,[58–60] covalently modified DNA and RNA nucleosides,[24,61–63] and small molecules or peptides exhibiting glycosylated, phosphorylated, or other distinct functional groups modifications.[64–66] Early-stage drug discovery requires extensive characterization of drug metabolism, including descriptions of drug-related metabolites, potential reactive intermediaries, and degradation products.[52,57,67] To characterize this array of metabolically diverse drug intermediates, structure-oriented monitoring approaches were developed to screen predictable transformations frequently occurring *in vivo*, such as hydroxylation, dehydrogenation, and demethylation, among other typical phase I or II conjugation reactions.[68,69] These approaches have been used to discover numerous drug bi-products[70–74] and nerve agents and their degradation products.[75,76]

1.4 Innovations in LC-MS/MS Technology for Structure-Oriented Discovery

The demand for class-specific monitoring approaches drove the development of multiple innovations in LC-MS/MS technology. The following sections review the concept of structure-oriented LC-MS approaches and the analytical paradigms possible for monitoring compounds through elemental, structural moiety, or functional group similarity. The five most common approaches are discussed, including (1) constant ion monitoring, (2) fragmentation filtering, (3) in-source fragmentation, (4) mass defect filtering, and (5) isotope pattern filtering. These approaches represent a combination of instrument-dependent monitoring techniques and post-hoc computational filtering processes. Only some of these methods are universally applicable to all LC-MS/MS instrument platforms;

however, the simultaneous application of one or more of these techniques maximizes the coverage of complex matrices. Finally, a perspective on the shared challenges of the field and prospective areas of development is discussed.

1.4.1 Structural-Oriented Screening Concept

The causal relationship between molecular structure and MS/MS fragmentation patterns forms a foundational principle of mass spectrometry.[39,72] Upon disassociation in the collision zone of the instrument, ionized compounds fragment in complex but reproducible patterns dependent on the activation technique, such as collision-induced dissociation (CID) and higher-energy collision dissociation (HCD). Typically, molecular dissociation occurs at labile chemical bonds, cleaving the precursor species along its proverbial weak points.[77,78] Strong covalent bonds may be dissociated with higher energetics to fully fragment the entire precursor structure.[78] The observed fragmentation patterns are ascribable to the originating compounds and descriptive of their molecular moiety. In conventional metabolomics workflow, these acquired MS/MS spectra are queried against a reference spectral library to infer chemical identity.[79] With limited chemical and spectral libraries cataloguing the expanse of modified biomolecules, secondary metabolites, and xenobiotics, alternative methods must be used to discriminate target molecules of interest from unrelated metabolites.

Shared patterns in fragmentation are often exhibited among species with related structural moiety and functional groups (**Table 1.1**). A class of compounds typically produces common neutral loss and fragment ions based on similarities in their core structure,[52,63,80,81] as illustrated by the following examples. Glycerophospholipids may be reliably distinguished by the neutral loss of a polar head group, such as the loss of a phosphocholine (183 Da)[82] or phosphoethanolamine (141 Da)[83] structural moiety.

Modified DNA nucleosides have been historically monitored by the neutral loss of deoxyribose (116 Da) structural moiety, regardless of the nucleic acid base.[23,61] Glucuronide-conjugated metabolites excreted in urine are commonly detected using multiple fragmentation pattern-based approaches, including the neutral loss of glucuronide moiety (176 Da), dehydrogenated equivalent (194 Da), and the observation of various diagnostic fragment ions in either positive (177 m/z, 159 m/z, 141 m/z) or negative (113 m/z, 85 m/z, 75 m/z) ion modes.[44,56,84] The combination of multiple fragmentation patterns is often used to expand the breadth of coverage of a discovery assay or improve the specificity to limit false positive hits.

It is important to acknowledge that not all structural moieties are amenable to LC-MS monitoring. The neutral loss of small functional groups such as hydroxyl and amino groups is ubiquitous among many chemical species. More extensive moieties that encompass the core of the structural backbone may be challenging to fragment reproducibly without specialized energetics to prevent further dissociation into smaller fragment ions in the collision cell. Finally, when monitoring for the detection of a fragment ion of a characteristic moiety, the ion in question must be ionizable. For electrospray ionization, the molecule must possess suitable electrophilic or nucleophilic sites to promote efficient ionization. The use of chemical derivatization can improve the stability and ionization efficiency of selected compounds. Using selective reagents, chemical derivatization may serve an additional purpose and enable monitoring of annealed functional groups or protecting groups.[85,86] Often, no shared fragmentation patterns exist to describe a group of chemically-related compounds prohibiting application of fragmentation monitoring approaches. Alternative monitoring methods are routinely used to facilitate molecular screening using unique characteristics of the target class of compounds. The breakdown and degradation products of numerous classes of drugs and metabolites are well-

described and readily monitorable without explicit fragmentation events. Similarly, isotope pattern tracing enables screening for compounds with unique elemental compositions. Many essential trace elements exhibit unique natural isotope patterns distinguishing their detection from background metabolites.[70]

1.4.1 Constant Neutral Loss and Precursor Ion Scan Monitoring

In quantitative small molecule analytics, multiple reaction monitoring (MRM) using triple quadrupole mass spectrometers represents a powerful technique for sensitive metabolic profiling (**Figure 1.3A**). In the first quadrupole (Q1), a select precursor m/z is isolated and transmitted to the collision cell (Q2) for fragmentation. The third quadrupole (Q3) is set to monitor a defined fragment ion m/z to be detected by the final ion detection system. While many analytical applications of triple quadrupole-based platforms utilize only targeted MRM applications, alternative acquisition modes enable unbiased ion detection over a wide mass range. Configuring the quadrupole to scanning mode enables the transmission of the multiple ion masses through Q1 or Q3 providing a survey of multiple compounds in a complex mixture. Application of these scanning modes in specific configurations enable the discovery of molecules with shared molecular substructures via shared neutral loss or product ion formation.

Table 1.1 - Common Diagnostic Fragmentation Patterns of Toxicologically Relevant Compounds

Metabolite Class	Neutral Loss	Moiety Loss	Accurate Mass of Neutral Loss	Product Ion (+/-)	Accurate Mass of Product Ion
Glutathione Adducts[45,87]	$C_5H_7NO_3$	Pyroglutamic acid	129.0426	$C_{10}H_{13}N_3O_6^-$ $C_{10}H_{11}N_3O_5^-$	272.0888 254.0782
Mercapturic Acid Conjugates[31,42]	$C_5H_7NO_3$	N-Acetyl-L-Cysteine	129.0426	$C_5H_8NO_3S^-$ $C_5H_6NO_3^-$ $C_4H_6NO^-$ $C_2H_4NO_2$-	162.0225 128.0348 84.0450 74.0242
Glucuronide Conjugates[44,56]	$C_6H_8O_6$ $C_6H_{10}O_7$	Glucuronide Dehydrated Glucuronide	176.0321 194.0427	$C_5H_6O_3^-$ $C_4H_6O_2^-$ $C_2H_3O_3^-$	113.0244 85.02948 75.0082
Sulfate Conjugates[54]	SO_3	Sulfate	79.9568	HSO_4^- HSO_3^- SO_3^-	96.9595 80.9646 79.9568
Covalently Modified Nucleosides [61,62,88]	$C_5H_5N_5O$ $C_5H_5N_5$ $C_5H_8O_4$ $C_5H_6N_2O_2$ $C_5H_8O_3$ $C_4H_5N_3O$	Guanine Adenine Ribose Thymine Deoxyribose Cytosine	151.0494 135.0545 132.0423 126.0429 116.0473 111.0433	$C_5H_5N_5O^+$ $C_5H_5N_5^+$ $C_5H_6N_2O_2^+$ $C_4H_5N_3O^+$	152.0567 136.0618 127.0502 112.0506

Constant neutral loss (CNL) monitoring is a popular approach for the characterization of unknown compounds that utilizes a precursor-product neutral loss offset to scan an *m/z* range for molecules of interest (**Figure 1.3B**).[26,30,52,62] In this scan mode, Q1 scans over a range of *m/z* values, typically set at 1 Da intervals. Following fragmentation, Q3 also scans over a mass range with fixed intervals defined by neutral loss offset between the precursor Q1 isolation. Using the CNL scan mode, a diagnostic neutral loss may be monitored over a chromatographic gradient in which the detected peaks represent ions that potentially exhibit a structural moiety or functional group of interest. Precursor ion (PI) monitoring scans a defined precursor *m/z* range but transmits a fixed product ion mass to the detector (**Figure 1.3C**).[54,56,86] Similar to CNL monitoring, Q1 scans over a *m/z* range in 1 Da intervals; however, in PI monitoring, Q3 is set to isolate and detect a fixed fragment ion mass. Using the PI scan mode, chromatographic peaks represent compounds that possess a distinctive structural fragment descriptive of the origin molecule class. Using a triple quadrupole platform, the simultaneous application of a CNL or PI experiment for small molecule discovery with product ion scanning for structural annotation is challenging. Hybrid triple quadrupole-linear ion trap mass spectrometers overcome this limitation by the unique properties of the Q3 linear ion trap mass analyzer. At the cost of sensitivity, detected peaks from a CNL or PI experiment may be isolated for a full fragment scan to infer chemical identity.[42]

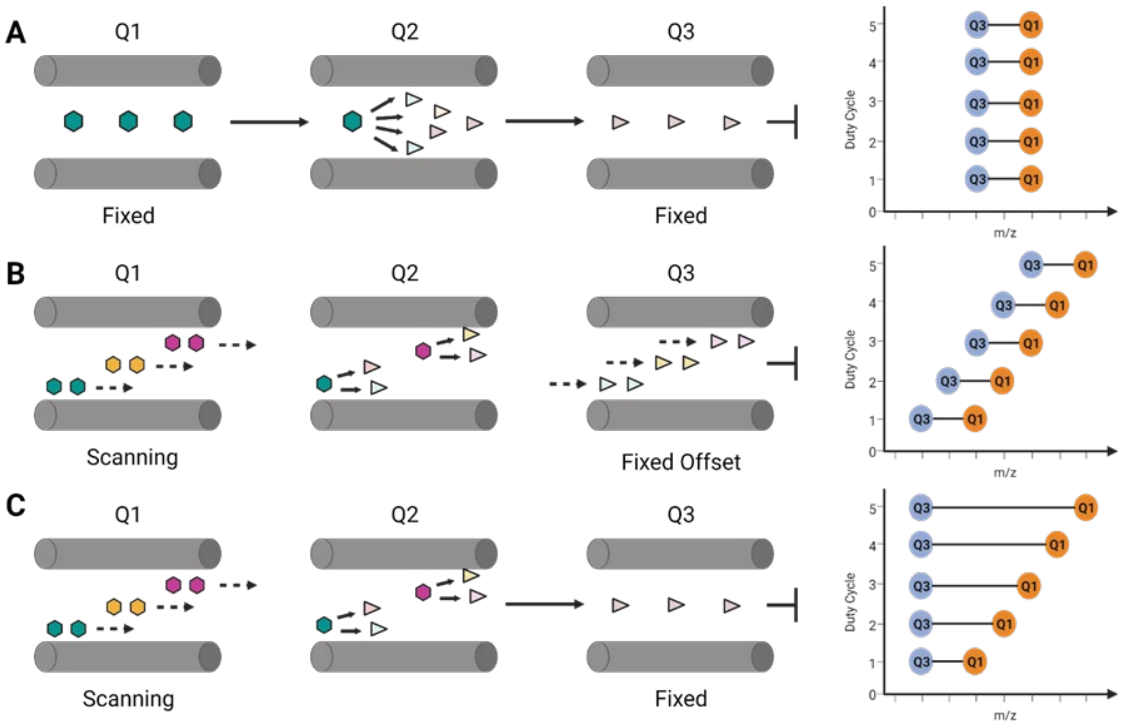


Figure 1.3 - Triple quadrupole scan types for structure-oriented monitoring. (A) Multiple reaction monitoring on defined target. Q1 and Q3 isolate a fixed mass for selected ion monitoring. Each duty cycle repeats this isolation schema. (B) Constant neutral loss scan. Q1 scans over a mass range and Q3 scans over a mass range using neutral loss-derived offset from the precursor isolation mass. Each duty cycle the isolation mass of Q1 and Q3 are incremented by 1 Da. (C) Precursor ion scan. Q1 scans over a mass range and Q3 isolates a fixed ion mass. Each duty cycle the isolation mass of Q1 is incremented 1 Da while Q3 remains fixed.

There are a multitude of applications of CNL and PI monitoring in the fields of toxicology, pharmacology, and drug discovery. Xenobiotics, endogenous chemicals, and drug intermediaries are reported to react with DNA, proteins, and host metabolic processes in living organisms.[89–91] The reactive tendency of these toxins impedes the ability to monitor their concentration in biological systems directly but the analysis of the adducted and trapped bi-products are commonly employed to monitor for exposures.[45,92] Analysis of GSH-trapped reactive metabolites via CNL monitoring of the pyroglutamic acid moiety (129 Da)[40,93] and PI monitoring of the γ-glutamyl-dehydroalanyl-glycine fragment

in negative ion mode (272 *m/z*)[94,95] has been employed to measure reactive metabolites in urine derived from xenobiotics,[40] cigarette smoke,[96,97] drug-induced toxicity.[45,48,93–95,98] Further transformation of GSH via transpeptidase cleavage and subsequent acetylation produce mercapturic acids-conjugates that are commonly assessed via CNL monitoring to evaluate reactive metabolite exposure.[30,42,52,99] Biomonitoring of phase II metabolites produced through glucuronidation[56,84,99,100] and sulfation[54,55] processes are routinely monitored through CNL and PI approaches. Chemical detoxification is finite with non-trivial levels of various reactive metabolites interacting with host systems by means of covalent modification. Covalently modified RNA and DNA are routinely monitored to assess the long-term health burdens of chemical toxicity on biological systems. Monitoring of the diagnostic neutral loss of the ribose and deoxyribose moieties in hydrolyzed RNA and DNA, respectively, provide a powerful approach for the unbiased characterization of the epigenetic and carcinogenic impacts of harmful chemical exposure.[26,62]

The development and application of CNL and PI monitoring techniques greatly expanded the understanding of the multitude of conjugation and bi-products produced by chemical exposure events and their association with clinical phenotypes. While a powerful analytical technique, there are several important limitations of CNL and PI-based discovery LC-MS/MS approaches. The low-resolution mass detection of quadrupole mass analyzers restricts precise elemental composition calculations of putative hits. Subsequently, the low mass specificity increases the false discovery potential of CNL and PI approaches in comparison to high resolution mass spectrometry-based techniques.[47] The use of multiple diagnostic neutral loss or fragment ions reduces the false discovery potential of an assay but the inclusion of multiple monitors limits the effective duty cycle of these experiments.[101] Despite these limitations, CNL and PI monitoring approaches continue to see regular applications in many research settings.

1.4.2 Fragmentation Filtering with Data Dependent and Data Independent Acquisition

Improvements in the scan rate and sensitivity of high-resolution mass spectrometry (HRMS) have expanded its potential for discovery metabolomics and small molecule analytics. Using the data dependent acquisition (DDA) scan mode, independent precursor ions are selected for isolation and fragmentation in an intensity-dependent manner over the course of an LC-MS/MS experiment. In a classical shotgun metabolomics experiment, these observed product ion spectra are queried against a reference library to infer chemical identity.[79] Library entries for modified biomolecules, secondary metabolites, and natural products are limited and require extensive manual curation to obtain structural information. Post-acquisition fragmentation filtering can be used to monitor relevant precursor-product ion relationships that serve as diagnostic signals for the detection of the analyte type of interest. High mass accuracy and full fragment ion scans offer improved specificity for structure-oriented monitoring and comprehensive chemical annotation.[47]

Fragmentation filtering in DDA-LC-MS/MS experiments serves as an effective curation tool to monitor diagnostic fragmentation patterns amongst the thousands of detected ions.[47,63,72,102,103] Acquired MS/MS spectra are systematically monitored for defined relationships between the isolated precursor ion *m/z* and observed fragmentation patterns using stringent mass tolerance filters (**Figure 1.4A**). Several software tools are available for applying fragmentation filtering, including vendor-specific solutions, such as Compound Discoverer (Thermo Scientific) and UNIFI (Waters) or open-source alternatives like MZmine[104] and Neutral Loss MSFinder.[81] However, many research applications of fragmentation filtering utilize custom in-house software[69,105] or strictly dependent on manual curation.[88] Applications of fragmentation filtering have proven

beneficial in deconvoluting the chemical complexity of many matrices throughout toxicology, pharmacology, and drug discovery. Similar to constant ion monitoring approaches, exact mass fragmentation filtering approaches are used to monitor for glutathione conjugates to monitor phase I metabolites associated with drug metabolism and hazardous chemical exposures.[47,67,101] Contemporary fragmentation filtering approaches for monitoring covalently modified DNA and RNA leverage high mass accuracy and multi-stage activation MS^n spectra for more specific compound annotation.[20,24,63] Fragmentation filtering has proven a powerful tool in the discovery of many classes of natural products with potential health, environmental, and economic benefits. There is increasing interest in characterizing the natural products of traditional medicines and multiple methods have been developed to survey the broad spectrum of expanse of therapeutically relevant compounds, including: chlorogenic acids,[106] quinochalcone C-glycosides,[107] and malonylginsenosides.[37]

The scan rate of modern instrumentation is insufficient to simultaneously isolate and fragment the hundreds of molecular ions that co-elute during an LC-MS/MS analysis of a complex matrix. Development of data independent acquisition (DIA) approaches aim to overcome this limitation by incrementally performing wide-window precursor isolation scans over an experimental duty cycle, fragmenting all molecular ions in the mass range.[108] Current computational pipelines aim to extract the numerous intact precursor ion peaks from the LC-MS full scan spectra and match them to any co-eluting diagnostic fragment ion peaks produced from the corresponding wide-window isolation MS^2 spectra to detect target analytes of interest (**Figure 1.4B**).[109,110] Application of structure-oriented monitoring techniques to DIA acquisition approaches are beginning to emerge with promising outcomes for trace-level discovery. Independent experiments have demonstrated its applicability for the discovery of covalently modified DNA,[111] mercapturic acid

conjugates,[53] and glucuronide conjugates.[44] A thorough comparison of DDA- and DIA-LC-MS/MS approaches has not been evaluated for clinically-relevant molecular classes. Internal evaluations conducted in the authors' laboratories suggest that these techniques are complementary, with DDA- and DIA-based acquisition approaches providing comparable results, but each offers a unique subset of putative targets of interest.

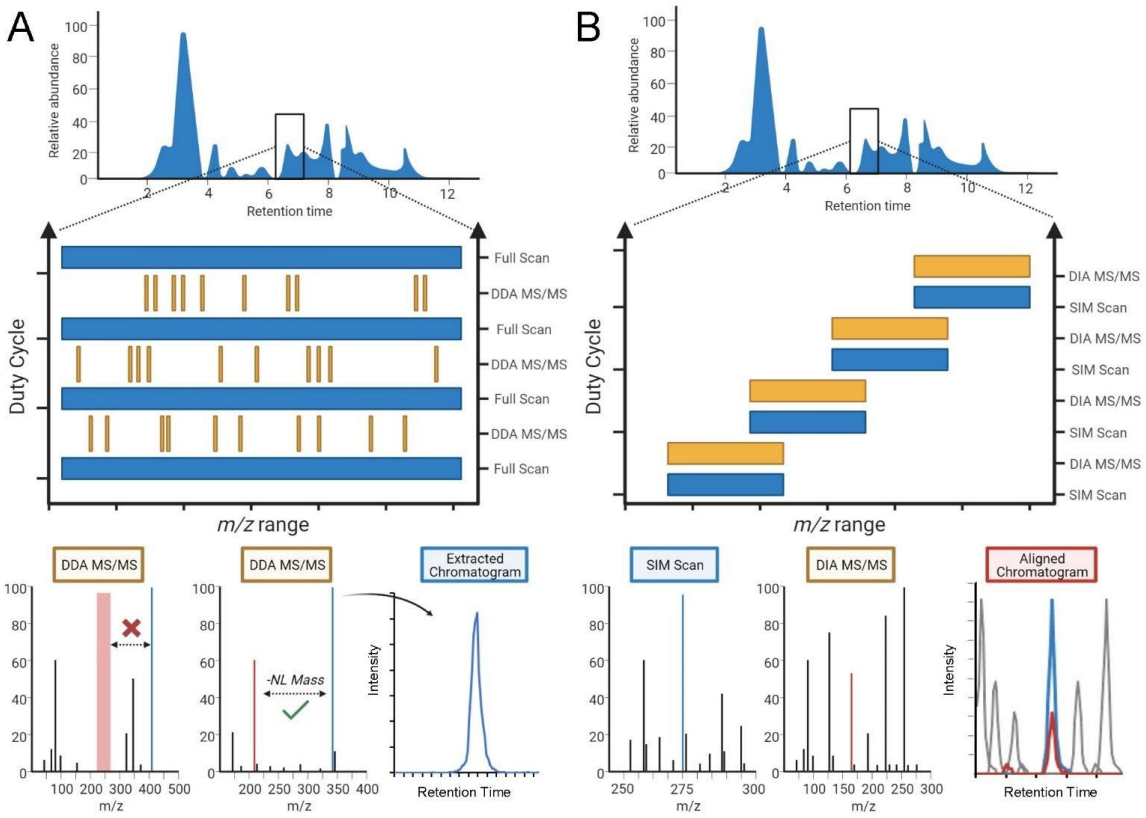


Figure 1.4 - LC-MS/MS acquisition paradigms for fragmentation filtering. (A) LC-DDA-MS/MS acquisition design. Results are collected using an intensity-dependent selection of precursors over the full precursor mass range. Acquired MS/MS are monitored for fragmentation patterns and precursor ions of interested are extracted for downstream processing. (B) LC-DIA-MS/MS acquisition design. Results are collected over a selected precursor mass range with wide-window isolation MS/MS acquisition. Precursor-product ion pairs of interest are detected and monitored for chromatographic alignment.

Although fragmentation filtering approaches offer strong specificity and flexibility for the discovery of novel and undescribed biomolecules in association with chemical exposure and drug metabolism, they also come with limitations. Not all chemical species exhibit diagnostic fragmentation patterns. In some cases, critical fragment ions formation rate is below the instrument's limit of detection. In high complexity matrices exhibiting large dynamic range application of fragmentation filtering can be challenging. For DDA-based approaches, the application of fragmentation filtering is dependent on the collection of tandem mass spectrum on precursors of interest. Many potential hits may fail to trigger a fragmentation event in a high complexity matrix excluding the possibility for fragmentation monitoring.[112] Contemporary DIA-based methods claim to overcome this obstacle with wide-window isolation strategies but face analogous challenges in high complexity matrices. Simultaneous fragmentation of large packets of ions in high dynamic range samples may distort fragment ion abundance complicating chemical annotation of unknown molecules.[112] In DIA-based fragmentation filtering approaches, wide window fragmentation can also increase the risk of false detection by erroneous misassignment of unrelated fragment ions to a precursor of interest.[44]

1.4.3 In-Source Collision-Induced Dissociation

In-source fragmentation is an often-unavoidable aspect of electrospray ionization that occurs as ions travel from atmospheric pressure at the ion source to the near vacuum of the mass analyzer. Under ideal conditions, electrospray ionization predominantly produces intact protonated or deprotonated molecular ions; however, a wide range of ion source fragments may form during transmission, largely dependent on the configuration and values of the ion optic voltages.[113,114] The loss of labile functional groups such as hydroxyl- (18 Da) and amine-groups (17 Da) are commonly observed derivatives from

electrospray ionization of small molecule metabolites but substantially larger structural rearrangements are reported.[115] Researchers observed that the compilation of these in-source fragments resembled CID fragment spectra patterns produced using low activation energies (**Figure 1.5A**).[113,116] This front-end fragmentation strategy, commonly referred to as in-source collision-induced dissociation (ISCID), provides valuable structural information without the requirement of conditional precursor isolation using in contemporary LC-MS/MS acquisition paradigms. Labile functional groups and other readily observed fragmentation patterns can be monitored directly in the survey scan of a discovery experiment to infer chemical identity. Optimizing ion optics energetics enables structure-oriented monitoring to discover compound classes that reproducibly produce diagnostic in-source fragment ions.[45,64,75] In contrast to constant ion monitoring or fragmentation filtering approaches, ISCID approaches to structure-oriented monitoring are not dependent on collision cell-based fragmentation of precursor molecules. As a front-end ionization-based discovery strategy, ISCID is an accessible monitoring approach compatible with multiple instrument configurations.

As a non-selective fragmentation strategy, ISCID approaches produce partial fragmentation of multiple molecular ions simultaneously following ionization. While the observed fragmentation is often less complex than ion trap-type and beam-type CID activation,[113,116] numerous fragment ions may be generated from a single molecular ion. Following source fragmentation, the resulting ion patterns must be deconvoluted to distinguish fragment signals produced from independent molecular ions co-eluting during an experiment (**Figure 1.5B**). In structure-oriented monitoring experiments, discovery of molecular signatures of interest requires the simultaneous detection of two or more peaks corresponding to an intact molecular ion and a diagnostic neutral loss or fragment ion species.[81] In some vendor software, mass transitions between a molecular ion and in-

source fragment ion may be specified to conditionally fragment molecular signatures of interest. The MS/MS spectra produced from in-source fragment ions are colloquially referred to as "*pseudo*-MS^3" because they provide information on the substructure of an ion without using multistage activation.[117] This isolation strategy is particularly beneficial for MS platforms without ion trapping capabilities, such as hybrid quadrupole-time of flight platforms, that are otherwise incapable of performing MS^n fragmentation strategies. While many research applications of ISCID-LC-MS/MS for structural moiety monitoring focus on the detection of neutral loss transitions, diagnostic product ions may be monitored in a similar manner. Isolation of these fragments enhances the specificity of these assays, providing an extra level of discrimination in complex biological matrices.[45]

Initial applications of ISCID primarily focused on the characterization of macromolecule modifications and forensically-relevant small molecule substances using direct analysis in real time (DART) methodology.[118] With increasing adoption of HRMS technologies, LC-MS-based metabolomic and small molecule applications are increasing, notably in fields of clinical and forensic toxicology, environmental sciences, and natural products research. Using a constant ISCID ionization approach, researchers developed a selection ion monitoring (SIM) approach to trace the exact mass of deprotonated γ-glutamyl-dehydroalanyl-glycine fragments (272.0888 m/z) of glutathione conjugates rivaling the sensitivity of PI monitoring on triple quadrupole-based instruments.[45] Analogous methods have been developed for the detection of glucuronide-conjugates.[119] Development of ISCID methods for the detection of drugs-of-abuse and other illicit substances associated with sporting regulations, including polysaccharide-based plasma volume expanders dextran and hydroxyethyl starch,[120,121] rapid screening of drugs and toxicants in forensic toxicology settings,[122,123] and many other broad coverage screening assays.[124] Currently, applications of ISCID methods in toxicology, pharmacology, and drug

discovery are sparse. In plant biology, considerable development efforts are being invested for the characterization of therapeutically and commercially relevant natural products, such as, glycosylated secondary metabolites,[66] ginsenoside-related natural products,[37,125] and chlorogenic acid derivatives.[126] A thorough evaluation of ISCID-based methodology to discover novel and undescribed toxicologically or pharmacological relevant compounds is needed.

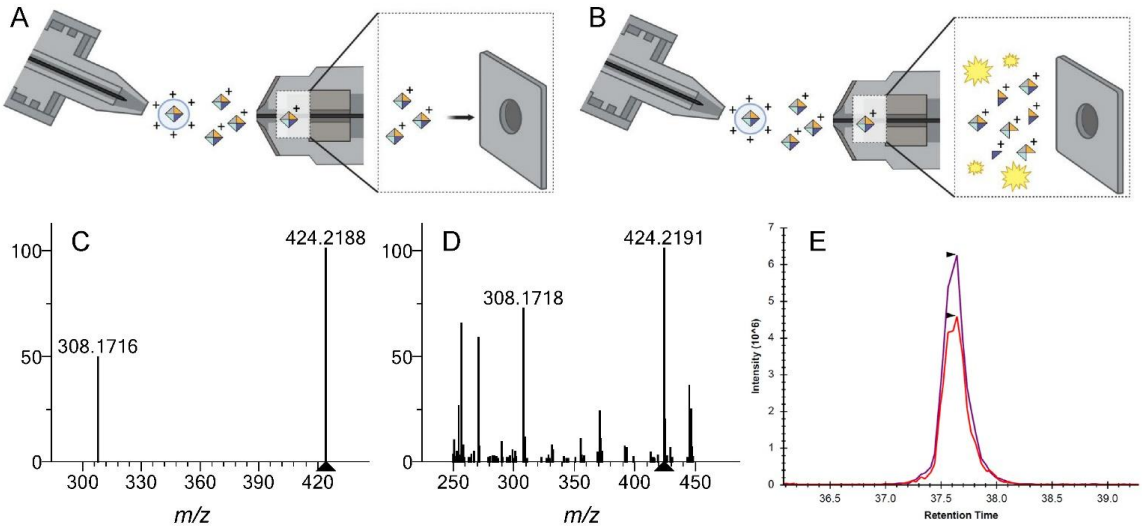


Figure 1.5 - Acquisition design for ISCID on LC-ESI-MS platform. (A) Front-end ionization and ion optics without ISCID. (B) Front-end ionization and ion optics optimized for in-source fragmentation. (C) Reference spectra for modified DNA adduct HNE-dG with HCD activation at 0% collision energy, acquired from Mass Bank of North America (https://mona.fiehnlab.ucdavis.edu/). (D) Mass spectra of HNE-dG ion in hydrolyzed calf thymus DNA background. (E) Chromatographic alignment of intact HNE-dG precursor (m/z 424.2188, purple) and in-source fragment (m/z 308.1716, red).

Structure-oriented monitoring using an ISCID approach offers a comprehensive and flexible approach to discover classes of compounds using diagnostic in-source fragment patterns. Not all chemical species are compatible with this front-end fragmentation approach and require higher CID or HCD collision energies to produce

informative patterns. Suitable energetics parameters have been described for several classes of compounds[127] but optimizing these settings for previously unstudied classes represents a considerable undertaking. ISCID approaches risk false negative detection if comprehensive source voltages and ion optics parameters are not available for dissociation of the target class of molecule upon ionization. Additionally, application of source fragmentation approaches is challenging in high dynamic range matrices. Dissociation of ions at the source reduces the effective intensity in the acquired spectrum. ISCID approaches are less suitable for trace-level signatures observable near the detection limits of the instrument.[121] These issues are compounded in high complexity matrices in which many ions are simultaneously dissociating at the source, potentially suppressing the signal of low intensity ions.

1.4.4 Mass Defect Filtering

The shared limitation of all fragmentation monitoring approaches is the need for reproducible and predictable disassociation patterns to discover unknown compounds of interest. Not all members of a chemical class will exhibit such properties and unpredictable fragmentation will occur despite a shared common substructure. Monitoring multiple fragmentation patterns can improve the breadth of coverage for a class of compounds but will decrease the specificity of such approaches. In the fields of pharmacology and drug discovery, these constraints limited application of fragmentation monitoring for the discovery of novel drug degradation and adduction products which exhibit greater fragmentation heterogeneity. With the expansion of HRMS into small molecule analytics, mass defect filtering (MDF) emerged as a contemporary method to discover structurally similar compounds not dependent on shared fragmentation patterns.[68,73] This technique employs computational filtering based on a narrow mass range offset from nominal

masses observed for compounds of similar molecular formulas to selectively monitor for common and predictable structural modifications from a target compound **(Figure 1.6A)**. MDF acts as a powerful denoising algorithm to enable sensitive detection of possible modifications of a target compound from unrelated background metabolites. As a computational method, MDF is not dependent on predefined instrument configurations to perform structure-oriented monitoring enabling flexible post-hoc applications and multiple MDF applications to expand coverage.[69,71,128]

MDF utilizes the concept of fractional mass filtering to remove non-informative, background and interfering ions from a discovery LC-MS experiment.[129] Modern HRMS platforms offer high mass resolution scan modes with many instruments capable of elemental composition analysis. With increases in mass accuracy, the fractional mass of an ion (the non-integral part of the *m/z*) became as important as the nominal mass for discriminating small molecules. Addition or subtraction of a functional group from a compound will not only modify its nominal mass but its fractional mass as well. The alteration in fractional mass, the mass defect, is an insightful marker when monitoring a target ion. Researchers observed that the mass defect of many common chemical modifications coalesced within 50 mDa of the origin compound.[68] Hydroxylation, methylation, and many other common chemical transformations fall within this fractional mass filtering range (**Table 2**). Utilizing this observation, researchers proposed using a mass defect window to filter ions over a discrete *m/z* range.[73,129] Only ions within the mass defect window pass through the filter to be utilized in downstream data processing and peak detection workflows (**Figure 1.6B-E**). Using a MDF approach, large volumes of data may be quickly curated to putative targets of interest theoretically associated with the origin compound of interest. Altered forms, modifications, and degradation products may be monitored simultaneously without the need for selected ion dissociation. Mass defect-

triggered DDA acquisition is implemented in some vendor acquisition software to enable the collection of additional structural information in unbiased experiments.

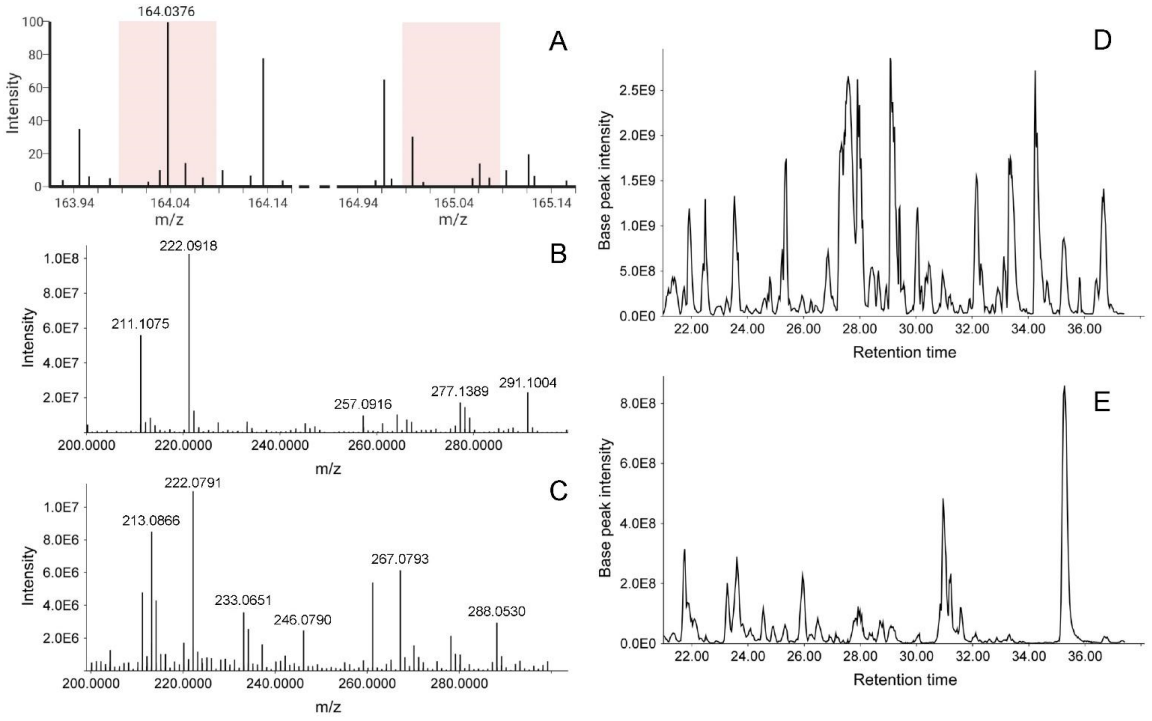


Figure 1.6 - Examples of mass defect filtering in LC-MS results. (A) MDF of precursor ion mass of unconjugated mercapturic acid in positive ion mode using a 50 mDa tolerance. Detected ions not observed in the highlighted window are discarded from the analysis. (B) Unaltered mass spectra of conjugated mercapturic acid, 3-HPMA at 222.0795 m/z. (C) Mass defect filtered spectra of 3-HPMA. (D) Base peak chromatogram of global profiling of urine matrix. (E) Transformed base peak chromatogram of urine matrix using MDF on unconjugated mercapturic acid exact mass.

Data mining applications of MDF are routinely practiced in many fields of the biological sciences, notably in drug discovery and natural products chemistry. The first application of MDF demonstrated the discovery of a novel drug-related species following reduction and methylation in a complex matrix[73] and was gradually extended to many other drug discovery experiments.[71,72,128,130–132] Utilizing MDF as a complementary

discovery technique has proven an effective tool when the structural modification distorts canonical fragmentation pathways. GSH-trapped reactive metabolites are traditionally screened using fragmentation-based screening approaches, such as CNL; however, distinct adduction species fail to produce diagnostic patterns. Partnering MDF with fragmentation-based approaches has demonstrated the ability to detect unique GSH adduct species missed using only fragmentation monitoring approaches.[67,133,134] In the fields of exposomics, application of MDF has proven valuable for monitoring exposure of xenobiotics and other foreign chemicals through glucuronidation, cysteinylation, sulfation, or other phase I and phase II metabolites transformations.[135–137] Application of MDF in natural products chemistry has proven impactful for the characterization of therapeutically relevant compounds in traditional medicines and flavor compounds in many food and beverages.[69,138]

Applications of MDF span decades with discovery experiments providing valuable insights into drug metabolism, toxin clearance, and chemical composition of complex tissues and biofluids. With an expansive body of application reference, the limitations of MDF are well described[68] and additional curation steps are required to remove unrelated chemical species from discovery results. The application of a 50 mDa fractional mass filter facilitates the curation of a discovery experiment toward molecules associated with a target of interest but many false positives will pass through this filter.[68] Subsequent elemental composition analysis, MS/MS fragmentation-based curation, or isotope pattern analysis are required to disambiguate novel discoveries from the background metabolome. Additionally, using a stringent 50 mDa filter will restrict the discovery of many larger conjugation and adduction products expected in the metabolism of xenobiotics, drugs, or chemical toxins.[129] The application of multiple defect filters will expand the

breadth of coverage of an MDF approach but will require increasing curation to mitigate the consequential increase in false positive detections.

Table 1.2 - Common Phase I and Phase II Metabolic Transformations and Mass Defect Shifts

Biotransformation Type	Molecular Formula Change	Exact Mass Change (Da)	Mass Defect Shift (mDa)
Decarboxylation	[M-CO_2]	-43.9898	10.2
De-ethylation	[M-C_2H_4]	-28.0313	-31.3
De-methylation	[M-CH_2]	-14.0157	-15.7
Dehydrogenation	[M-H_2]	-2.0156	-15.6
Amination	[M+NH-O]	-0.9840	16
Deamination	[M+O-NH]	0.9840	-16
Reduction	[M+H_2]	2.0156	15.6
Methylation	[M+CH_2]	14.0157	15.7
Oxidation	[M+O]	15.9949	-5.1
Dehydration	[M-H_2O]	18.0106	10.6
Acetylation	[M+C_2H_2O]	42.0101	10.1
Glycine Conjugation	[M+C_2H_3NO]	57.0215	21.5
Sulfation	[M+SO_3]	79.9568	-43.2
Taurine Conjugation	[M+$C_2H_5NO_2S$]	107.0041	4.1
Glutamine Conjugation	[M+$C_5H_8N_2O_2$]	128.0586	58.6
Glucuronidation	[M+$C_6H_8O_6$]	176.0321	32.1
Glutathione Conjugation	[M+$C_{10}H_{15}N_3O_6S$]	305.0682	68.2

1.4.5 Isotope Ratio Mass Spectrometry

To perform the most thorough investigation into the chemical metabolism of xenobiotics the use of isotopically-labeled reference compounds is required. With the defined incorporation of stable-labeled isotopes (^2H, ^{13}C, ^{15}N) the detection of novel and undescribed adduction or conjugation products may be intuitively monitored via isotope ratio mass spectrometry. Isotope pattern filtering (IPF) biases a discovery experiment to conjugated compounds or bi-products exhibiting distinct isotope patterns between the

detected monoisotopic molecular ions (M) and their stable-label incorporated forms (M+1, M+2). Application of IPF extends controlled experiments using stable isotope-labeled reference materials. Halogens and metal ions naturally possess multiple stable isotopic forms. Chlorine, bromine, iron, gallium, and other rare elements found in biomolecules exhibit multiple stable isotopic forms that occur in a distinct abundance. Isotope-driven discovery methods are routinely utilized to detect incorporation of rare elements chelated or conjugated to protein, DNA, RNA, or other metabolic intermediates. Isotope ratio mass spectrometry is a widespread approach used in multiple disciplines including food sciences,[139,140] marine biology,[141] toxicology,[142] pharmacology,[70] and natural products chemistry.[143]

Utilizing the natural abundance of stable isotopes as a detection criterion forms the basis of IPF monitoring approaches. Stable-isotope ratios are highly reproducible properties of an element and represent a fundamental measurement of mass spectrometry. Common proteomic and metabolomic software monitor ion abundance when computing the isotopic envelope of a molecular ion to properly infer charge state and discriminate interfering signals. Analysis of biomolecules incorporated with stable-labeled isotopes, halogens, or other trace elements requires modification of isotope detection heuristics to accommodate the impact of variable isotope occurrences. Restricting an analysis to ions exhibiting a defined ratio between select isotope pairs provides a quick mechanism for filtering shotgun experiments to a subset of putative targets. In a discovery context, the chemical formulas of molecules of interest are unknown and the observed isotope ratios often deviate from the expected values. Specifying suitable ratio tolerances enables isotope-driven discovery of a broader range of chemicals. Utilizing HRMS, high resolution mass differentials in isotope transitions, in addition to isotopic ratio filtering, may be used to increase specificity of an analysis.[70]

Isotope ratio mass spectrometry and IPF have been used to directly monitor stable isotope-labeled and natural stable isotopes for over 50 years. Utilizing isotopically-labeled GSH, multiple classes of drugs and reactive xenobiotic bi-products have been classified using an IPF strategy to monitor for explicit isotope ratios.[144–148] Administration of stable isotope-labeled medications enables a more holistic approach for monitoring of drug metabolism, allowing for monitoring of multiple phase I and phase II conjugation products.[70] Many oral medications and chemotherapeutics utilize halogenated or metal-based functional groups to catalyze specific reactions. Brominated compounds exhibit two stable isotope forms (^{79}Br and ^{81}Br) in near equal proportions that may be directly monitored for facile detection of bromine-tagged products.[149,150] Chelated metal complexes serve highly specialized tasks in biological organisms, including catalyzing enzymatic reactions, interaction intermediaries, and molecular scavengers. Detection of metallo-metabolites, metal complexes, and other metal-chelating metallophores may be detected using a stable isotope-driven discovery approach.[149,151]

Isotope ratio mass spectrometry and IPF are routinely used throughout the analytical sciences to verify molecular signatures, guide the discovery of novel compounds, and validate putative ion assignments. Unavoidable barriers inhibit the application of isotope ratio-driven discovery approaches and prohibit its usage in many experiments. Acquisition and administration of stable-labeled drugs and xenobiotics often represent a design obstacle, requiring appropriate clearance and safeguards to ensure safe experimentation. Observational studies of occupational or lifestyle chemical exposures are often not amenable to isotope-driven discovery. While the analysis of natural stable isotopes remains a viable experimental design paradigm in many

exposomics experiments, the prevalence of such elements restricts IPF techniques to select experimental contexts.

1.4.6 Combination of Multiple LC-MS/MS Characterization Methods

Characterizing the overwhelming degree of complexity induced by chemical exposure is beyond the capacity of any singular analytical technique. Multiple structure-oriented monitoring approaches have been developed to expand the discovery possibilities in an LC-MS/MS experiment. These approaches represent instrument-driven or computational techniques to guide the discovery of compounds exhibiting shared structural moiety, functional groups, or chemical composition. Each of these approaches possess unique strengths and limitations that restrict their detection capabilities to molecules exhibiting explicit characteristics (**Table 1.3**). To expand the coverage of discovery experiments, applying multiple structure-oriented monitoring approaches in tandem is becoming increasingly common. Using HRMS, fragmentation filtering, ISCID, mass defect filtering, and IPF can simultaneously be accomplished using a DDA or DIA discovery strategy. Constant neutral loss and precursor ion scanning approaches can also be combined with ISCID approaches for specific structural monitoring in pseudo-MS^3 screening approaches.

As complementary analytical approaches, applying multiple structure-oriented monitoring methods can overcome the limitations of a select methodology.[125] For example, each technique reviewed in this article has been used to characterize GSH adducts unbiasedly.[40,45,47,94,133,144] Fragmentation filtering approaches monitoring the pyroglutamic moiety or γ-glutamyl-dehydroalanyl-glycine fragment ion are potent tools, but these disassociation patterns are not universal for GSH adducts. Simultaneous application of mass defect filtering can expand a discovery assay to detect novel products that may not

exhibit canonical fragmentation pathways.[67] Combination of MDF with DIA-based fragmentation filtering can help reduce the burden of data processing an interpretation. Curating the total feature list to only precursors within the filtering window narrows the discovery space and lessens the time required to validate diagnostic precursor-fragment co-elution patterns. The simultaneous application of in-source fragmentation with fragmentation filtering-based validation has proven a valuable tool to minimize false discoveries by requiring putative targets to pass multiple discovery thresholds.[37,45] The putative discovery target must exhibit a diagnostic ion pair following ISCID acquisition and exhibit the same fragmentation pattern following targeted mass difference acquisition of the prospective precursor compound.

The combination of multiple structure-oriented monitoring approaches is unexplored in many fields of toxicology, pharmacology, and drug discovery. While these acquisition strategies and computational workflows represent powerful discovery tools, the limited resources available for automated data processing are a barrier to widespread application. Existing software and algorithms often exhibit poor compatibility to accommodate and utilize multiple sources of structural information. Standardization of acquisition and data processing workflows will aid the development of new software development for flexible data analysis. As with all small molecule analytics, the interpretation of discovery results produced from structure-oriented approaches proves a bottleneck in the analytical pipeline. Improved integration with emerging chemical annotation and structural prediction software is paramount for the continued development of these fields.

Table 1.3 - Summary of Structure-Oriented LC-MS/MS Monitoring Approaches

Discovery Method	Strengths	Weakness
Constant Neutral Loss / Precursor Ion Monitoring	• Sensitive fragmentation monitoring approach • Intuitive data interpretation • Accessible analytical technology	• Low mass resolution (limited specificity) • Limited duty cycle to monitor multiple patterns • Acquisition of complete product ion scan requires additional acquisition
Fragmentation Filtering	• Highly specific compound discovery by high resolution mass filtering and ion isolation. • Post-hoc application and optimization of fragment detection parameters. • Direct collection of structural fragments improves downstream structural annotation and identification	• Reduced performance in high complexity matrices with wide dynamic ranges • Limited duty cycle to acquire fragmentation spectra on all co-eluting ions (DDA only) • Complex data processing with considerable manual interpretation of results required (DIA only)
In-Source Collision-Induced Dissociation	• Enables fragmentation-based discovery without precursor isolation event • Fragmentation of in-source fragments provides pseudo-MS3 structural insights • Well-suited to labile moiety and functional groups	• Optimization of ion optic energetics required • In-source fragmentation reduces absolute intensity and reduces ability to detect trace-level ions • Increases complexity of LC-MS results, increasing chance of false discovery detection
Mass Defect Filtering	• Discovery of related compounds not dependent on shared fragmentation pathways • De-noising of raw LC-MS/MS results that improves manual interpretation and downstream processing • Selective fragmentation of precursors of interest improves information content of experiment	• Extensive curation of results required • Discovery of compounds of interest outside mass defect window is impossible • Add-on instrument software is required to selectively fragment precursor ions in filter window
Isotope Pattern Filtering	• Specific compound discovery not dependent on shared fragmentation pathways • Comprehensive monitoring of exogenous chemicals and bi-products in complex mixtures • Accessible data analysis and interpretation	• Only compatible with natural molecules incorporated with select bioelements or metals • Application of stable-labeled isotopes is costly and not amenable to all experiment types. • Putative compounds of interest may exhibit isotope pattern ratios outside parameterized tolerances

1.5 Limitations of Current LC-MS/MS Analytical Workflows

Limited data processing tools represent a critical barrier to the accessibility of focused LC-MS workflows. Structure-oriented acquisition and data processing are curation techniques for enhancing the detection and discrimination of ions of interest from a complex background matrix. Commercial and open-source software solutions are emerging to alleviate the burden of manual data processing, but considerable development is still required. Integrating these tools with emerging peak detection algorithms, downstream spectral library searching, and chemical annotation tools is paramount for expanding to any unexplored class of compounds. Ensuring the compatibility with multiple structure-oriented acquisition strategies and computational curation techniques will broaden the potential scope of these experiments.

As with all small molecule analytics, chemical identification is a critical bottleneck in the focused LC-MS workflow. In contrast to metabolomic analysis of primary metabolites, the outcomes of many structure-oriented experiments are not expected to populate commercial or open-source spectral libraries. As such, inferring the identity of novel drug degradation products, undiscovered metabolite modifications, or rare natural products is almost entirely dependent on manual spectral annotation. Additionally, each approach discussed in this review is susceptible to false discovery driven by electrical noise, background contaminant signals, or unrelated isomeric or isobaric structural moiety or functional groups. Improvements to predictive spectra annotation tools[152,153] or spectral scoring tools that quickly filter poor quality results will greatly improve the outcomes of structure-driven experimentation.[154]

Development of community-driven data-sharing guidelines represent a paradigm shift in the mass spectrometry community. Important data mining technological advances

enable spectral annotation of uncommon metabolites and natural products through spectral similarity searching and spectral networking approaches. Only select research journals and periodicals require disseminating raw results as part of the FAIR (findable, accessible, interpretable, and reusable) data management guiding principles.[155] Few metabolomic studies utilize publicly available repositories to document results, and archival of structure-oriented discovery experiments is even rarer. Improving the accessibility and reusability of spectral results from structure-oriented LC-MS experiments offer a promising mechanism for elucidating structural information of rarely studied metabolic signatures.

1.6 Conclusion and Thesis Goals

Development of a universal metabolic profiling technique is beyond the scope of current analytical technology. The tremendous chemical diversity and high dynamic abundance range of all molecular species in a complex biological specimen prevent their direct analysis. To overcome these limitations, systematic methodologies are continually being developed to provide comprehensive characterization of discrete aspects of metabolism and molecular physiology. The compartmentalization of metabolomics has given rise to new analytical methods that focus on the detection of defined classes of chemicals, monitoring specific metabolic processes, or describing outcomes of chemical interactions. Structure-oriented LC-MS methodology expands the sensitivity and specificity of discovery assays for the characterization of structurally related chemical species in complex matrices. Monitoring common fragmentation pathways, predictable modifications, or shared elemental composition enables the rapid differentiation of ions of interest from the surrounding background matrix. These methods are well suited for the characterization of uncommon molecular signatures often overlooked in standard metabolomic workflows.

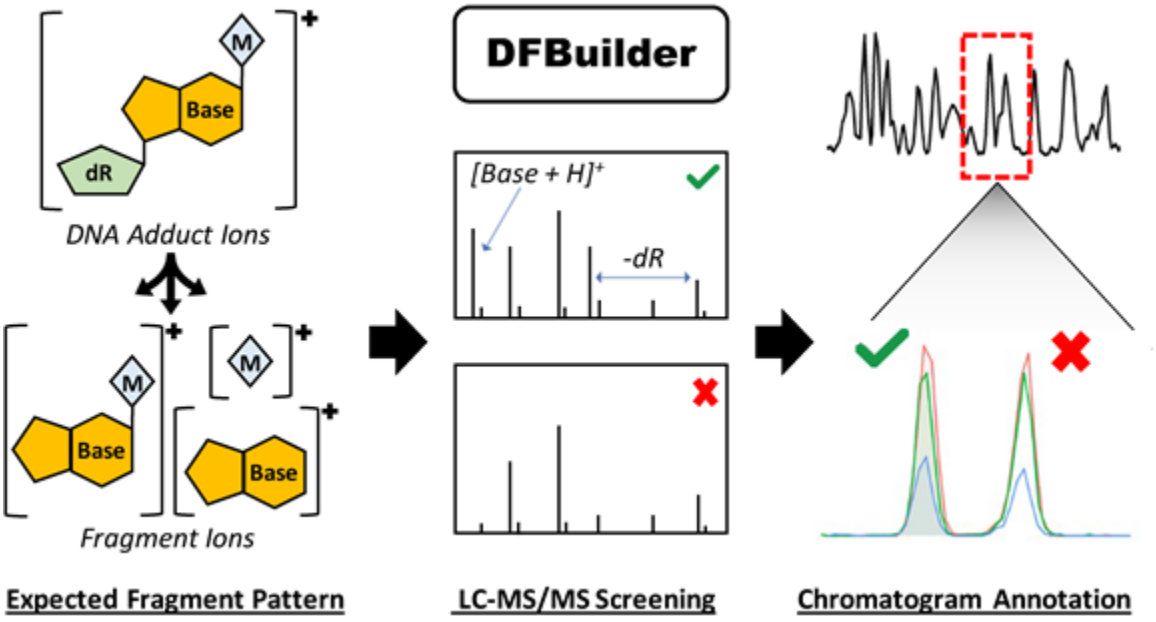
The development of structure-oriented methods expands the versatility of LC-MS analytical platforms for the characterization of the outlying elements of metabolism. Rare and trace-level compounds, otherwise missed in shotgun metabolomics approaches, require an expanded repertoire of analytical techniques to ensure their detection in complex matrices. Continued improvements to LC-MS technologies and computational workflows are required to increase the detection rates of trace analytes in complex mixtures. In Chapter 2 of this Thesis, a novel computational workflow for the automated detection of DNA adducts is presented. The development of the DFBuilder module[63] in

MZmine provides a parameterized implementation of fragmentation filtering to LC-MS/MS experiments. The application of this module for the discovery of undescribed DNA adducts is discussed. Chapter 3 of this Thesis introduces a novel analytical workflow for the unbiased detection of urinary mercapturic acids. This workflow combines a novel LC-MS/MS analytical method in conjugation with the combined application of fragmentation filtering and MDF. The application of this workflow to characterize mercapturic acids associated with cigarette use is discussed. In Chapter 4 of this Thesis discusses current computational methods for the characterization and annotation of unknown chemical signatures. Using natural product reference material and spectral tree MS^n acquisition, emerging algorithms to predict chemical structures are reviewed. In the final chapter of the thesis provides a perspective on the state of analytical approaches used in systems toxicology and the future directions of the field. Emerging computational methods and their potential applications for the detection of the molecular biomarkers of exposure are discussed.

Chapter 2: Extension of Diagnostic Fragmentation Filtering for Automated Discovery in DNA Adductomics

The following chapter introduces a novel MZmine module for the automated processing of structure-oriented LC-MS/MS results by fragmentation filtering. The original article was authored by Kevin Murray and edited with the assistance of all co-authors.

Murray, K. J.; Carlson, E. S.; Stornetta, A.; Balskus, E. P.; Villalta, P. W.; Balbo, S. Extension of Diagnostic Fragmentation Filtering for Automated Discovery in DNA Adductomics. Anal. Chem. 2021, 93 (14), 5754–5762. https://doi.org/10.1021/acs.analchem.0c04895.



2.1 Introduction

Exposures to most genotoxic chemicals induce covalent modifications of DNA that, when not repaired, represent a major risk factor for disease pathogenesis and the development of cancer.[156–159] A wide array of distinct adduction products result from alkylation, oxidation, and deamination reactions involving modifying agents.[160] The simultaneous identification of multiple DNA adducts in complex biological matrices represents a major bottleneck for the comprehensive investigation of the genotoxic effects of exposures. The high sensitivity and structural elucidation capability of LC-MS/MS enables the rapid characterization of known and unknown modifications.[25,159,161] Application of this technology for the discovery of DNA modifications led to the development of DNA adductomics, a systems biology approach designed to identify and annotate the multitude of known and novel adduction products induced by the effects of the combination of harmful exposures on DNA.[23]

A variety of LC-MS/MS techniques may be applied to characterize DNA adducts present in common biological samples, including tissues, blood, or cell cultures.[23] Prior to analysis, DNA is isolated—typically hydrolyzed to nucleosides—enriched, and purified to ensure sufficient signal for MS analysis. Often, clean-up steps such as solid phase extraction are used to remove unmodified bases and other sample components that confound the results.[162] The most common LC-MS/MS techniques for DNA adducts analysis monitor known adduct precursor-to-product ion transitions using selected reaction monitoring or parallel reaction monitoring experiments.[163] Although offering high sensitivity for quantitation of low abundant ions, these techniques require an *a priori* defined precursor mass, precluding the detection of novel DNA adducts.[159,161] Alternatively, untargeted screening of precursor-product ion transitions is increasingly

applied for the unbiased assessment of complex matrices. It has been observed that the neutral loss of the 2´-deoxyribose moiety during collision CID is a highly selective signature of deoxynucleosides.[164] Monitoring this predictable fragmentation pattern via CNL scans serves as the foundation of many DNA adductomic strategies, including DIA and DDA methodologies. A DIA approach provides near-full coverage of complex matrices by simultaneously assessing multiple compound fragmentation patterns using wide-window precursor isolation.[111] These techniques, however, offer limited structural information of low abundant ions. Additionally, DIA methods are prone to false discovery resulting from the co-fragmentation of multiple ion species in each successive MS/MS scan.[165] Robust sample-specific spectrum libraries ease the deconvolution and detection of putative DNA adducts but the generation of these libraries require considerable time and effort. DDA strategies provide more precursor structural information necessary for compound identification. However, insufficient scanning speed of complex sample matrices limits the ability of this approach to fragment low abundance ions and discovery of trace-level adducts.[111]

A DDA-CNL-MS^3 DNA adductomic approach with high-resolution/accurate mass MSn fragmentation has recently been developed.[88,161] The structural elucidation potential offered by MS^2 and MS^3 fragmentation provides valuable insight into unknown modifications and simultaneously controls for false-positive detection using accurate mass monitoring. However, the success of this approach has been limited by the traditional data analysis strategy, which uses the appearance of an MS^3 event as an indicator of a putative DNA adduct. This approach limits the flexibility of the methodology because alternative neutral losses and product ions in the MS^2 spectra cannot be identified and is dependent upon the fidelity of the instrument in the triggering of the MS^3 data acquisition upon observation of the criteria neutral loss.

Diagnostic fragmentation filtering is a computational method that closely resembles CNL-MS/MS data acquisition methods.[102] The algorithm searches for the presence of diagnostic pattern in all MS^2 spectra, exporting the corresponding precursors of interest. In contrast to instrument CNL methods, diagnostic fragmentation filtering is applied post-acquisition and facilitates repeated analysis with new fragmentation patterns without the need for sample re-injection. Application of this approach in the field of microbial natural products chemistry demonstrated the potential of fragmentation monitoring to detect known and novel products in complex biological matrices.[102] Similarly, diagnostic fragmentation scanning approaches are utilized in LC-MS/MS proteomics to characterize peptides and proteins with complex post-translational modifications, such as glycopeptides.[166,167] To the best of our knowledge, application of diagnostic fragmentation filtering for the discovery of DNA adducts is yet to be explored. Instrument-independent data analysis for DNA adducts overcomes the limitations of previous screening strategies and enhances the application of DNA adductomics to new experiments.

In this study, we adapted the diagnostic fragmentation filtering approach to enable an automatic data processing workflow for the discovery of DNA adducts using a newly developed MZmine module,[168] called DFBuilder. The module scans all input LC-MS/MS spectra for any number of user-specified fragmentation patterns and exports a feature list of targeted-extracted ion chromatograms (EIC) for precursors of interest. The resulting feature lists are further processed to ensure only high-quality peaks remain and any detected duplicate, isotope, and in-source fragments are removed. The strength of our workflow is its capacity for fully automated and reproducible data analysis via batch processing relative to manual processing. All workflow components and corresponding parameterization are saved in an XML file, which can be easily reprocessed or shared between experiments.[168] Our approach is only dependent on computer processing

resources and enables the scaling up of experimental designs that would otherwise require days of manual data processing time. Here, we establish the effectiveness of our workflow for the automated discovery of DNA adducts and demonstrate its potential to expand screening strategies in future experiments.

2.2 Experimental Section

2.2.1 Chemical Standard Mixture

O^6-Methyl-2´-deoxyguanosine (O^6-me-dG) **(1)**, 8-oxo-7, 8-dihydro-2´-deoxyguanosine (8-oxo-dG) **(2),** N^6-hydroxymethyldeoxyadenosine (N^6-Me-dA) **(6)**, 1, N^6-etheno-2´-deoxyadenosine (ε-dA) **(7)** were purchased from Sigma-Aldrich (St. Louis, MO). N^2-Ethyl-2´-deoxyguanosine (N^2-ethyl-dG) **(3)**, (6R/S)-3-(2´-deoxyribos-1´-yl)- 5,6,7,8-tetrahydro-6-hydroxypyrimido[1,2-a]-purine-10(3H)one (OH-PdG) **(4)** and O^2-[4-(3-pyridyl)-4-oxobut-1-yl]thymidine (O^2-POB-dT) **(8),** D_5-ethyl-2´-deoxycytidine (D_5-ethyl-dC) **(9)** were prepared as described.[169–172] 6-(1-Hydroxyhexanyl)-8-hydroxy-1, N^2-propano-2´-deoxyguansine (HNE-dG) **(5)** was generously donated by Dr. Fung-Lung Chung of Georgetown University Medical Center. The nine standards were dissolved in 20% methanol and combined at a final concentration of 10 fmol/µL, respectively. The standards mixture was prepared in triplicate for LC-MS analysis. All solvents were LC-MS grade and were purchased from Sigma-Aldrich.

2.2.2 DNA from HeLa Cells Exposed to *pks*$^+$ *E. coli*

The previously acquired data was obtained with permission and a complete research protocol has been previously described.[173] The BAC*pks* island (*pks*$^+$) and empty pBeloBAC (*pks*$^-$) bacterial artificial chromosomes were used to generate derivative parent strains of *E. coli* harboring colibactin biosynthesis genes. HeLa cells were transiently

infected with each strain and genomic DNA was isolated for DDA-CNL-MS^3 analysis. Results from the DNA adductomics analysis of one replicate of *pks*$^+$ and *pks*$^-$ were used in the DFBuilder workflow.

2.2.3 LC-MS Parameters

All analyses were conducted using identical chromatographic conditions and MS instrument settings, unless otherwise described. An UltiMate™ 3000 RSLCnano HPLC system (Thermo Scientific, Waltham, MA) was interfaced to an Orbitrap Fusion™ Tribrid™ MS (Thermo Fisher Scientific, San Jose, CA). One microliter of the authentic DNA standard mixture and five microliters of *E. Coli* DNA extracts were injected onto the analytical platform equipped with a 5 µL injection loop. Solvent blanks were analyzed before and after acquisition to assess contamination and sample carryover between injections. Chromatographic separation was performed using a custom-packed capillary column (75 µm ID, 20 cm length, 10 µm orifice) using a commercially available fused-silica emitter (New Objective, Woburn MA) containing Luna C18 (Phenomenex Corp. Torrance, CA) stationary phase (5 µm, 120 Å). The LC solvents were (A) 0.05% HCO_2H in H_2O and (B) CH_3CN solutions. The flow rate was 1000 nL/min for 5.5 min at 2% B, then decreased to 300 nL/min with a 25 min linear gradient from 2 to 50% B, an increase to 98% B in 1 min, with a 4 min hold and a 5 min equilibration at 1000 nL/min to the starting conditions. The injection valve was switched at 5.5 min to remove the sample loop from the flow path during the gradient. A Nanospray Flex ion source (Thermo Fisher Scientific) was used with a source voltage of 2.2 kV and capillary temperature of 300°C. The S-Lens RF level setting was 60%.

Untargeted DDA-CNL-MS^3 analyses were performed with full-scan detection followed by MS^2 acquisition and constant neutral loss triggering of MS^3 fragmentation.

Full-scan detection was performed using the Orbitrap detection at a resolution of 60,000, automatic gain control (AGC) targeted setting of 2×10^5, and a maximum ion injection time setting of 118 ms. Full scan ranges of 300 – 1000 *m/z* and 150 – 1000 *m/z* were used for the *pks*$^+$ infected cells and the authentic standards, respectively. MS^2 spectra were acquired with quadrupole isolation of 1.5 *m/z*, fragmentation of the top 10 most intense full scan ions with Orbitrap detection at a resolution of 15,000, an AGC setting of 5×10^4, and a maximum ion injection time of 200 ms. The analysis of authentic standards utilized CID fragmentation with a constant collision energy of 30% and maximum ion injection time of 75 ms. The analysis of *pks*$^+$ infected cells utilized a HCD fragmentation with a stepped collision energy of 5%, 15%, and 25% and maximum ion injection time of 200 ms. Data-dependent parameters were as follows: a triggering threshold of 2.0×10^4, repeat count of 1, exclusion duration of 15 s. An exclusion mass list of the most intense ions observed in calf thymus DNA (ctDNA) were excluded from fragmentation in the analysis of HeLa cells treated with *pks*+ *E. coli* (±5 ppm). No masses were excluded in the analysis of the authentic standards. MS^3 HCD fragmentation scans (2.5 *m/z* isolation width, collision energy of 30%) with Orbitrap detection at a resolution of 15,000 was triggered upon observation of neutral losses of 116.0474, 151.0494, 135.0545, 126.0429 and 111.0433 *m/z*. A minimal product ion signal of 1.0×10^4 was used. All spectra were acquired with the EASY-IC lock mass (202.0777 *m/z*) enabled.

2.2.4 Data Processing

Raw data files were converted to mzML format and centroid mode using MSConvert (ProteoWizard) and imported into MZmine.[168,174] All MZmine data processing utilized a mass tolerance of 5 ppm. Automated detection of DNA adducts was performed using the DFBuilder module. Data-dependent DFBuilder parameters were matched to the

scan range and chromatographic parameters of each experiment, respectively. Diagnostic ion thresholds were matched to the CNL-MS^3 triggering threshold, detailed above. Extracted ion chromatogram and retention time (RT) tolerance of 1.5 and 0.5 min were used for the analysis of authentic standards and *pks* infected cells, respectively. An exclusion list of contaminant and background signals was constructed by analyzing blank injections. EICs were deconvoluted using the Local Minimum search algorithm, including a chromatographic threshold of 30% and minimum peak top/edge ratio of 5. Duplicate peaks and isotopes were grouped with a retention time tolerance of 0.1 min. Putative adduct peaks were aligned between raw data files using the Join Aligner algorithm with a tolerance of 0.3 min. Compound cation adducts, neutral losses, fragments, and complexes were annotated. Missing peak values were estimated using Gap-filling with a retention time tolerance of 0.3 min. Features detected in all three DNA adducts standards mixtures were retained in the final feature list, and putative adduct features unique to the *pks*$^+$ transiently infected HeLa cells were retained in the feature list. The final feature list was exported to a CSV file for manual review. A complete list of processing parameters is summarized in **Table 2.1**.

Table 2.1 - MZmine2 Processing Parameters for DNA Adduct Detection

Task (Parameter -> Value)	Standard's Experiment	*pks*$^{+}$ *Experiment*
Raw Data Import		
DFBuilder		
Scan Range	6.00-40.00 min	
m/z tolerance	5 ppm	
Diagnostic Feautre List	CSV of Nucleobase Transitions	
Mininimum Ion Intensity (Relative Abundance)	25000	10000
Retention Time Tolerance	1.5	0.5
Exclusion List (T/F)	T	F
Chromatogram deconvolution - Local Minimum		
Chromatographic threshold	30%	
Min relative height	10%	
Min absolute height	1.00E+04	1.00E+05
Min ratio of peak top/edge	3	5
Peak duration range (min)	0.40-3.00	0.25-1.00
Peak filter		
Conditions	Keep only features with MS/MS scan	
Duplicate peak filter		
m/z tolerance	5 ppm	
RT tolerance (min)	1.5	0.1
Isotopic peaks grouper		
RT tolerance (min)	0.1	
Max charge	3	
Join Aligner		
m/z tolerance	5 ppm	
Weight for m/z	2	
RT tolerance (min)	1.5	0.5
Weight for RT	4	
Peak Finder (gap-filling)		
Intensity tolerance	100%	
m/z tolerance	5 ppm	
RT tolerance	1.5	0.5
Feature list row filters	Detected in 3 replicates	Unique to pks^{+}
Adduct/Fragment/Complex Search		
Export		

2.3 Results and Discussion

2.3.1 Workflow Overview

We developed the DFBuilder module to apply the diagnostic fragmentation filtering algorithm as part of a fully automated data processing workflow for the analysis of any data-dependent MS/MS data set. The workflow components and processing design are presented in **Figure 2.1**. The DFBuilder module monitors for user-defined product ions and neutral losses in tandem MS/MS spectra in input raw data files and builds EICs for precursors of interest (**Table 2.2**). To minimize false discovery rates, we designed DFBuilder to utilize mass tolerances, signal thresholds, and retention time limits to ensure only high-quality spectra and peaks are detected (**Figure 2.2, Top Panel**). An optional exclusion list removes contaminants detected in blank injections or previously defined signals (**Table 2.3**). The DFBuilder module exports the detected precursor targeted-EIC as feature lists for further downstream processing in the analytical workflow (**Figure 2.2, Bottom Panel**). Alternatively, the module may be used stand-alone with precursor *m/z*, retention, and tandem MS/MS scan information of interest exported to a CSV file for manual review.

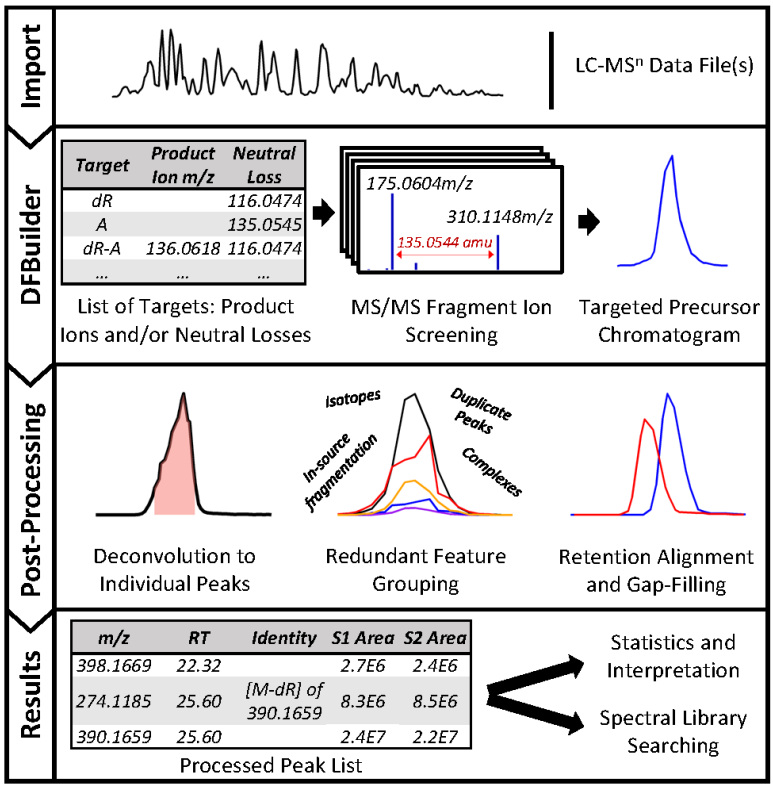
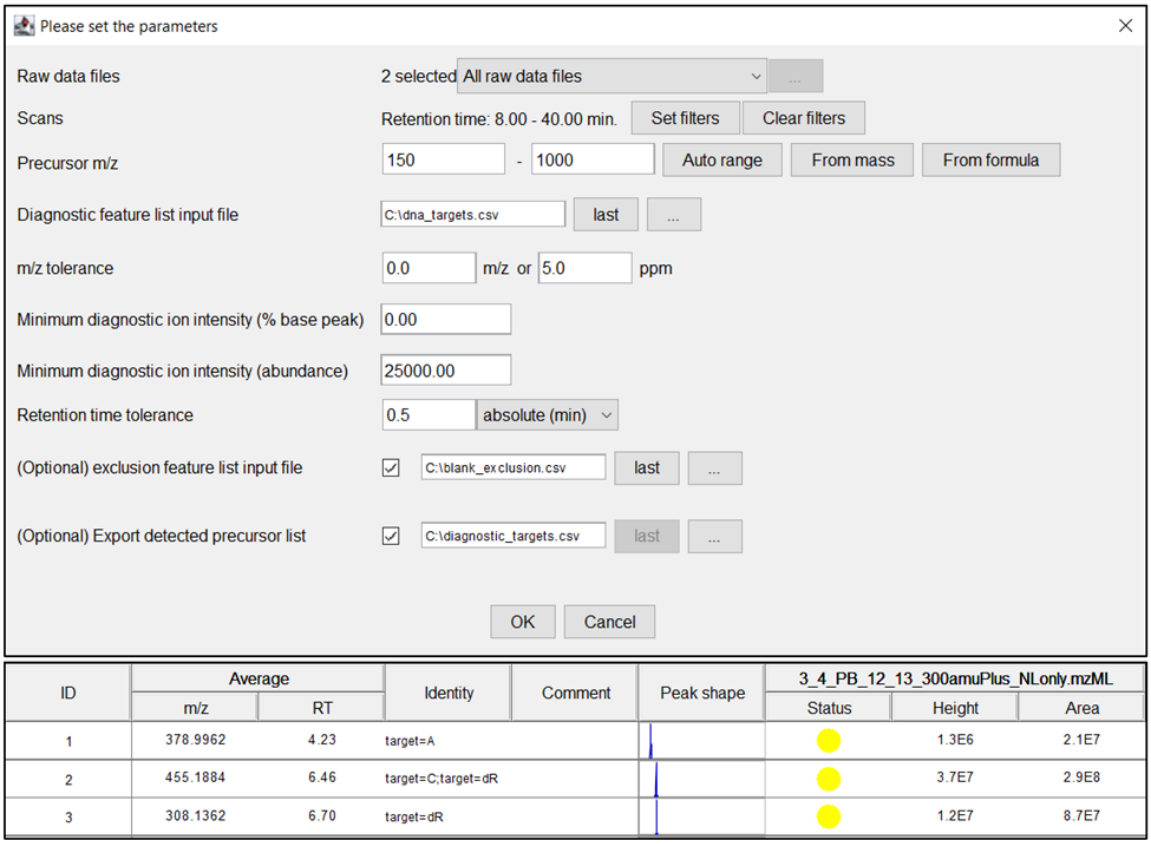


Figure 2.1 - Data analysis workflow using DFBuilder. Using a list of diagnostic fragmentation patterns, LC-MSⁿ data files are screened for putative DNA adducts and targeted-EICs are built for precursors of interest. Chromatographic features are further processed to produce a peak list of high-quality, reproducible putative DNA adducts. The final result of the workflow is a feature list of quantitative peak metrics and associated fragment spectra for each input data file that may utilized for additional statistical analyses and identification purposes.

Table 2.2 - Example Diagnostic Feature List CSV File Format

{Target Name}*	**{Product Ion (*m/z*)}***	**{Neutral Loss (amu)}***
"Target A - Product Ion"	136.0618	
"Target B - Neutral Loss"		135.0545
"Target C - Combined Search"	136.0618	135.0545
"Target D - Multiple Targets"	136.0618; 152.0567	135.0545; 151.0494

* No headers included in final CSV file. Observed headers are for information purposes only.



ID	Average		Identity	Comment	Peak shape	3_4_PB_12_13_300amuPlus_NLonly.mzML		
	m/z	RT				Status	Height	Area
1	378.9962	4.23	target=A				1.3E6	2.1E7
2	455.1884	6.46	target=C;target=dR				3.7E7	2.9E8
3	308.1362	6.70	target=dR				1.2E7	8.7E7

Figure 2.2 - MZmine2 interface of the DFBuilder module; Top Panel: Parameter setup window; Bottom Panel: Feature list export of targeted precursor-EIC results.

Table 2.3 - Example Exclusion Feature List CSV File Format

{Excluded *m/z*}*	{Start RT (min)}*	{End RT (min)}*
252.1091	0	55
376.1228	0	55
485.1878	6	20
557.1617	30	40

* No headers included in final CSV file. Observed headers are for information purposes only.

The DFBuilder module serves as the entrance to a complete analytical workflow for DNA adduct discovery and annotation using the MZmine platform. Following diagnostic fragmentation filtering, additional processing curates the resultant feature list to remove low-quality or redundant peaks. Then, chromatographic deconvolution eliminates non-reproducible features from downstream analysis and provides more accurate peak quantitation. Grouping duplicate hits, in-source fragments, and compound complexes simplifies the results to a feature list of monoisotopic precursors. Finally, retention time alignment corrects spectral drift commonly observed between repeated injections, and feature annotations inadvertently excluded during the previous steps may be estimated with Gap-filling. The end product of our DFBuilder workflow is a feature list of high-quality, non-redundant putative DNA adducts. Following data analysis, peak quantitation estimates may be exported for external statistical review, and compound identities assessed manually or with assistance from spectral library searching algorithms. The ability to *post-hoc* update diagnostic patterns and processing parameters for optimized analysis is the main strength of our workflow compared to instrument-based approaches. Through the use of open-source data formats, our approach is not limited to vendor-specific raw files and may be utilized with multiple mass detector types, including ion traps, quadrupole time-of-flight, and other hybrid instruments. The DFBuilder module and workflow are compatible with batch-based analysis to ensure computational repeatability in large-scale studies. In the following sections, we demonstrate the capability of diagnostic fragmentation filtering to detect DNA adducts in complex mixtures and the potential of our workflow to serve as part of an automated pipeline for DNA adductomics analysis. Although the following experiments focus on the detection of DNA adducts, diagnostic fragmentation filtering and the DFBuilder workflow may be applied to any class

of molecule with diagnostic fragmentation, including lipids, glycoproteins, and other complex modified compounds.

2.3.2 Automated Detection of Authentic Standards using the DFBuilder Workflow

For validation of the DFBuilder workflow, we analyzed a mixture of nine DNA adduct standards (**Figure 2.3**). The authentic standard mixture contains adducts of all four nucleobases which possess polarities with enough difference to cover a wide chromatographic range. The loss of the 2´-deoxyribose moiety (116.0474 amu) is the most selective transition for the detection of DNA nucleoside adducts using CNL screening.[175] Recent studies have demonstrated that neutral loss monitoring of the nucleobases, including cytosine (111.0433 amu), thymine (126.0429 amu), adenine (135.0545 amu) and guanine (151.0494 amu), expands the coverage to DNA nucleobase adducts.[88] The application of the DFBuilder module applied here used diagnostic fragmentation filtering of the neutral losses of deoxyribose and the four nucleobases to monitor for the presence of nucleoside and nucleobase DNA adducts in solution. Prior to the analysis of the standard mixture, a solvent blank injection was analyzed to assess the presence of low abundant contaminant signals in our system. In total, 54 MS/MS spectra exhibited diagnostic fragmentation. These contaminant signals populated an exclusion list for the subsequent analysis.

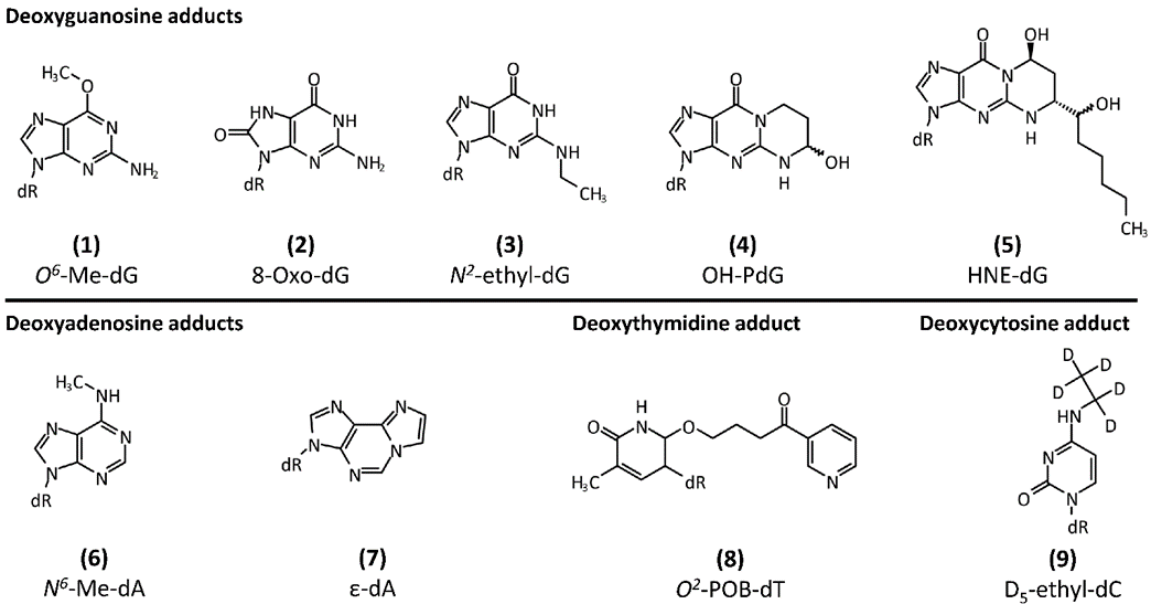


Figure 2.3 - Structures of the DNA adducts in the authentic standard mixture, dR = 2'-deoxyribose.

The application of the DFBuilder module to analyze the authentic standard mixtures detected 48 tandem MS/MS exhibiting a diagnostic neutral loss. Of these spectra, 47 exhibited a neutral loss of 2´-deoxyribose and the final signal displayed a neutral thymine nucleobase loss. Empirical evaluation of the resultant targeted-EICs revealed many of the detected spectra represented duplicate hits from repeated MS/MS fragmentation of a single precursor ion. After filtering poor-quality and duplicate peaks, 18 putative DNA adduct signals remained. Precursor annotation indicated six of these signals characterized in-source fragments or sodium adducts $[M+Na]^+$ formed during ionization. At the completion of our workflow, 12 monoisotopic precursors were categorized as putative DNA adducts in the untargeted analysis of the authentic standard mixture (**Table 2.4**). Diagnostic fragmentation filtering detected all nine authentic standards from the characteristic 2´-deoxyribose neutral loss induced during CID fragmentation. The identity for each standard was confirmed by manual inspection of the acquired MS^2 and MS^3

spectra. No apparent relationship between the authentic standards and the three unidentified signals could be made. Manual evaluation of the acquired spectra confirmed the presence of a 2´-deoxyribose neutral loss among these features. The precursor of *m/z* 211.1329 $[M+H]^+$ at 27.13 min was determined to be a persistent background contaminant that failed to trigger a fragment spectra in the blank injection. The remaining signals of *m/z* 296.1353 $[M+H]^+$ and 398.1669 $[M+H]^+$ at 28.51 and 22.32 min, respectively, represent potential impurities or contaminants in our standard mixture based on their reproducible detection and fragmentation pattern.

2.3.3 Reproducible Detection of Colibactin-Induced DNA Adducts

Colibactin is a genotoxic secondary metabolite produced by commensal and extra-intestinal pathogenic strains of *E. coli* harboring the *pks* genomic island (*pks*$^+$). Infection with colibactin-harboring microbes influences tumorigenesis in multiple mouse models of colitis-associated colorectal cancer.[176,177] Despite its strong relationship with disease, the genotoxic mechanism driving pathogenesis has remained uncertain, although recent research has revealed colibactin alkylates DNA *in vivo,* producing large genomic crosslinks and DNA adducts.[88,173] A DDA-CNL-MS^3 adductomics approach identified two diastereomeric adducts at *m/z* 540.1765 $[M+H]^+$ exclusively in cells transiently infected with *pks*$^+$ *E. coli* compared to negative controls (*pks*$^-$ *E. coli*).[173] Here we have utilized data generated in this study to demonstrate the potential of our workflow to detect DNA modifications in complex matrices.

Table 2.4 - Results of Automated DFBuilder Workflow on Mixture of Authentic DNA Adduct Standards

No.	*m/z*	Retention time (min)	Compound Identity	Neutral Loss	Neutral Mass	Neutral Formula [Predicted*]
1	261.1605	9.34	D5-ethyl-dC	dR	260.1532	C11H17D5N3O4
2	266.1247	16.62	N6-Me-dA	dR	265.1174	C11H15N5O3
3	276.1091	12.43	ε-dA	dR	275.1018	C12H13N5O3
4	282.1197	21.5	O6-me-dG	dR	281.1124	C11H15N5O4
5	284.099	15.76	8-oxo-dG	dR	283.0917	C10H13N5O5
6	296.1352	26.08	N2-ethyl-dG	dR	295.1279	C12H17N5O4
7	324.1303	16.96	OH-PdG	dR	323.123	C13H17N5O5
8	390.1659	25.6	O2-POB-dT	dR	389.1586	C19H23N3O6
9	424.2189	37.65	HNE-dG	dR	423.2116	C19H29N5O6
10	211.1329	27.13	Unknown #1	dR	210.1256	C12H18O3*
11	296.1353	28.51	Unknown #2	dR	295.128	C12H17N5O4*
12	398.1669	22.32	Unknown #3	dR	397.1596	C16H23N5O7*

* Neutral formula predicted using natural elemental and heuristic constraints with a 5 ppm mass tolerance.[178]

We applied the automated DFBuilder workflow for the discovery of colibactin-induced DNA adducts in HeLa cells transiently infected with *pks*$^+$ and *pks*$^-$ *E. coli*. Tandem MS/MS fragmentation filtering utilized all previously mentioned DNA-specific neutral losses. In total, 212 MS/MS spectra exhibited a diagnostic neutral loss. After removing non-reproducible peaks, duplicate hits, isotopes, and ion complexes, 56 features remained. Three putative DNA adduct features were unique to *pks*$^+$ infected cells and exhibited no signal among negative controls. In agreement with the original findings, we detected the two colibactin-induced DNA adducts of *m/z* 540.1765 at 16.92 and 17.50 min, respectively (**Figure 2.4**). Our workflow discovered one additional putative adduct at 7.68 min exhibiting an *m/z* of 342.1680 $[M+H]^+$ (**Figure 2.5**). As this putative adduct was not reported in the original publication, we further evaluated its properties. The putative adduct exhibited a neutral loss of adenine (135.0545 amu) in the acquired MS/MS fragment spectra (**Figure 2.5B**) and triggered a CNL-MS^3 spectrum for the resulting product peak of *m/z* 207.1134 (**Figure 2.5C**). The high-resolution accurate mass measurement of 342.1680 $[M+H]^+$ yielded a molecular formula of $C_{16}H_{19}N_7O_2$ (calculated, 342.1673) with 11 degrees of unsaturation. Tandem MS/MS fragmentation exhibited one major fragment ion of *m/z* 207.1134 $[M+H\text{-Adenine}]^+$. Product ion CNL-MS^3 fragmentation displayed three additional fragment ions of *m/z* 179.0825 $[M+H\text{-}C_2H_4]$, 152.0715 $[M+H\text{-}C_3H_5N]$, and 124.0766 $[M+H\text{-}C_4H_5NO]$. A mass tolerance of 10 ppm was used for the assignment of the fragment ions.

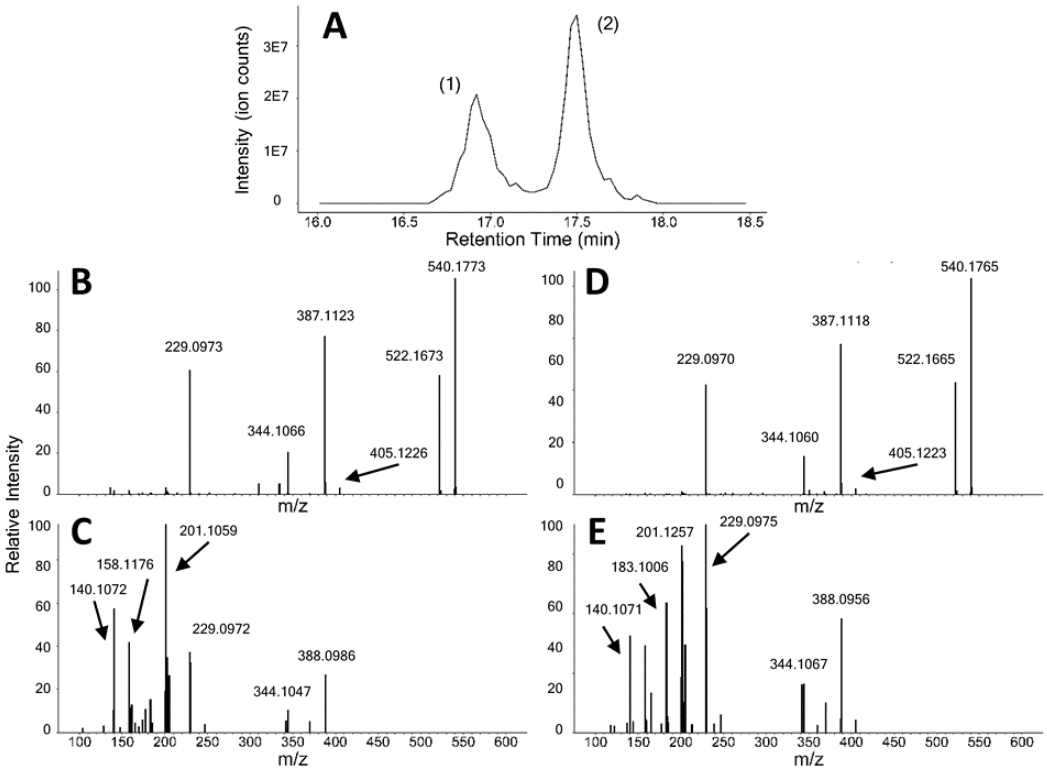


Figure 2.4 - LC-MS characteristics of m/z 540.1765 colibactin-induced DNA adduct stereoisomers; (A) EIC of precursor m/z 540.1765 [M+H]+ at 16.88 min (stereoisomer 1) and 17.50 min (stereoisomer 2); (B) MS^2 precursor fragmentation of stereoisomer 1 exhibiting the neutral loss of adenine; (C) CNL-triggered MS^3 on product ion m/z 405.1226 of stereoisomer 1; (D) MS^2 precursor fragmentation of stereoisomer 2 exhibiting the neutral loss of adenine; (E) CNL-triggered MS^3 on product ion m/z 405.1223 of stereoisomer 2.

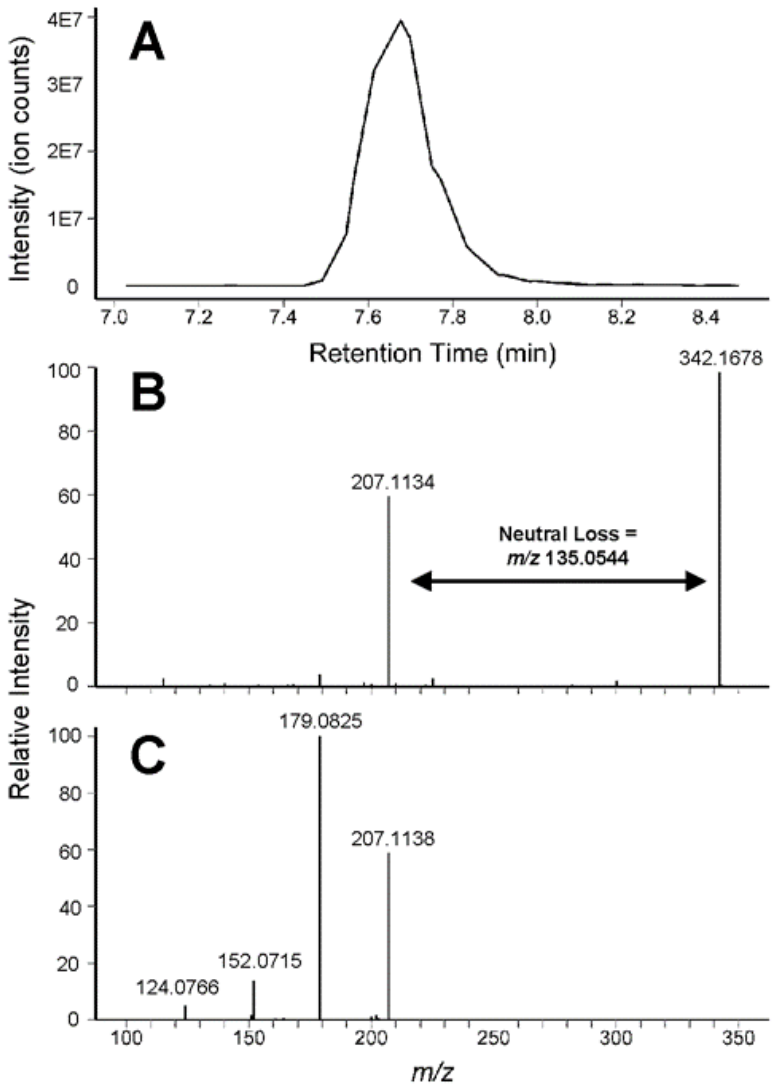


Figure 2.5 - LC-MS characteristics of novel colibactin-induced DNA adduct; (A) EIC of precursor m/z 342.1673 [M+H]+ at 7.71 min; (B) MS^2 precursor fragmentation exhibiting the neutral loss of adenine; (C) CNL-triggered MS^3 on product ion m/z 207.1134.

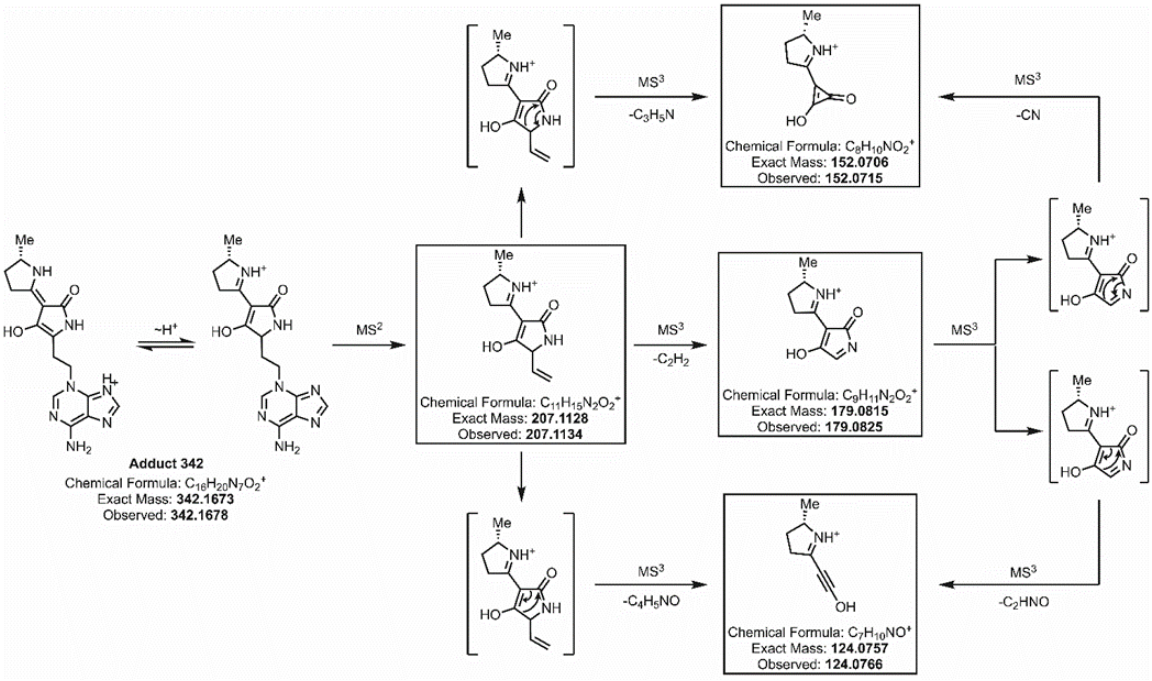


Figure 2.6 - Proposed structure of adduct 342 and mechanism for MS^2 and MS^3 fragmentation. The proposed structures of all observed ions are enclosed and accompanied by their chemical formula, exact mass, and the m/z experimentally observed by LC-MS analysis.

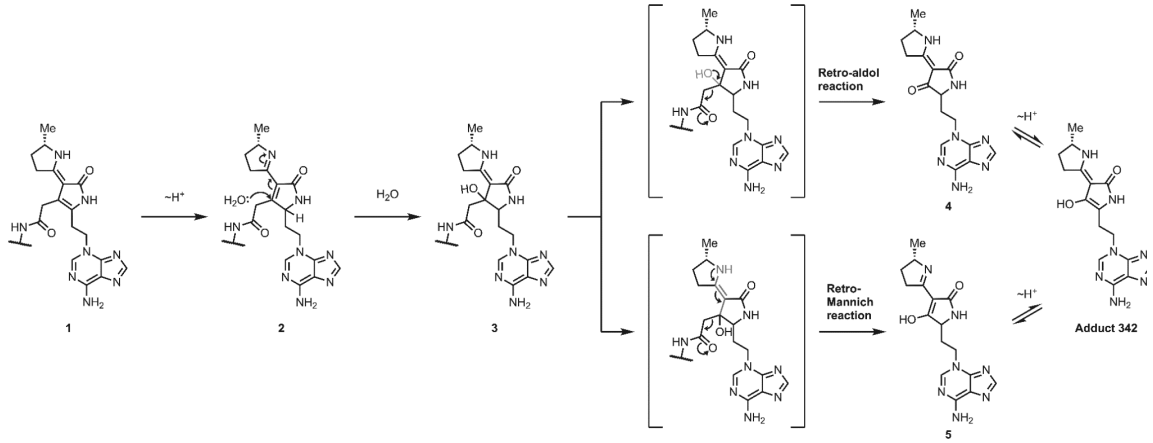


Figure 2.7 - Proposed mechanism for the production of adduct 342. After initial tautomerization of 1 to compound 2, water attacks into the electrophilic, unsaturated carbonyl system to produce 3. This intermediate then decomposes by either a retro-aldol or retro-Mannich type reaction, releasing compounds 4 or 5, respectively. Final aromatization of either compound yields the proposed structure of adduct 342.

Based on these mass spectra, we propose the *m/z* 342.1680 parent ion structure (adduct 342; **Figure 2.6**) to be a 4-hydroxy-pyrrolidin-2-one, which closely resembles the structures of the *m/z* 540.1765 adducts and other precolibactin metabolites.[173,179] In Figure 4, we provide a fragmentation mechanism that not only rationalizes the appearance of all detected MS^2 and MS^3 ions but also predicts their structures, which all support our proposal for adduct 342. The most striking feature of adduct 342 is the enol hydroxyl group located where a carbon side chain normally resides. We posit the hydroxyl group is introduced through formal hydrolysis either via a retro-aldol or retro-Mannich type mechanism (**Figure 2.7**). Importantly, this mechanism suggests adduct 342 is a novel decomposition product of colibactin and its DNA crosslinks and offers another explanation for colibactin's notorious instability. However, in the absence of internal standards and additional experimentation, we are currently unable to unambiguously confirm the structures of adduct 342 and its fragment ions. Additionally, lack of standards limits our ability to quantify adduct 342 and understand its relative contribution to colibactin DNA adduct decomposition and its potential as a biomarker.

2.3.4 Expanding the Detection of Putative DNA Adducts

The strength of diagnostic fragmentation filtering is the ability to alter analysis parameters post-acquisition and probe any diagnostic fragmentation patterns characteristic of the compound class of interest.[102] The DDA-CNL-MS^3 screening approach used to date biases adduct discovery to low molecular weight, singly-charged ions and is limited to previously described neutral loss fragmentation pathways.[111] It has been recently observed that the ionization of DNA cross-links (two or more nucleosides connected through a covalent modification) and other bulky covalent modifications can be dominated by double-charge ionization and can exhibit more complex fragmentation

patterns not predictably characterized by single neutral loss transitions (e.g. the 2´-deoxyribose moiety).[180] In addition, the fragmentation of depurination DNA adducts lacking basic sites will result in the production of the charged nucleic acid (e.g. N^7-ethylguanine). To demonstrate the potential of our DFBuilder workflow to evaluate new fragmentation patterns, we re-analyzed the previous HeLa cell colibactin-*pks* experiment with updated monitoring criteria. We supplemented our list of fragmentation patterns to include product ions of each nucleobase: cytosine (112.0506 *m/z*), thymine (127.0502 *m/z*), adenine (136.0618 *m/z*), and guanine (152.0567 *m/z*).

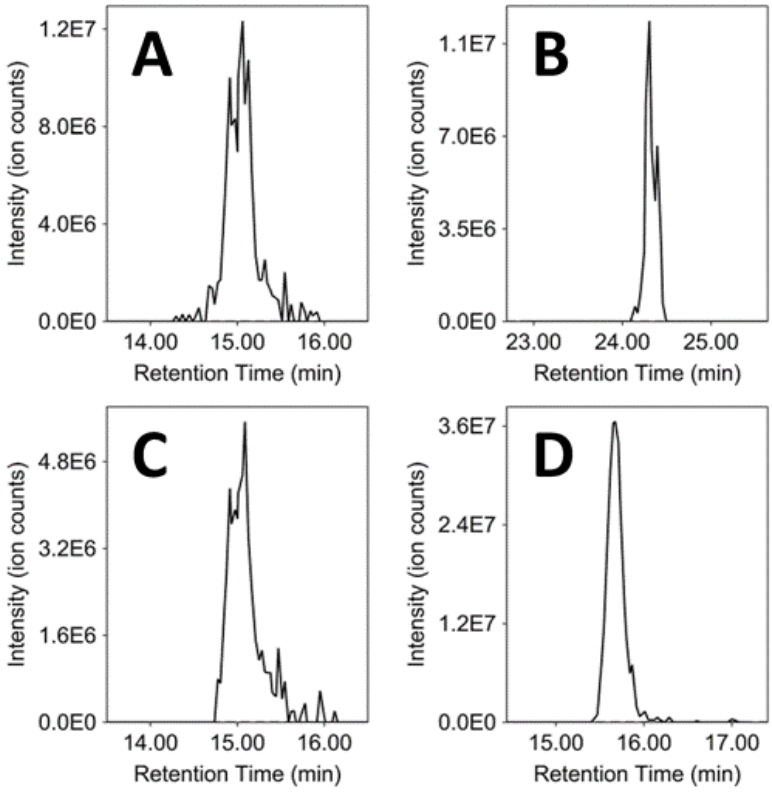


Figure 2.8 - EICs of putative colibactin-induced DNA adducts detected by adenine fragment monitoring exhibiting an m/z of; (A) 333.5243 $[M+3H]^{3+}$ at 14.91, (B) 423.6687 $[M+2H]^{2+}$ at 24.33 min, (C) 499.7857 $[M+2H]^{2+}$ at 15.09 min, and (D) 535.2576 $[M+3H]^{3+}$ at 15.67 min.

Using the modified pattern list, a total of 1,069 MS/MS spectra exhibited diagnostic fragmentation. After removing non-reproducible peaks, duplicate hits, isotopes, and ion complexes, our workflow detected 343 putative DNA adducts in the *pks*$^{+}$ and *pks*$^{-}$ infected HeLa cell DNA. The inclusion of nucleobase product ions resulted in a six times increase in the number of detected spectra relative to a CNL search. Over one-third of these signals represented doubly $[M+2H]^{2+}$ or triply charged $[M+3H]^{3+}$ ion species. In contrast, only 2 features detected via CNL monitoring were multiply charged. A total of seven putative DNA adduct features were found to be unique to *pks*$^{+}$ infected cells. In addition to the three CNL-detected features, four novel putative adducts were detected via the presence of an adenine product ion upon fragmentation, exhibiting an *m/z* of 333.5243 $[M+3H]^{3+}$ at 14.91 min, 499.7857 $[M+2H]^{2+}$ at 15.09 min, 535.2576 $[M+3H]^{3+}$ at 15.67 min, and 423.6687 $[M+2H]^{2+}$ at 24.33 min (**Figure 2.8 A-D**). In contrast to the neutral loss-derived putative adducts, all of the newly detected putative DNA adducts were multiply charged. Two of these features, *m/z* of 499.7857 $[M+2H]^{2+}$ and 333.5254 $[M+3H]^{3+}$, likely represent multiply charged variations of the same compound due to their identical neutral mass of 997.5568 amu and overlapping elution patterns (**Figure 2.8 A and C**). With only tandem MS/MS fragmentation, elucidating chemical structures for these signals is challenging. Hundreds of molecular formulae are possible for each precursor, even with high-resolution accurate mass measurement. These bulky putative adducts could represent DNA interstrand cross-links missed though CNL monitoring alone. However, these features do not correspond with any proposed colibactin-associated DNA crosslink structures.[180,181] Unlike the near-universal loss of deoxyribose for singly charged DNA adducts, the propensity of multiply charged cross-linked DNA adducts to fragment to protonated nucleobases has not been established. Therefore, it is possible these signals are false positive signals from background contamination. In the absence of a comprehensive

evaluation of the sensitivity and specificity of the nucleobase product ions, we are unable to assess the false discovery rate of this method. However, it should be noted that the ion signals attributed to the putative multiply charged cross-linked DNA adducts were not present in the negative control sample. With a quadrupole isolation window of 1.5 *m/z*, co-fragmentation of independent precursors may inadvertently occur during an analysis. A thorough evaluation will be required to fully assess the diagnostic potential of product ions for the discovery of DNA adducts.

The strength of post-acquisition monitoring of characteristic neutral loss and product ions is the expedient evaluation of any fragmentation pattern for class-specific small molecule discovery. Although the results presented here focus on the detection of DNA adducts using a hybrid Orbitrap mass analyzer, this approach may be applied to any other compound class with diagnostic fragmentation using any mass analyzer capable of generating MS/MS fragmentation in either positive or negative ion polarity. We foresee this approach being especially useful for the detection of other conjugated and covalently modified compounds indicative of chemical exposure, such as the cleavage of the thioether bond that is characteristic of mercapturic acids conjugates in negative ion mode LC-MS/MS (129.0426 amu).[53] Additionally, the DFBuilder approach may be used to discriminate specific moieties of diverse compound classes, such as saturated or unsaturated fatty acids. Using complex diagnostic pattern combinations, users may simultaneously discriminate chain length neutral losses and head group fragment ions for selective detection and characterization of glycerophospholipids and other complexes.

2.4 Conclusion

We developed the DFBuilder MZmine module and automated workflow for the detection of DNA adducts using diagnostic fragmentation filtering. Our method screens any data-dependent MS/MS spectra for fragmentation patterns (neutral losses and product ions) and builds targeted precursor EICs in the corresponding chromatographic region. Using the suite of characterization, filtering, and alignment tools available in MZmine, the resulting peak list is automatically reduced to a list of high-confidence, reproducible putative DNA adduct signals. In contrast to existing techniques, our method can be used with any tandem mass spectrometer and may be applied post-acquisition for repeated evaluation of experimental results. Using a mixture of authentic standards, we validated our workflow for the automated detection of DNA adducts with minimal false detection. We demonstrated the ability of our approach to reproduce results and discover novel putative adducts in complex mixtures. Using the DFBuilder workflow, data analysis of LC-MS/MS DNA adductomics experiments is simplified and enables further development of fragmentation filtering for the detection of DNA adducts and other compounds.

The source code repository for the DFBuilder module and our executable MZmine 2.54 forked repository is freely available on Github (https://github.com/Kevin-Murray/mzmine2). The DFBuilder module will be incorporated into main branch of MZmine in the next available update. The DNA adduct standards data used in this project are available for public download from the Metabolomics Workbench.

Chapter 3: Development of a UHPLC-MS/MS Method for Profiling of Mercapturic Acids in Positive Ion Mode

The following chapter introduces a novel analytical method for discovering and profiling mercapturic acid conjugates. This method employs a UHPLC-MS/MS method using positive ion mode detection to global profile urine metabolites. A combination of fragmentation filtering and mass defect filtering is applied to detect known and novel mercapturic acids. Computational chemical prediction assists in result curation and proposes candidate structures for biomarkers associated with tobacco cigarette usage. The following chapter is planned to be submitted to *Analytical Chemistry* as an original research article.

3.1 Introduction

Recurrent exposure to hazardous chemical agents is directly related to adverse health consequences, including neurological disorders[182,183], metabolic aberrations[184,185], and cancer[186]. The reactivity of volatile organic compounds (VOC) and endogenously produced electrophiles is of particular concern when categorizing the health risk associated with the exposures. Non-canonical covalent modifications of nucleophilic biomolecules, such as DNA, RNA, and proteins, are associated with adverse clinical phenotypes[183] and are an acknowledged source of genetic mutations[162]. Understanding the strength and toxicity of chemical exposures is essential to regulatory and occupation agencies for defining safety recommendations and remedial guidelines.[187,188]

Chemical exposure events are often characterized by the formation of detoxification products In the case of VOC or endogenous electrophiles, these processes tend to neutralize these reactive compounds by turning them into conjugates readily excreted in urine.[43,92,119,189,190] Mercapturic acids are one such class of detoxification products commonly detected in urine in the hours or days following exposure.[31,52,53,189] Their formation is mediated by Phase I metabolic modification of reactive electrophiles, Phase II conjugation to glutathione, and transformation into a water-soluble urinary metabolite for elimination.[191] Monitoring of mercapturic acids using targeted LC-MS/MS platforms is a common approach for characterizing exposure to explicit chemical agents and identify susceptibility to associated health risk.[190] Acrolein, acrylonitrile, and crotonaldehyde exposure from tobacco cigarettes are measured through the urinary concentration of 3HPMA, 2CyEMA, and 3HMPMA, respectively.[192] Classifying the total electrophilic burden imparted by exogenous and endogenous chemicals is far more challenging. Tobacco cigarette smoke is reported to contain over 7,000 chemicals,[193]

many of which are eliminated through glutathione-mediated detoxification. The most common technique for discovering novel and undescribed mercapturic acids utilizes a CNL detection approach that monitors the neutral loss of the N-Acetyl-L-Cysteine moiety (-129 Da) in LC-MS/MS experiments.[52] Chromatographic peaks exhibiting this diagnostic fragmentation pattern may be further interrogated to obtain structural configuration and quantitative insights for characterizing an exposure event or discriminating experimental populations.

However, the application of unbiased analytical methods for characterizing novel and undescribed mercapturic acids faces many obstacles. The unit mass resolution of triple quadrupole-based platforms limits the specificity of the analysis, increasing erroneous false detections. Additionally, annotating and identifying putative mercapturic acid signatures is more challenging without accurate mass product spectra. The increased popularity of HRMS has driven the development of new analytical approaches to characterize global mercapturic acid profiles. High-resolution mass measurements enable more specific monitoring of the N-Acetyl-L-Cysteine moiety (-129.0426 Da).[53] Emerging applications utilize multiple fragmentation patterns and chemical library-driven detection techniques to increase the discovery capabilities of unbiased methodology.[31] At present, all reported discovery methodologies use negative ion mode detection. Early reports in this field suggested that positive ion mode is less sensitive than negative ion mode,[52] and few documented applications using positive ion mode are available for review. While sensitivity is undoubtedly an important consideration when selecting an analytical approach, the availability of spectral library resources and structural annotation tools is worth considering. Commercially and publicly available mass spectral libraries are heavily populated with positive ion mode submissions. As such, the success of structural prediction algorithms underperform in negative ion mode submissions relative to positive

ion mode equivalents.[152] Consequently, it is worthwhile to revisit the application of positive ion mode for detecting and annotating known and unknown mercapturic acid conjugates.

We developed a UHPLC-MS/MS analytical workflow using positive ion mode detection to characterize urinary mercapturic acids. This method utilizes nanoflow chromatography to maximize the sensitivity of positive ion mode profiling of known and unknown mercapturic acids. We employed a DDA approach with a conditional neutral loss-triggered MS^3 acquisition to obtain further confidence in our detections and gain further structural insights. To expand the detection scope, we utilized MDF to discover mercapturic acids that fail to exhibit characteristic fragmentation patterns.[68] We applied our approach to classify mercapturic acid signatures associated with tobacco cigarette usage in a proof-of-principle evaluation. With the assistance of structural prediction software, we demonstrate that our approach can distinguish known and undescribed mercapturic acids from the background urinary metabolome.

3.2 Experimental Section

3.2.1 Chemicals and Reagents

D_5-N-acetyl-S-(phenyl)-L-Cysteine (D_5-PhMA), D_3-N-Acetyl-S-(2-carbamoyl-ethyl)-L-cysteine (D_3-2CaEMA), D_3-N-Acetyl-S-(2-cyanoethyl)-L-cysteine (D_3-2CyEMA), D_3-N-Acetyl-S-(2-hydroxypropyl)-L-cysteine (D_3-2HPMA), D_3-N-Acetyl-S-(3-hydroxy-propyl)-L-cysteine (D_3-3HPMA), D_3-N-Acetyl-S-(3-hydroxypropyl-1-methyl)-L-cysteine (D_3-3HMPMA) were purchased from Toronto Research Chemicals (Toronto ON, Canada). UHPLC-grade acetonitrile, methanol, and hydrogen chloride were purchased from Fisher Scientific (Hampton, NH, USA). UHPLC-grade formic acid was purchased from Sigma-Aldrich (St. Louis, MO, USA).

3.2.2 Study Participants and Urine Samples

Study participants were recruited as part of an ongoing study conducted at the University of Minnesota. Urine samples were collected throughout a 24-hour period and stored at -80 °C until the UHPLC-MS/MS analysis. Urine samples were normalized by total collection volume prior to sample preparation. This study was approved by the University of Minnesota Institutional Review Board.

3.2.3 Sample Preparation

Urine samples were prepared as previously described.[192] Urine samples (200 µL) were acidified using 40 µL of a 30% aqueous HCL solution and vortexed gently. A mixture of 10 ng D_5-PhMA, 20 ng D_3-2CaEMA, 20 ng D_3-2CyEMA, 10 ng D_3-2HPMA, 50 ng D_3-3HPMA, and 50 ng D_3-3HMPMA was added to each sample. Oasis MAX mixed mode reverse phase anion exchange solid phase extraction cartridges (60 mg, 60 u, 2 mL reservoir size) were obtained from Waters Corp. (Milford, MA, USA). Before sample introduction, the plate was preconditioned with MeOH, water, and 2% aqueous NH_4OH solutions. The samples were applied and washed with 0.7 mL 2% aqueous NH_4OH and 0.7 mL MeOH, then dried with nitrogen gas for 20 min. The plate was washed with 0.7 mL 2% aqueous formic acid before sample collection. Finally, the unfractionated urine extracts were collected using 0.7 mL 90% MeOH in 2% formic acid wash. The samples were transferred from the 96-well collection plate to fresh 1.2 mL silanized vials and dried to completeness in a SpeedVac without heat. The dried samples were stored at -80 °C until ready for LC-MS/MS analysis.

3.2.4 UHPLC LC-MS/MS Analysis

Analytical separation and detection were performed on an UltiMate 3000 RSLCnano UHPLC system (Thermo Scientific, Waltham, MA) interfaced to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA). All dried urine extracts were reconstituted using a load solvent mixture of 0.1% aqueous formic acid. Samples were spiked with a mixture Pierce™ Peptide Retention Time Calibration Mixture (Rockford, IL, USA) at a final 100 fmol/µL concentration to monitor chromatographic and retention time performance throughout the experiment. A 1% sample equivalent load in 1 µL was injected on the analytical platform equipped with a 10 µL injection loop. Chromatographic separation was performed using a self-packed C18 column (Dr. Maisch GmbH ReproSil-PUR 1.9 µm 120 Å C18aq, 100 µm ID x 45 cm length) maintained at 55 °C for the duration of the experiment. The mobile phases were (A) 0.1% aqueous formic acid and (B) 0.1% formic acid in ACN solutions. Chromatographic separation was performed using a linear gradient starting at 0% B and increased to 50% B at 40 min, 90% B at 60 min, and held for 8 min followed by a return to starting conditions. The analysis was operated at a constant flow rate of 325 nL/min. A Nanospray Flex ion source (Thermo Fisher Scientific) was used with a source voltage of 2.1 kV and an ion transfer tube temperature of 250 °C.

Discovery DDA-CNL-MS^3 analyses of experimental samples were performed using full-scan detection followed by data-dependent MS^2 acquisition and a conditional neutral loss-triggered MS^3 acquisition. All analyses were performed in positive ion mode. Full-scan detection was performed using Orbitrap detection at a resolution of 120,000, AGC targeted setting of 4×10^5, a maximum ion injection time of 50 ms, and an S-Lens RF setting of 40%. Scan ranges of 170 *m/z* – 600 *m/z* were used for full-scan detection.

MS^2 spectra were collected using a DDA design with a 2 sec cycle time in centroid mode with a 15 sec dynamic exclusion list. Fragment spectra were acquired with quadrupole isolation of 0.8 *m/z*, Orbitrap detection at a resolution of 30,000, an AGC setting of 1×10^5, and a 100 ms maximum injection time. The analysis utilized HCD fragmentation at a fixed collision energy of 30%. MS^3 spectra were collected using a conditional neutral loss design that acquired an additional fragmentation scan on ions exhibiting a neutral loss of 131.0582 Da or 105.0426 Da from the selected precursor. MS^3 spectra were acquired with a full-scan MS isolation of 1.5 *m/z* and MS^2 isolation of 2.0 m/z, Orbitrap detection at a resolution of 30,000, AGC setting of 2×10^5, and a 200 ms maximum injection time. The MS^3 analysis utilized HCD fragmentation at a fixed collision energy of 50%.

3.2.5 Data Processing

Raw instrument files were converted to mzML format and centroid using MSConvert.[174] Discovery of putative mercapturic acids was performed using a combination of fragmentation filtering and MDF[68] using the open-source software MZmine.[168] Fragmentation filtering was performed using the DFBuilder module[63] to monitor diagnostic neutral loss and fragment ions characteristic of mercapturic acid conjugates. Putative mercapturic acid detection required the observation of a neutral loss of 131.0582 Da ($C_5H_9NO_3$) or 105.0426 Da ($C_3H_7NO_3$) in acquired MS/MS spectrum. To conduct the MDF analysis, a mass defect range was defined using a 50 mDa window based on the fractional mass of protonated N-Acetyl-L-Cysteine (164.0376 *m/z*). The MDF analysis only retained features exhibiting a fractional mass between 987.6 mDa and 87.6 mDa.

A complete list of the processing parameters is summarized in **Table 3.1**. Fragmentation filtering using the DFBuilder module was performed for the specified

neutral loss using a 5 ppm mass tolerance and minimum fragment ion intensity of 1 x 10^4 with a 1% base peak threshold. A targeted chromatogram was drawn around detected precursor masses using a 1.0 min retention time window. To conduct an MDF analysis, a classical metabolomics data processing workflow was employed. Masses of interest were detected using the centroid detector and noise threshold of 1 x 10^5. Chromatogram traces were computed using the ADAP chromatogram builder module with a mass tolerance of 5 ppm.[194] For both fragmentation filtering and MDF analyses, chromatographic peaks were detected using the Local Minimum search algorithm using a 30% chromatographic threshold, 0.1 min retention time minimum search range, a minimum peak height of 1 x 10^5, a minimum peak top/edge ratio of 10, and minimum peak width tolerance of 0.1 min. Duplicate peaks were removed using a 0.1 min retention time tolerance. Feature lists were deisotoped using the 13C isotope filter module. All samples and pooled replicate feature lists were aligned using the Join Aligner module with a mass tolerance of 5 ppm and retention time tolerance of 0.4 min. Commonly observed in-source fragments, adducts, and multimer complexes were detected and removed from consideration. Missing peak annotations were replaced using the Gap-Filling algorithm. Acquired MS/MS spectra were matched against the NIST mass spectral library (version 2020) to remove unrelated metabolic signatures. Finally, the remaining features were exported as CSV files for downstream analysis. For the MDF analysis, the resultant feature list was filtered to retain features in the detailed fractional mass filtering range.

Table 3.1 - MZmine processing parameters for the detection of mercapturic acid conjugates

Task (Parameter -> Value)	Fragmentation Filtering	MDF Analysis
Mass Detection		
Feature Detection		
Chromatogram Builder Algorithm	DFBuilder	ADAP
m/z tolerance	5 ppm	
Min Ion Intensity	1.00E+04 and 1% Basepeak	1.00E+05
Neutral Loss Monitor (Da)	-131.0582 or -105.0426	-
Chromatogram building range (min)	1.0	60.0
Chromatogram deconvolution - Local Minimum		
Chromatographic threshold	30%	
Min absolute height	1.00E+05	
Min ratio of peak top/edge	10	
Peak duration range (min)	0.1 - 2.0	
Duplicate peak filter		
RT tolerance (min)	0.05	
Isotopic peaks grouper		
RT tolerance (min)	0.1	
Max charge	3	
Join Aligner		
m/z tolerance	5 ppm	
Weight for m/z	2	
RT tolerance (min)	0.4	
Weight for RT	4	
Adduct / Neutral Loss / Complex Search		
RT tolerance (min)	0.02 min	
NIST Spectral Library Search		
Min Dot Product	700	
Peak Finder (gap-filling)		
Intensity tolerance	1%	
RT tolerance	0.4	
Export		

3.2.6 Post-processing and Statistics

Post-processing of the discovery results was conducted to remove non-reproducible features and curate false discoveries from the list of putative mercapturic acid signatures. All post-processing and statistics were conducted using the R programming language.[195] Feature abundances were Log2-transformed to improve the normality assumption of the statistical models. Features not detected in at least 80% of the sample cohort were removed from downstream analyses. Missing abundance values were imputed using a *k*-nearest neighbors (*k* = 10) approach using the *impute* R package.[196] Using the pooled quality control samples run periodically throughout the experiment, analytical variation was normalized using a linear mixed-effects model approach with an `injection order` fixed effect and `batch number` random effect using the *lme4* R package.[197] A quadratic injection order variable accounted for the non-linearity observed in the iRT peptide standards (**Figure 3.1**). Following normalization, features that failed to exhibit a coefficient of variation of less than 20% among the pooled urine samples were excluded from the analysis. Elemental formulas were calculated using the *MassTools* R package.[198] Formula were predicted using elemental heuristics adhering to the Seven Golden Rules paradigm[199] with a minimal elemental composition of N-Acetyl-L-Cysteine (5C 9H 1N 3O 1S) and 2 degrees of unsaturation with mass tolerance of 5 ppm. Features were excluded from the analysis if no formula could be predicted under these constraints.

Unsupervised clustering of mercapturic acid profiles was conducted using principal components analysis. Variables were scaled for the assessment. Variable loading and score plots were interrogated to evaluate the unbiased stratification of the results. Multivariate differences in putative mercapturic acid profiles between smoker and non-

smoker populations were assessed using an orthogonal partial least squares discriminant analysis (OPLS-DA) approach using MetaboAnalyst 5.0.[200,201] The model fit was assessed using 1,000 permutations. Variable importance scores were computed for the top ten features of the correlated component only. Univariate associations of mercapturic acid profiles and smoking status were estimated using the Welch's T-test and a Benjamini-Hochberg multiple testing correction.[202]

3.2.7 Structural Annotation and Chemical Prediction

Molecular formulas of putative mercapturic acid signatures were assessed using *in silico* fragmentation performed in SIRIUS (v 5.8.3).[152] Chemical classification was assessed using CANOPUS to evaluate the probability discovered features could be predicted as an N-acyl-alpha amino acid.[203,204] Fragmentation fingerprints and structural predictions were computed using the CSI:FingerID.[205,206] Putative structural assignments were made considering $[M + H]^{1+}$ only to all available databases using a 5-ppm mass tolerance. Many previously reported mercapturic acid structures are not reported in the available databases queried in SIRIUS.[31] A custom database was created for the missing queries using the SMILES notation for these chemicals.

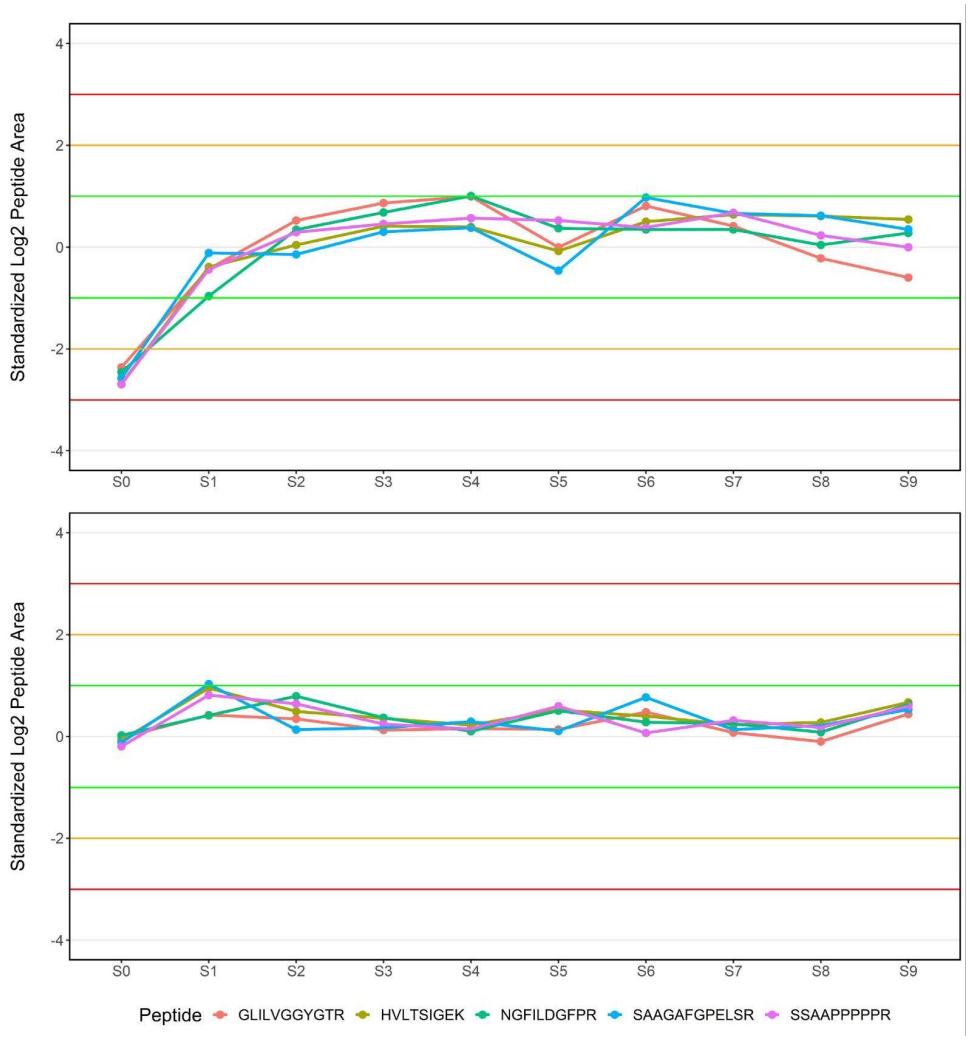


Figure 3.1 - Performance of the Pierce iRT peptides external standards in pooled quality control samples. (Top) Standardized Log2 peak area of iRT peptides mapped by injection order. (Bottom) Standardized Log2 peak area following linear mixed-effects modeling using a quadratic injection order fixed effect and batch random effect. Horizontal lines indicate one (green), two (yellow), and three (red) deviations from the standardized mean of zero.

3.3 Results and Discussion

3.3.1 Research Strategy

Historically, the discovery of mercapturic acid conjugates utilized a CNL approach monitoring for the neutral loss of 129.0426 Da ($C_5H_7NO_3$) representative of the N-Acetyl-L-Cysteine moiety in negative ion mode.[52] Emerging methods using HRMS monitor this neutral loss using a DIA-based acquisition approach to discover unexpected conjugates.[31,53] Multiple reports suggest the sensitivity of this class of molecules is higher in negative mode,[52] and there are few reported examples of the detection of mercapturic acid conjugates using positive ion mode. To overcome this limitation, we developed a UHPLC-MS/MS approach to increase the detection limits of a positive ion experiment and evaluated the discovery potential of our methodology. To establish an unbiased method, we employed a combination of fragmentation filtering and MDF to discover mercapturic acid conjugates in urine matrices and associated their quantitative profiles with smoking status (**Figure 3.2**).

We used a mixture of internal standards to test and optimize our methodology. Evaluation of reference material did not indicate the previously reported neutral loss of the N-Acetyl-L-Cysteine moiety (129.0426 Da) as being a prominent fragmentation pattern of mercapturic acid conjugates in positive ion mode. A repeated neutral loss of 131.0582 Da ($C_5H_9NO_3$) was observed among multiple evaluated compounds (**Figure 3.3**). This mass shift reflects an additional H_2 loss in the fragment ion following the disassociation of the N-Acetyl-L-Cysteine moiety. This structural rearrangement has yet to be reported as a common observation for mercapturic acid conjugates in negative mode. Compounds such as 2CyEMA did not produce this characteristic neutral loss with HCD activation but exhibited a related loss of 105.0426 Da ($C_3H_7NO_3$) derived from the loss of the N-Acetyl

and carboxyl function groups (**Figure 3.3**). These two fragmentation patterns were monitored in a neutral loss-triggered MS^3 analysis to obtain greater insights into the fragmentation patterns of the conjugated group in positive ion mode. We observed a few repeating fragmentation patterns from the analysis of the multistage activation MS^3 spectra for these compounds. The formation of a fragment ion of *m/z* 58.9950, representative of a protonated C_2H_2S functional group, was the most common pattern observed using both neutral loss-triggered MS^3 acquisition schemas. The formation of this fragment ion was not universally reproduced among the analyzed internal standards.

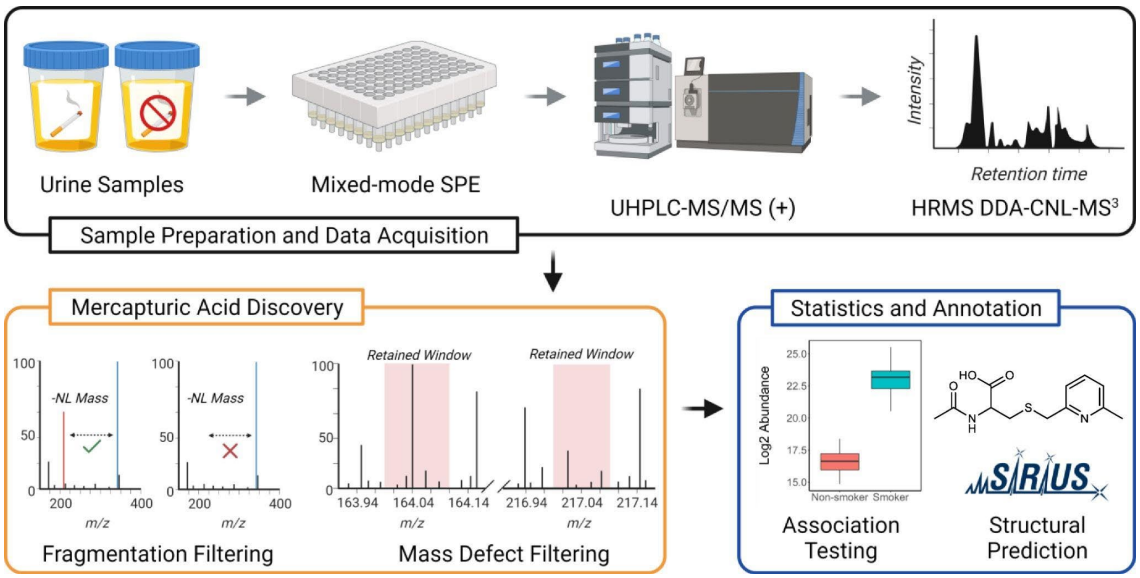


Figure 3.2 - Analytical pipeline for the detection of mercapturic acid conjugates in human urine. Samples are prepared using mixed-mode SPE and analyzed using a UHPLC-MS/MS platform generated in positive ion mode. Putative mercapturic acids are detected using a combination of fragmentation filtering and MDF. Following association testing, unknown statistically significant features are interrogated for potential structures.

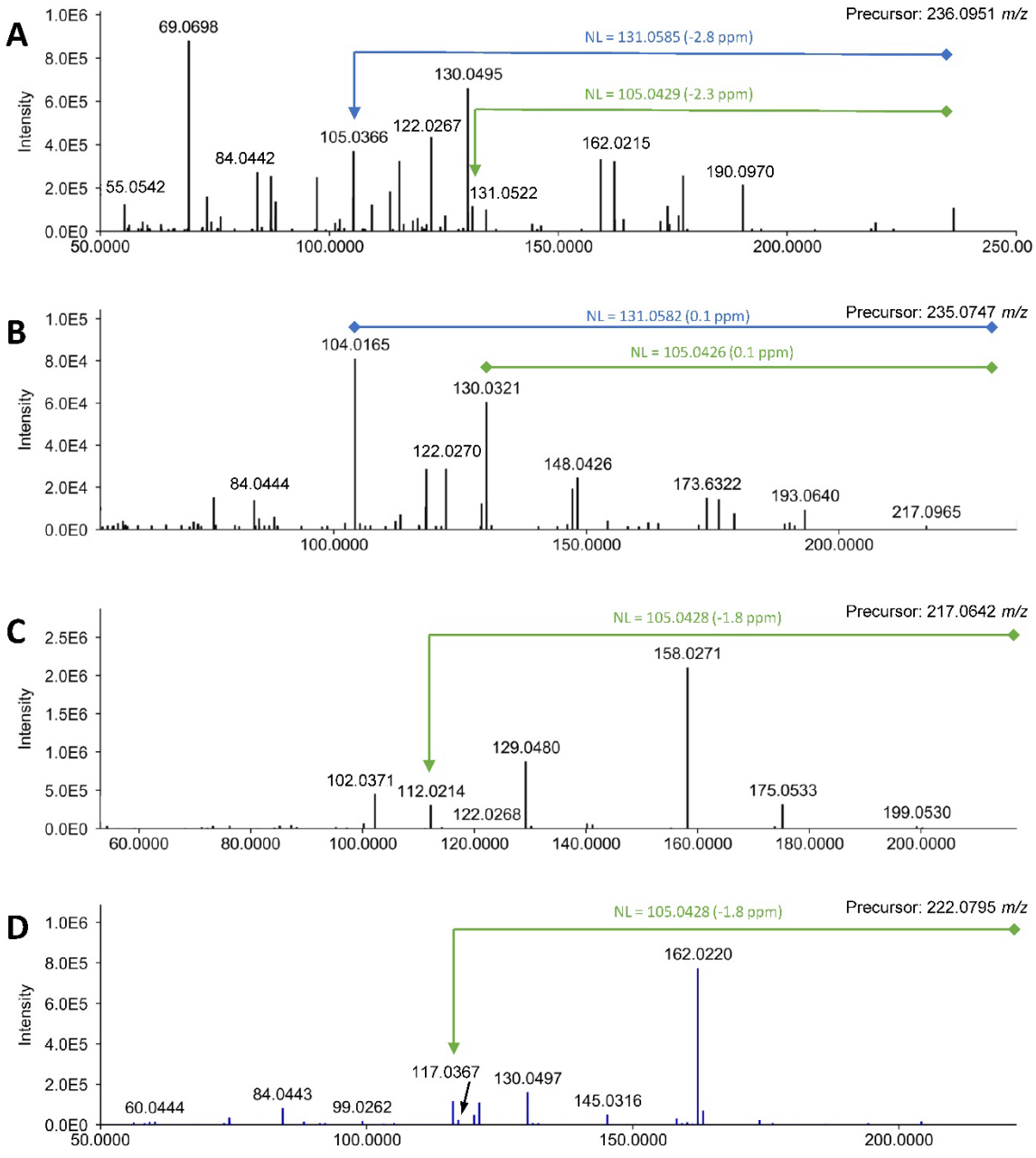


Figure 3.3 - Fragmentation spectra of mercapturic acid reference standards: (A) 3HMPMA, (B) 2CaEMA, (C) 2CyEMA, and (D) 3HPMA. Notable fragment ions and neutral loss of N-Acetyl-L-Cysteine moiety (-131.0426 Da) and combined N-Acetyl and Carboxyl functional group losses (-105.0426 Da) are indicated.

Multiple reports suggest that the neutral loss descriptive of the N-Acetyl-L-Cysteine moiety is not universal among all mercapturic acid conjugates.[30,31,53] To overcome this obstacle, an alternative approach using MDF was employed. In contrast to fragmentation-based approaches, MDF enables the discovery of unknown mercapturic acid conjugates by restricting the fractional mass range of the full-scan to a narrow 50 mDa window around the protonated mass of N-Acetyl-L-Cysteine (164.0376 *m/z).* MDF denoises and curates the resultant feature list of a global profiling experiment to discover mercapturic acid conjugates that exhibit a predictable transformation from this base structure. A primary limitation of an MDF-based discovery approach is the elevated false discovery from unrelated metabolite signatures observed in the mass defect filtering window. To control false detection, we restricted the discovery criterion to molecules that could be assigned an accurate mass molecular formula of a putative mercapturic acid conjugate.

3.3.2 Discovery of Mercapturic Acid Conjugates

An average of 100 unique chromatographic features exhibited a diagnostic neutral loss detected through our fragmentation filtering approach, ranging between 48 and 146 over the complete sample cohort. Cumulatively, this represented 1,043 unique chromatographic features representing 618 distinct mass values. The NIST MS Search software determined that the two neutral losses utilized in this experiment are commonly observed in peptide species found in urine. A small collection of these features represented independent chemical signatures putatively identified as attributable to acetylated amino acids, such as N-Acetyl-L-Tyrosine, or small di- and tripeptide species. These features were removed from consideration when possible. Following elemental formula prediction and filtering, only 230 unique putative mercapturic acid signatures were considered from the fragmentation filtering analysis. Of these features, 58 signatures were

detected using the neutral loss of N-Acetyl-L-Cysteine, 107 were detected using the loss of N-Acetyl and Carboxyl functional groups, and 65 exhibited both neutral losses in the acquired fragmentation spectrum.

To expand the discovery potential of our UHPLC-MS/MS approach, we employed MDF to detect mercapturic acid signatures as an alternative to fragmentation filtering. Using a discovery metabolomics-based workflow, 2,724 chromatographic features were reproducibly detected in at least 50% of the sample cohort. Applying an MDF approach, only 313 features exhibited a fractional mass within the specified filtering window. Following elemental formula prediction and spectral library-based filtering, only 117 putative mercapturic acid signatures remained in consideration. Roughly 25% of these masses overlapped with the results produced from the fragmentation filtering analysis. Evaluation of the non-overlapping component revealed that many of the putative mercapturic acid signatures failed to trigger an MS/MS spectrum during the experiment. In some cases, fragmentation spectra were acquired, but no diagnostic neutral loss was observed with the applied collision energies. Using MDF, we were able to discover this mercapturic acid that was missing because of the strict filtering thresholds.

The fragmentation filtering and MDF approaches discovered 316 unique putative mercapturic signatures. These features exhibited a broad mass range and elution profile over the UHPLC gradient (**Figure 3.4**). A collection of these features could be annotated as previously described mercapturic acids using accurate mass. Four of these features were validated using the internal standard spike-in mixture, including 2CyEMA, 2CaEMA, 3HPMA, and 3HMPMA (**Figure 3.5**). As previously reported,[207] multiple isomers were observed near the 3HMPMA peak, indicating that our method has potential to resolve isomeric mercapturic acid species. Continued gradient optimization will be required to achieve reproducible baseline separation of these compounds. Few fragmentation spectra

for mercapturic acid conjugates are available in positive ion mode, and structural validation is dependent on manual annotation. We prioritized our annotation efforts using the statistical results of the biomarker evaluation comparing smokers with nonsmokers.

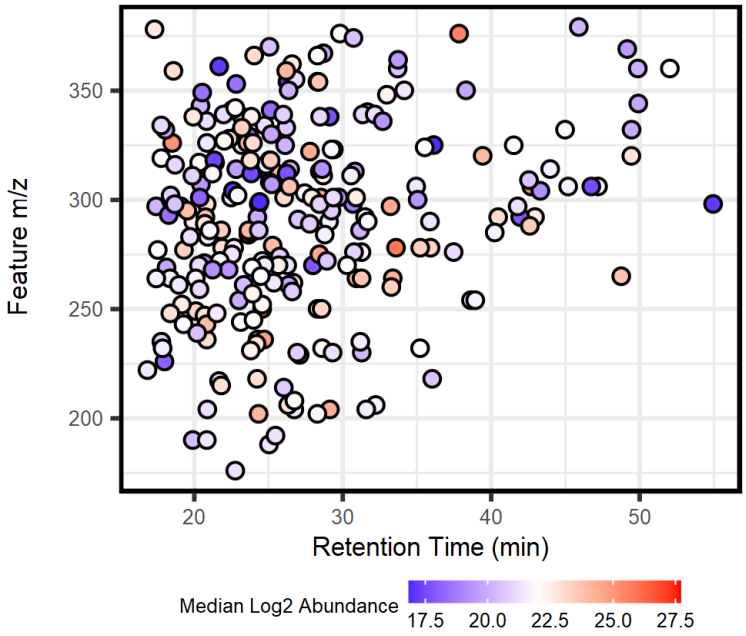


Figure 3.4 - Distribution of putative mercapturic acid discoveries (N = 316) across the UHPLC-MS/MS gradient. Feature points are colored by the observed median Log2 abundance.

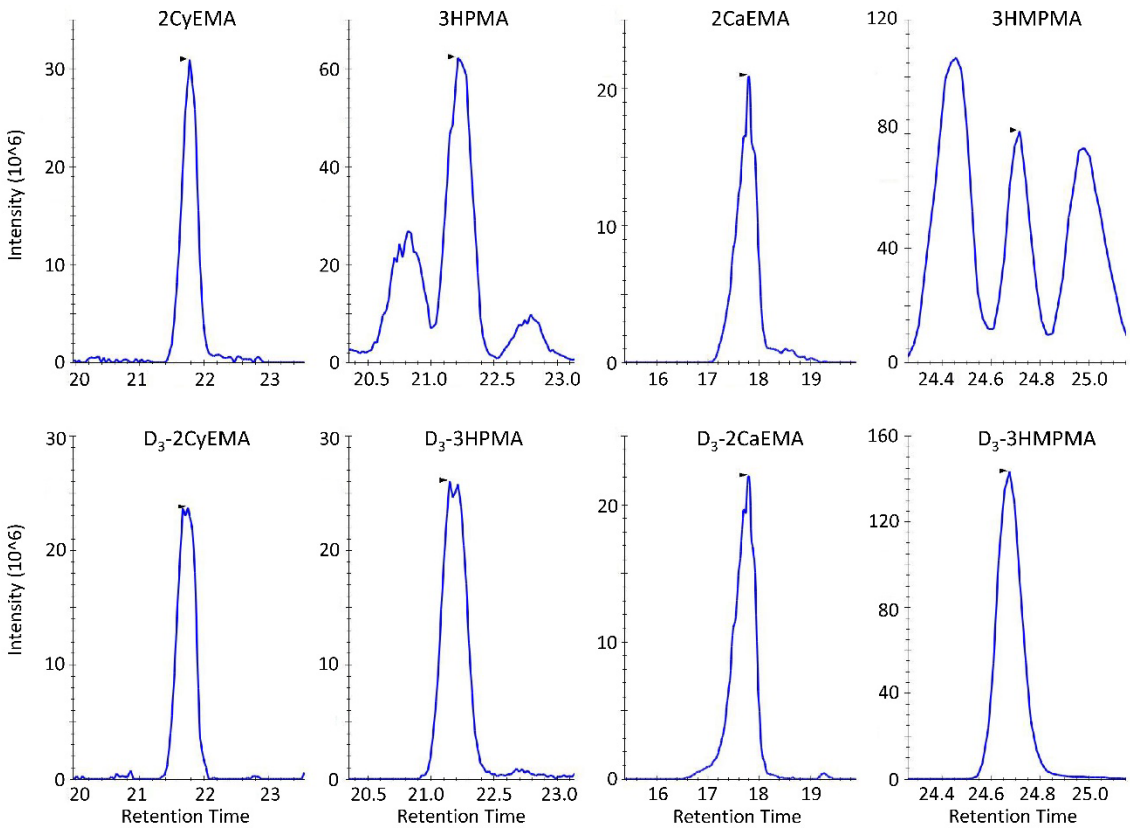


Figure 3.5 - UHPLC extracted ion chromatograms for internally validated mercapturic acids. All detected features exhibited a strong signal current and chromatographic performance.

3.3.3 Biomarkers of Tobacco Cigarette Smoke

The UHPLC-MS/MS method was validated on a proof-of-principle application using urine samples from a cohort of 20 smokers and 20 non-smoking individuals. Additional filtering was performed prior to statistical modeling to remove non-reproducible features. A minimum 80% detection rate was implemented to remove non-repeatable features, and missing values were imputed. A set of pooled quality control samples were analyzed intermittently throughout the experiment. After data normalization, putative mercapturic acid signatures were required to exhibit a coefficient of variation (CV) of less

than 20% among the pooled quality control samples following data normalization. In total, 214 putative mercapturic acid signatures (68%) satisfied these criteria.

Principal component analysis was performed to assess the unsupervised clustering of the cohort profiles (**Figure 3.6A**). The smoking populations formed sparse but distinct clusters. The first principal component (PC1) explained 25% of the variance but did not associate with smoking status. The second principal component (PC2) explained a lesser 9% of the total variation but drove the partitioning between the unsupervised clusters. A sizeable degree of within-group variation was observed from the unsupervised clustering, suggesting many additional cofactors beyond smoking status attributed to the variation in mercapturic acid profiles. Evaluation of the loadings suggested multiple signals were influential in the observed clustering results (**Figure 3.6B**).

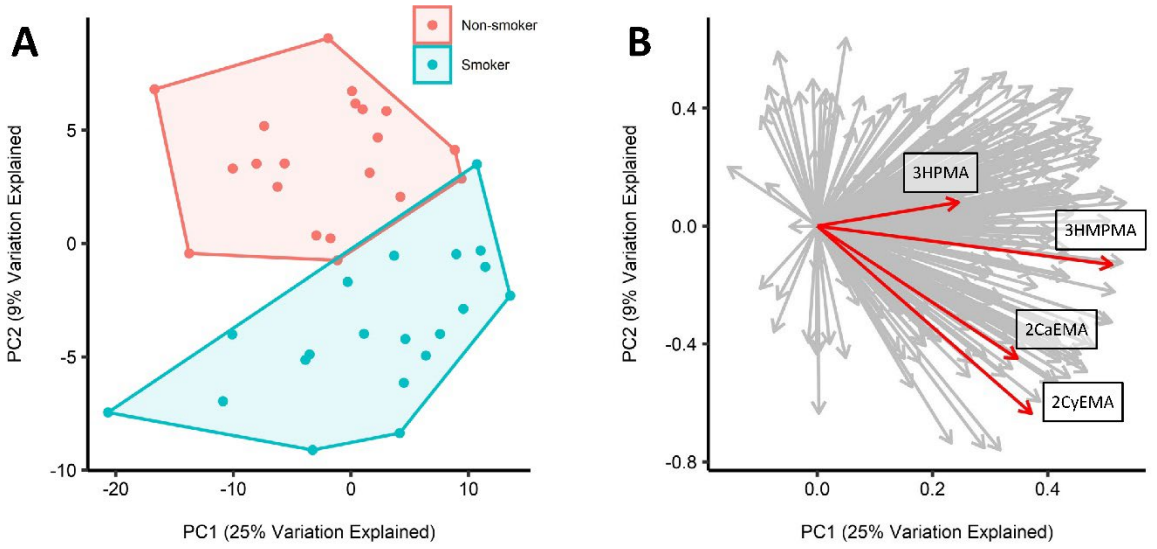


Figure 3.6 - Unsupervised clustering of mercapturic acids profiles by PCA. (A) Principal components score plots. Variables scaled before calculation. Convex hull computed for each group based on smoking status. (B) Principal components loadings plot. The four validated mercapturic acid loadings are highlighted in red.

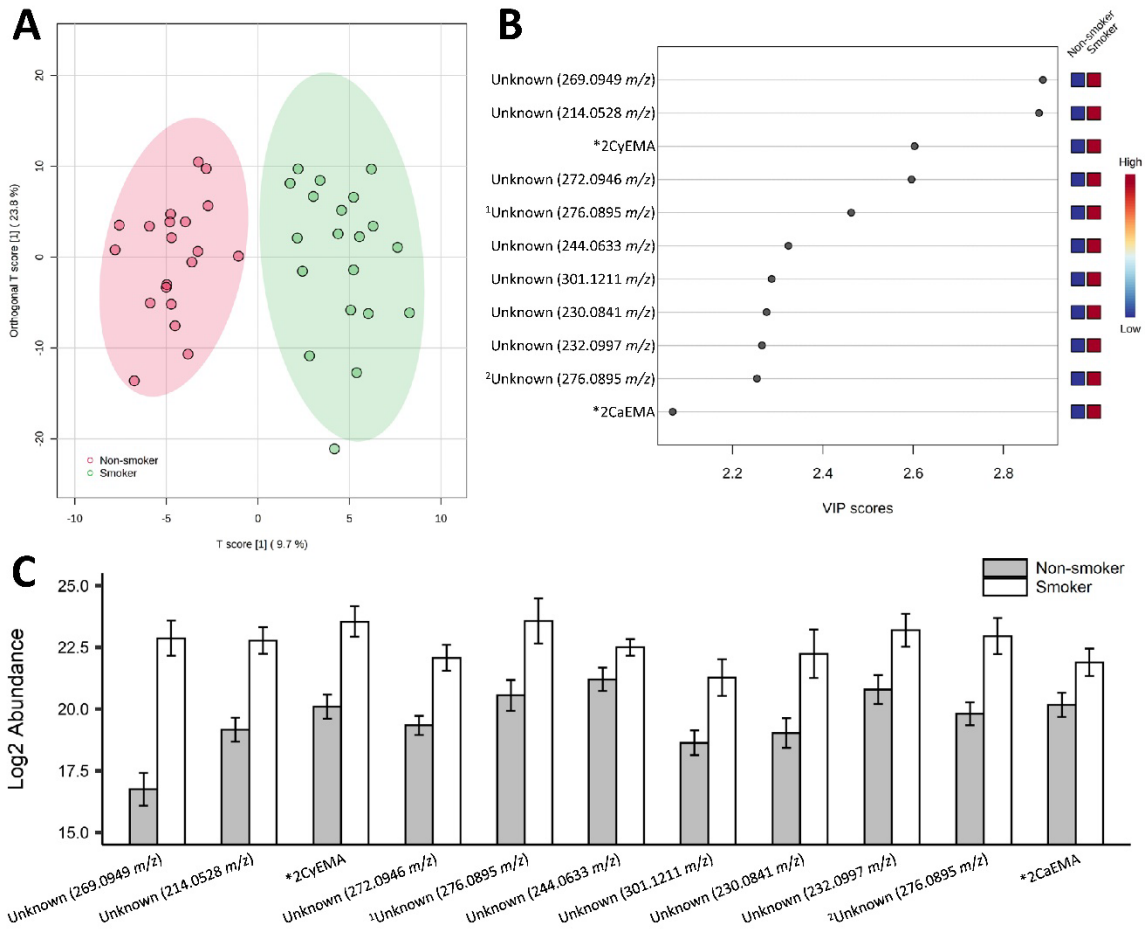


Figure 3.7 - Multivariate assessment of mercapturic acid profiles with smoking status. (A) OPLS-DA scores plot for the classification of smoker and non-smoker groups. Quantitative abundances derived from putative mercapturic acid peak heights (B) Variable importance plot ranking the most impactful features for categorizing smoker profiles. (C) Quantitative distributions of the topmost important features. Error bars represent 95% confidence intervals.

Supervised clustering using OPLS-DA demonstrated a strong classification potential (R^2Y = 0.878, Q^2 = 0.731) using the mercapturic acid profiles to discriminate smoking groups (**Figure 3.7A**). Evaluation of the variable importance plots and quantitative profiles (**Figure 3.7B-C**) indicated many putative mercapturic acids served as informative classifiers, including the validated and previously reported 2CyEMA, 2CaEMA, and 3HMPMA results.[192] A univariate evaluation of the mercapturic acid profiles indicated

these features are strongly associated with smoking status, with many exhibiting high fold changes with confident outcomes (**Figure 3.8**). The evaluation detected 38 putative mercapturic acid signatures with a significant association with smoking status following a multiple testing correction (α = 0.05). These features were selected for spectral annotation and chemical prediction to assess the likelihood of a mercapturic acid assignment and elucidation of any candidate structures.

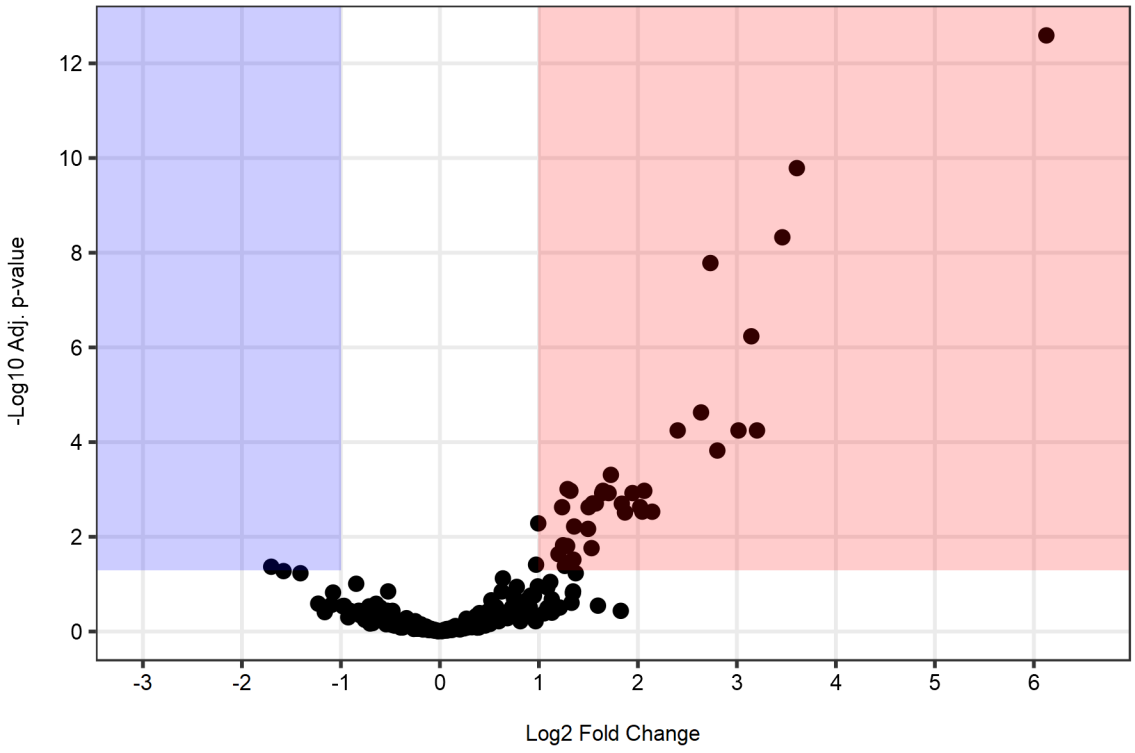


Figure 3.8 - Volcano plot of hypothesis testing of mercapturic acid profiles by smoking status. Statistical significance assessed using Welch's T-test with Benjamini-Hochberg multiple testing correction (α = 0.05). Results with statistically significantly increased abundances are highlighted in the red region and decreased abundances in the blue region.

3.3.4 Structural Annotation of Unknown Mercapturic Acid Conjugates

There are few mercapturic acid entries in available commercial or open-source mass spectral libraries, limiting the confident Level-2 identification (putative identity) of this class of molecules in discovery experiments.[208] Current annotation efforts suggest using the detection of multiple characteristic fragment ions to increase the discovery confidence to Level-4 (valid formula) or Level-3 (tentative structure) identification confidence,[31] but this approach has limitations. Many characteristic fragment ions observed for mercapturic acid conjugates are produced by other metabolites, notably peptide species that are abundant in urine matrices. While the observation of characteristic fragment ions may serve as a useful filter, it does little to assist annotation efforts of novel or undescribed targets detected in a discovery experiment. As such, we propose using SIRIUS with the CSI:FingerID and CANOPUS modules to assist with the curation of the putative mercapturic acid signatures. We evaluated discovery results by molecular formula assignment and summarized the proportion of explained peaks and total explained intensity predicted from *in silico* fragmentation. For classification assignment we used CANOPUS to evaluate probability that putative mercapturic acids could be characterized as N-acyl-alpha amino acids. Finally, we summarized the CSI:FingerID confidence and substructure scores of the top mercapturic acid assignment for each discovery.

We first evaluated the three validated and statistically significant mercapturic acids detected in this study: 2CyEMA, 2CaEMA, and 3HMPMA (**Table 3.2**). For each molecule the correct molecular formula was predicted, the majority of the observed peaks were annotated, and 75% of fragment ion intensity explained from the *in silico* fragmentation model. All fragmentation spectra exhibited patterns typically associated with N-acyl-alpha amino acids, with all probabilities greater than 80%. Interpretation of the structural

predictions required more careful consideration. The CSI:FingerID confidence score was low for all validated discoveries due to the common occurrence of structural isomers for these masses (results not shown). As such, we elected to focus on substructure coverage in combination with the estimated prediction scores to select the top candidate structure. For each reference standard, the correct structural prediction was not the most confident suggestion. However, the correct match was among the top suggestions. For example, 2CaEMA was predicted as N-Acetyl-S-(2-aminopropanoyl)-L-Cysteine and the correct assignment was the fifth suggested compound in the list. From these results, we established guidelines for interpreting the assignments of unknown mercapturic acids.

Evaluation of the 38 statistically significant findings of the univariate analysis suggested multiple discoveries could be mercapturic acids (**Table 3.2**). All findings could be assigned a plausible formula of an N-Acetyl-L-Cysteine conjugate with a high proportion of observed fragments ion annotated following *in silico* fragmentation. Using the CANOPUS module, many of the findings exhibited fragmentation patterns associated with the N-acyl-alpha amino acid classification. As noted in the evaluation of the validated mercapturic acid findings, the common occurrence of structural isomers with the reported formula limited the confidence of an unambiguous CSI:FingerID structural prediction. Regardless, 32 of 38 significant findings displayed a high substructure annotation rate with at least one mercapturic acid candidate structure. To avoid biasing the results, we only reported the most confident finding but determining the true structure will require further experimentation. The candidate structure represents Level-3 identifications supported by machine learning-based formula, classification, and structural predictions. It is important to note that the presented annotation by SIRUIS chemical prediction does not preclude the possibility of false detection. The proposed method provided evidence that some of the reported discoveries were less likely to be mercapturic acid conjugates. For example,

despite its strong association with smoking status, the feature of 214.0528 *m/z* did not exhibit fragmentation patterns associated with an N-acyl-alpha amino acid and could not be annotated to any previously reported mercapturic acid structure. This observation supports the use of a chemical prediction-based annotation approach to assist the curation of discovery results.

The experimental results were generated using a DDA-CNL-MS^3 approach to provide structural insights into conjugation group of putative mercapturic acids. Currently, the annotation of multistage activation spectra is not supported by SIRIUS; however, these MS^3 spectra can be manually imported and annotated as independent chemical structure. We first focused on the annotation of the unknown mercapturic acid of m/z 269.0949 which exhibited the strongest association with smoking status and a predicted formula of $C_{12}H_{16}N_2O_3S$. This formula is shared with the dipeptide phenylalanylcysteine but there was poor spectral similarity between these compounds. Numerous putative mercapturic acid structures exhibited a pyridine or phenylamine-based conjugation moiety for this feature with high substructure annotation. A neutral loss-triggered MS^3 from disassociation of the N-Acetyl-L-Cysteine moiety at *m/z* 138.0370 was collected for this molecule. The annotation of acquired spectra suggested a molecular formula of C_7H_7NS with 78% of the total fragment ion intensity explained by an *in silico* fragmentation. Many pyridine-based compounds were suggested from the structural prediction but a 3-pyridineethyl moiety exhibited strong substructure annotation of 77.0%. Using this observation we update the top candidate structure for this putative smoking biomarker to N-Acetyl-S-(3-pyridine-2-ethyl)-L-cysteine (**Figure 3.9**). However, the observed multistage fragmentation spectra lack the specificity to unambiguously annotate this structure. Alternative candidates for this discovery exhibited a phenylamine moiety. Numerous chemicals in tobacco cigarette

smoke exhibit pyridine or phenylamine moiety. As such, classifying the chemical source of this putative mercapturic acid compound will require further characterization efforts.

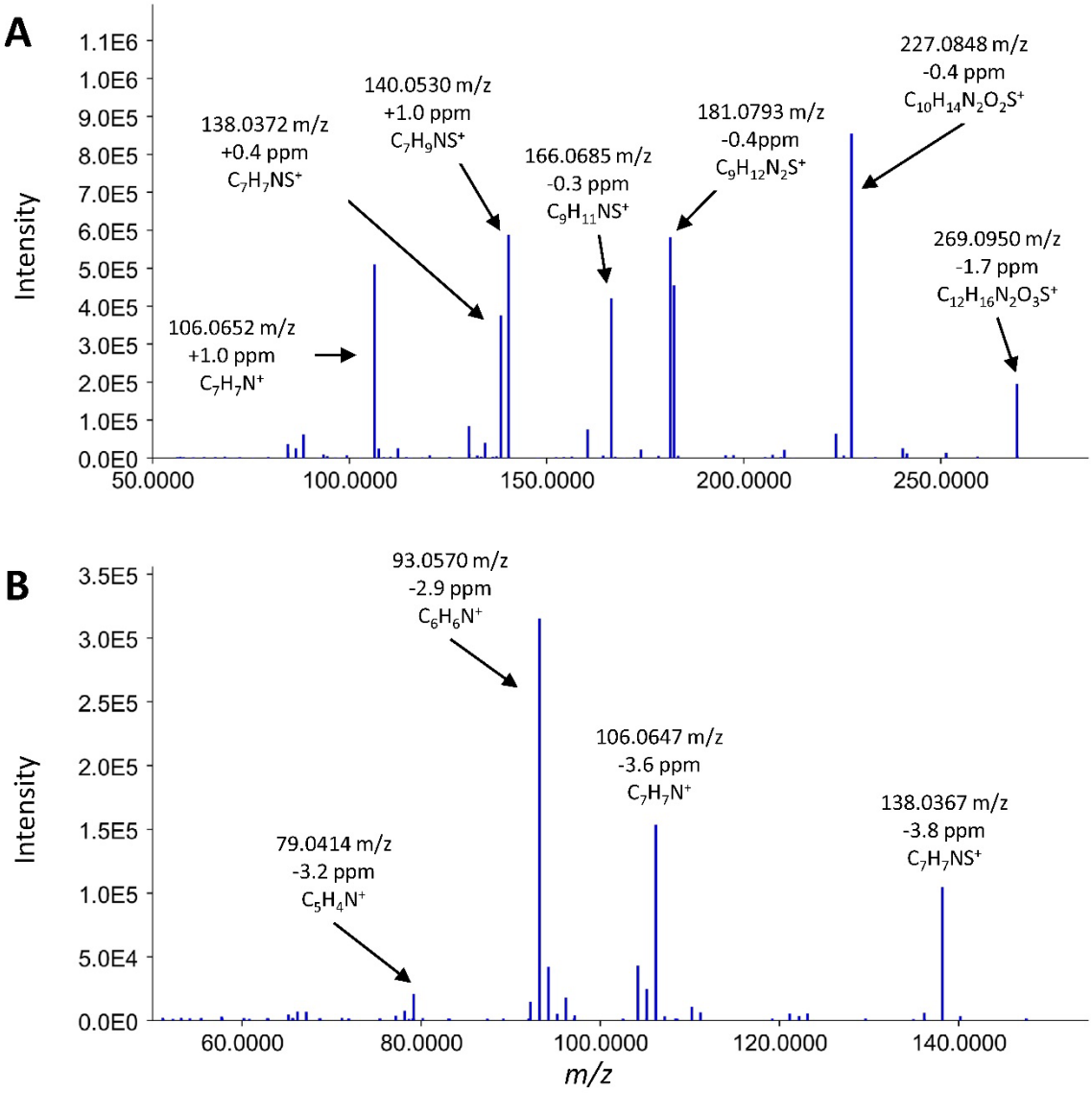


Figure 3.9 - Structural annotation of putative mercapturic acid: N-Acetyl-S-(3-pyridine-2-ethyl)-L-cysteine. (A) Annotation of MS^2 fragmentation spectra collected with HCD activation at fixed 30% collision energy. (B) Annotation of MS^3 spectra of product ion at *m/z* 138.0372 with HCD activation at fixed 50% collision energy. Substructure annotation predicted by SIRIUS v. 5.8.3.[152]

Table 3.2 - Structural Annotation of Mercapturic Acid Biomarkers using SIRIUS, CSI:FingerID, and CANOPUS

m/z	RT	Adj. P-Value	Ratio	Predicted Formula	Annotated Peaks (Explained Intensity)	Class Probability	Top Candidate Mercapturic Acid Structure	Substructure Coverage	PubChem ID
269.0949	18.1	2.6E-13	69.85	$C_{12}H_{16}N_2O_3S$	28/42 (97%)	68%	N-Acetyl-S-(6-methylpyridin-2-yl)-L-cysteine	67.03%	79267016
214.0528	26.0	1.6E-10	12.17	$C_9H_{11}NO_3S$	15/25 (92%)	*NA*	*NA*	*NA*	*NA*
217.0638	21.7	4.7E-09	11.00	$C_8H_{12}N_2O_3S$	27/30 (98%)	83%	**N-Acetyl-S-(2-cyanoethyl)-L-cysteine***	76.56%	114812
272.0946	28.9	1.7E-08	6.65	$C_{12}H_{17}NO_4S$	45/58 (90%)	96%	N-Acetyl-S-[(4-methoxyphenyl)methyl]-L-cysteine	41.52%	68736566
276.0895	30.8	5.8E-07	8.85	$C_{11}H_{17}NO_5S$	49/58 (97%)	100%	N-Acetyl-S-(4-hydroxy-2-oxocyclopentyl)-L-cysteine	66.53%	18738720
301.1211	29.7	2.4E-05	6.22	$C_{13}H_{20}N_2O_4S$	27/45 (87%)	82%	N-Acetyl-S-[2-[bis(prop-2-enyl)amino]-2-oxoethyl]-L-cysteine	61.81%	64141856
230.0841	31.3	5.7E-05	9.21	$C_{10}H_{15}NO_3S$	29/58 (83%)	98%	N-Acetyl-S-(pent-3-ynyl)-L-cysteine	92.05%	107774782
232.0997	35.2	5.7E-05	5.29	$C_{10}H_{17}NO_3S$	45/58 (96%)	98%	N-Acetyl-S-(3,3-dimethylallyl)-L-cysteine	96.15%	107774496
276.0895	31.3	5.7E-05	8.09	$C_{11}H_{17}NO_5S$	46/57 (97%)	100%	N-Acetyl-S-(4-hydroxy-2-oxocyclopentyl)-L-cysteine	65.59%	18738720
307.1316	25.2	1.5E-04	6.98	$C_{12}H_{22}N_2O_5S$	40/58 (81%)	93%	N-Acetyl-S-[1-(3-methoxypropylamino)-1-oxopropan-2-yl]-L-cysteine	68.57%	64139861
235.0742	17.8	4.9E-04	3.31	$C_8H_{14}N_2O_4S$	29/58 (77%)	98%	**N-Acetyl-S-(2-carbamoylethyl)-L-cysteine***	84.70%	157849
244.0633	23.1	9.8E-04	2.44	$C_{10}H_{13}NO_4S$	29/42 (77%)	*NA*	*NA*	*NA*	*NA*
234.0790	24.2	1.1E-03	4.18	$C_9H_{15}NO_4S$	39/58 (89%)	100%	N-Acetyl-S-(3-oxobutyl)-L-cysteine	85.34%	185147
270.0459	20.3	1.1E-03	2.49	$C_8H_{15}NO_5S_2$	16/58 (80%)	99%	N-Acetyl-S-[2-(methylsulfonyl)ethyl)-L-cysteine	82.23%	107774368
293.0796	18.3	1.1E-03	3.13	$C_{10}H_{16}N_2O_6S$	31/59 (70%)	100%	N-Acetyl-S-(2-acetamido-2-carboxyethyl)-L-cysteine	85.71%	22572358
204.0685	20.9	1.2E-03	3.11	$C_8H_{13}NO_3S$	15/32 (89%)	*NA*	*NA*	*NA*	*NA*
270.0901	26.3	1.2E-03	3.26	$C_{11}H_{15}N_3O_3S$	27/58 (73%)	*NA*	*NA*	*NA*	*NA*
297.0897	19.1	1.2E-03	3.85	$C_{13}H_{16}N_2O_4S$	14/25 (92%)	89%	N-Acetyl-S-(4-acetamidophenyl)-L-cysteine	84.11%	71380584
190.0529	20.8	2.0E-03	2.92	$C_7H_{11}NO_3S$	20/58 (69%)	69%	N-Acetyl-S-(ethenyl)-L-cysteine	59.24%	71371885
236.0582	20.8	2.0E-03	2.98	$C_8H_{13}NO_5S$	34/59 (83%)	99%	N-Acetyl-S-(2-carboxyethyl)-L-cysteine	90.34%	108073

218.0841	24.2	2.0E-03	3.57	$C_9H_{15}NO_3S$	46/58 (96%)	98%	N-Acetyl-S-(but-2-enyl)-L-cysteine	89.89%	77021037
290.1051	35.9	2.4E-03	4.06	$C_{12}H_{19}NO_5S$	44/58 (94%)	99%	N-Acetyl-S-[[1-(2-methoxy-2-oxoethyl)cyclopropyl]methyl]-L-Cysteine	67.51%	114000172
264.0895	18.4	2.4E-03	2.35	$C_{10}H_{17}NO_5S$	49/57 (98%)	100%	N-Acetyl-S-(2-methyl-2-carboxypropyl)-L-cysteine	86.56%	79623390
264.0895	20.1	2.4E-03	2.83	$C_{10}H_{17}NO_5S$	29/35 (56%)	100%	N-Acetyl-S-(1-carboxybutyl)-L-cysteine	83.16%	83041993
204.0685	26.7	2.9E-03	4.12	$C_8H_{13}NO_3S$	32/58 (80%)	99%	N-Acetyl-S-(2-propen-1-yl)-L-cysteine	91.95%	152467
208.0634	26.7	3.0E-03	4.42	$C_7H_{13}NO_4S$	9/25 (89%)	97%	N-Acetyl-S-(2-hydroxyethyl)-L-cysteine	56.77%	108009
278.1051	35.2	3.1E-03	3.65	$C_{11}H_{19}NO_5S$	42/50 (98%)	99%	N-Acetyl-S-[1-(carboxymethyl)butyl]-L-cysteine	79.41%	*NA*
188.0372	25.0	5.2E-03	1.99	$C_7H_9NO_3S$	17/33 (88%)	*NA*	N-Acetyl-S-(ethynyl)-L-cysteine	58.29%	88654551
278.1051	22.7	6.1E-03	2.56	$C_{11}H_{19}NO_5S$	53/58 (98%)	100%	N-Acetyl-S-[1-(carboxymethyl)butyl]-L-cysteine	80.86%	*NA*
264.0895	31.3	6.8E-03	2.82	$C_{10}H_{17}NO_5S$	46/58 (97%)	100%	N-Acetyl-S-(2-methyl-2-carboxypropyl)-L-cysteine	86.10%	79623390
286.0737	31.2	1.5E-02	2.37	$C_{12}H_{15}NO_5S$	32/51 (93%)	98%	*NA*	*NA*	*NA*
235.0201	31.2	1.6E-02	2.44	$C_7H_{10}N_2O_3S_2$	15/42 (84%)	61%	*NA*	*NA*	*NA*
275.0691	22.5	1.7E-02	2.89	$C_{10}H_{14}N_2O_5S$	32/44 (95%)	100%	N-Acetyl-S-(1-methyl-2,5-dioxopyrrolidin-3-yl)-L-cysteine	71.30%	113999095
278.1051	21.9	2.3E-02	2.29	$C_{11}H_{19}NO_5S$	41/58 (90%)	100%	N-Acetyl-S-[1-(carboxymethyl)butyl]-L-cysteine	77.36%	*NA*
264.0895	30.9	3.0E-02	2.55	$C_{10}H_{17}NO_5S$	46/58 (96%)	100%	N-Acetyl-S-(2-methyl-2-carboxypropyl)-L-cysteine	86.10%	79623390
236.0947	24.3	3.9E-02	1.96	$C_9H_{17}NO_4S$	43/58 (86%)	87%	**N-Acetyl-S-(3-hydroxypropyl-1-methyl)-L-cysteine***	83.77%	46780019
278.1051	35.9	4.1E-02	2.40	$C_{11}H_{19}NO_5S$	41/58 (95%)	100%	N-Acetyl-S-[1-(carboxymethyl)butyl]-L-cysteine	77.10%	*NA*
298.0771	28.4	4.3E-02	0.31	$C_{10}H_{19}NO_5S_2$	23/32 (97%)	*NA*	N-Acetyl-S-[(3-ethylsulfonyl)propyl]-L-cysteine	66.50%	65092758

*** Validated by internal reference**

3.4 Conclusion

We developed a UHPLC-MS/MS method to profile urinary mercapturic acid conjugates. In contrast to contemporary methodology, our novel approach utilizes nanoflow chromatography and positive ionization for the unbiased detection of detoxification products. Using a combination of fragmentation filtering and MDF, we detected hundreds of putative mercapturic acid signatures from the analysis of 40 urine samples collected from smoking and non-smoking study participants. Using multivariate and univariate statistical modeling, 38 of these discoveries were revealed to be significantly associated with tobacco cigarette use. Molecular annotation and chemical prediction of these results using SIRIUS and contemporary prediction software demonstrated many of these findings could putatively be annotated as mercapturic acid conjugates. Here, we propose many previously unknown detoxification products including the discovery of novel putative biomarkers of tobacco cigarette exposure.

Chapter 4: A Mass Spectral Database for Mercapturic Acid Conjugates

The following chapter introduces a mass spectral library for mercapturic acid conjugates developed to improve the annotation and characterization of these biomarkers of exposure in analytical studies. The library is comprised of mass spectra generated from authentic reference material of 20 mercapturic acid conjugates using multiple mass spectrometry platforms. This work was conducted in collaboration with the University of Minnesota Center for Metabolomics and Proteomics (CMSP) and Masonic Cancer Center Bioanalytical Services Center. This will be submitted as an original technical note or supplement to the research article to be derived from the efforts outlined in Chapter 3.

4.1 Introduction

Humans are repeatedly exposed to foreign xenobiotics through their local environment, occupation, and lifestyle choices.[209] These foreign materials represent various chemicals, including plant constituents,[18,37] cosmetics,[210,211] food and beverage additives,[212,213] pharmaceuticals,[2,214] industrial chemicals,[215,216] and environmental pollutants.[186,217] It is difficult to predict the number of foreign chemicals encountered daily, but it is estimated that humans are exposed to upwards of three million xenobiotics in their lifetime.[218] Following absorption, many of these chemicals undergo metabolic transformation to remove reactive functional groups and increase their hydrophilicity for excretion. In rare instances, toxic xenobiotic precursors or their by-products may react and modify essential biomolecules, such as DNA, RNA, or proteins.[1] Recurrent exposure to these toxic chemicals may produce adverse clinical phenotypes[184] and contribute to disease pathogenesis.[156] Molecular characterization of toxic foreign chemical entities and their physiological consequences provides valuable insights into the environmental contribution to human health.

Mercapturic acid conjugates represent one detoxification product commonly used to characterize chemical exposure. Mercapturic acid conjugates are produced through glutathione-mediated conjugation of reactive electrophiles. These intermediate products are further transformed into an N-Acetyl-L-Cysteine S-conjugate form that increases water solubility and are ultimately disposed of through urinary excretion.[191] Analytical detection and quantitation of urinary mercapturic acid conjugates using LC-MS/MS has been the method of choice for characterizing the molecular outcomes of many exposure types.[19,22,31,219] Most of these methods represent targeted methodology, characterizing one or more defined mercapturic acid conjugates associated with a known chemical

exposure. For example, acrolein exposure is monitored through 3HPMA formation,[220] acrylonitrile through 2CyEMA formation,[221] and acrylamide through 2CaEMA formation.[222] Applications of an untargeted methodology for characterizing global profiles of urinary mercapturic acid conjugates are sparse. For many years, these methods depended on unit mass resolution CNL methods that monitored for a neutral loss of N-acetylcysteine (-129 Da), a characteristic of mercapturic acid conjugate compounds.[30,52] While capable of detecting unknown and undescribed mercapturic acid conjugates in complex matrices, the limited mass accuracy of triple quadrupole-based LC-MS/MS platforms limited the annotation of these discoveries.[30] Methodologies that utilize HRMS are being actively developed and have been demonstrated to characterize vast arrays of mercapturic acid conjugates associated with common exposure types.[31,53]

In Chapter 3 of this thesis, a novel HRMS method for characterizing urinary mercapturic acid conjugates was proposed. In contrast to existing techniques, this method utilized nanoflow chromatography to maximize instrument sensitivity and positive ion mode detection to obtain richer fragmentation of unknown mercapturic acid conjugates. The proposed method was demonstrated to detect previously described mercapturic acid conjugates associated with tobacco cigarette exposure and discovered many previously undescribed products. However, we could not provide confident identifications of these novel discoveries. With limited reference libraries, we were forced to rely on accurate mass assignments and computationally predicted spectral annotations. The lack of a mass spectral library represents a deficiency in this field of study and a significant obstacle in applying any discovery-based methodology.

Entries for mercapturic acid conjugates in existing mass spectral libraries are rare, despite decades of research. Chemical entries are available for mercapturic acid conjugates in the Human Metabolome Database (HMDB)[223], but almost none possess

spectral information. This issue is observed for the MassBank of North America (MoNA - https://mona.fiehnlab.ucdavis.edu/) and the Global Natural Product Social Molecular Networking (GNPS) repositories as well.[224] The NIST mass spectral library (https://chemdata.nist.gov/) has only eight tandem MS entries for mercapturic acid conjugates in its 2023 release. Without suitable reference spectra to evaluate discovery results against, the development of global profiling techniques will be impeded, and the greater toxicology community will be unable to appreciate the true impact of these emerging methods.

This chapter outlines the creation and annotation of a high-resolution tandem mass spectrometry library of mercapturic acid conjugates. To ensure the generalizability of this library, we collected results using both positive and negative ion mode detection, CID and HCD activation using multiple collision energies. We acquired both MS^2 and MS^3 spectra of fragment ions carrying the adducted chemical moiety. Each mass spectral library entry represents a composite spectrum generated from dozens of individual measurements to minimize noise and unstable signal influence. We attempted to annotate all observed ions with their corresponding chemical formula. Following result annotation and curation, we evaluated common patterns among the mercapturic acid conjugate spectra to detect diagnostic fragmentation patterns that could assist in future discovery experiments. Finally, we evaluated the detection of mercapturic acid conjugates in a human urine matrix with a DIA-based LC-MS/MS approach with our spectral results serving as an analytical reference.

4.2 Experimental Section

4.2.1 Chemicals and Reagents

A collection of 20 mercapturic acid conjugate standards, including N-Acetyl-S-(methyl)-L-cysteine (MMA) **(1)**, N-Acetyl-S-(ethyl)-L-cysteine (EMA) **(2)**, N-Acetyl-S-(2-propen-1-yl)-L-cysteine (2PeMA) **(3)**, N-Acetyl-S-(n-propyl)-L-cysteine (1PMA) **(4)**, N-Acetyl-S-(2- hydroxyethyl)-L-cysteine (2HEMA) **(5)**, N-Acetyl-S-(2-cyanoethyl)-L-cysteine (2CyEMA) **(6)**, N-Acetyl-S-(2-hydroxypropyl)-L-cysteine (2HPMA) **(7)**, N-Acetyl-S-(3-hydroxypropyl)-L-cysteine (3HPMA) **(8)**, N-Acetyl-S-(2-carboxyethyl)-L-cysteine (2CoEMA) **(9)**, N-Acetyl-S-(3-hydroxy-1-methyl- propyl)-L-cysteine (3HMPMA) **(10)**, N-Acetyl-S-(2,3-dihydroxypropyl)-L-cysteine (23HPMA) **(11)**, N-Acetyl-S-(2-carbamoylethyl)-L-cysteine (2CaEMA) **(12)**, N-Acetyl-S-(phenyl)-L-cysteine (PhMA) **(13)**, N-Acetyl-S-(2-carboxy-1-methylethyl)-L-cysteine (2CoMEMA) **(14)**, N-Acetyl-S-(2-carboxypropyl)-L-cysteine (2CoPMA) **(15)**, N-Acetyl-S-(3,4-dihydroxybutyl)-L-cysteine (34HBMA) **(16)**, N-Acetyl-S-(benzyl)-L-cysteine (BzMA) **(17)**, N-Acetyl-S-(4-nitrophenyl)-L-cysteine (4NPhMA) **(18)**, N-Acetyl-S-[5-(acetylamino)-2-hydroxyphenyl]-L-cysteine (5AcAHPhMA) **(19)**, and N-Acetyl-S-(2,4-dinitrophenyl)-L-cysteine (24NPhMA) **(20)** were purchased from Toronto Research Chemicals (Toronto, ON, CA). UHPLC-grade acetonitrile and LC-MS grade formic acid were purchased from Fisher Scientific (Waltham, MA, USA).

4.2.2 Mass Spectral Library Generation

All analytical standards were solubilized in MeOH to achieve 10 mg/mL concentrations. Prior to analytical characterization, the solutions were diluted to a final concentration of 50 pmol/µL in a solution of 50:50 water and acetonitrile with 0.1% formic

acid. A blank reference was analyzed for each monitored mass to assist in noise peak categorization. Analytical characterization and detection were performed on an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Scientific, Waltham, MA). The analysis was operated at a constant flow rate of 3 µL/min using a 50 µL syringe and an integrated syringe pump. A heated electrospray ionization (HESI) source was used with a source voltage of 3.5 kV and 2.5 kV for positive and negative ion detection, respectively. The ion transfer tube temperature was 275 °C.

Mass spectral library spectra were collected using full-scan detection followed by targeted multistage activation MS^n acquisition on each reference standard independently. All data collection was performed using Orbitrap detection at a resolution of 120,000 in both positive and negative ion modes. Full-scan detection was performed over a mass range of 70 – 300 *m/z* using an AGC targeted setting of 4×10^5, a maximum ion injection time of 50 ms, and an S-Lens RF setting of 30%. Targeted MS^2 collection was performed using each reference standard's charged monoisotopic mass with a quadrupole isolation of 1.0 *m/z*, an AGC setting of 5.0×10^4, and a maximum injection time of 250 ms. Multistage activation MS^3 collection was performed if a fragment ion produced from the neutral loss of 131.0582 Da ($C_5H_9NO_3$) or 105.0426 Da ($C_3H_7NO_3$) in positive ion mode detection or 129.0426 Da ($C_5H_7NO_3$) in negative ion mode was detected for the respective reference standard. Targeted MS^3 acquisition was performed using a full scan quadrupole isolation of 1.0 *m/z* and MS^2 quadrupole isolation of 2.0 *m/z*, an AGC setting of 5×10^4, and a maximum injection time of 250 ms. For both MS^2 and MS^3 acquisitions, both CID and HCD activation spectra were collected using a normalized fixed collision energy of 0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 75%, and 90%. For MS^3 acquisition, the initial MS^2 fragmentation was performed using HCD activation at 30% or 50% fixed collision energy dependent on which energy produced more intense

detection for the fragment ions of interest. Data collection was performed continuously for 5 min using the following settings to acquire approximately 20 fragmentation spectra replicates at each activation-collision energy setting, in positive and negative ion modes, respectively.

4.2.3 UHPLC LC-MS/MS Analysis

Analytical characterization of mercapturic acid conjugate references in human urine matrices utilized identical UHPLC methods as described in Chapter 3. Analytical separation and detection were performed on an UltiMate 3000 RSLCnano UHPLC system (Thermo Scientific, Waltham, MA) interfaced to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA). All dried urine extracts were reconstituted using a load solvent mixture of 0.1% aqueous formic acid. Samples were spiked with a mixture Pierce™ Peptide Retention Time Calibration Mixture (Rockford, IL, USA) at a final 100 fmol/µL concentration to monitor chromatographic and retention time performance throughout the experiment. A 1% sample equivalent load in 1 µL was injected on the analytical platform equipped with a 10 µL injection loop. Chromatographic separation was performed using a self-packed C18 column (Dr. Maisch GmbH ReproSil-PUR 1.9 µm 120 Å C18aq, 100 µm ID x 45 cm length) maintained at 55 °C for the duration of the experiment. The mobile phases were (A) 0.1% aqueous formic acid and (B) 0.1% formic acid in ACN solutions. Chromatographic separation was performed using a linear gradient starting at 0% B and increased to 50% B at 40 min, 90% B at 60 min, and held for 8 min followed by a return to starting conditions. The analysis was operated at a constant flow rate of 325 nL/min. A Nanospray Flex ion source (Thermo Fisher Scientific) was used with a source voltage of 2.1 kV and an ion transfer tube temperature of 250 °C.

Discovery analyses of experimental samples were performed using full-scan detection followed by data-independent MS^2 acquisition. All analyses were performed in positive ion mode. Full-scan detection was performed using Orbitrap detection at a resolution of 120,000, AGC targeted setting of 4×10^5, a maximum ion injection time of 50 ms, and an S-Lens RF setting of 40%. Scan ranges of 170 *m/z* – 600 *m/z* were used for full-scan detection. Fragmentation spectra were collected using a DIA design with 20.0 *m/z* staggered-isolation window schema. Fragment spectra was performed using Orbitrap detection at a resolution of 30,000, normalized AGC target setting of 4×10^5 with a 60 ms maximum ion injection time, HCD activation with a 30% fixed collision energy.

4.2.4 Mass Spectral Library Assembly

A mass spectral library of mercapturic acid conjugates was compiled from consensus spectra computed from each reference standard for each activation type and collision energy in positive and negative ion mode, respectively. All library operations and statistics were conducted using the R programming language.[195] Raw mass spectral data was converted to mzML format using MSConvert and centroid using the continuous wavelet transformation algorithm.[174] Mass spectra formatting was conducted using the *mzR* R package.[174] Fragment ions between spectra were aligned using a mass tolerance of 10 ppm in positive ion mode and 15 ppm in negative ion mode. To compute consensus spectra, only fragment ions detected in at least 75% of the collected spectra at a minimum 0.1% base peak intensity were considered. Consensus spectra fragment ion intensity was calculated from mean abundance of all collected spectra. Noise ions detected in the blank references were excluded unless they exhibited a 10:1 signal-to-noise ratio in the standards reference relative to the blank. A NIST compatible mass spectral library entry was created for each consensus spectra that included relevant molecule and

instrumentation meta data where necessary. All fragment ions and neutral losses were annotated with assistance of the *MassTools* and *rcdk* R packages.[198,225]

4.2.5 Data Processing and Mercapturic Acid Verification

Human urine analyses generated from the DIA analysis were analyzed using the open-source software Skyline.[226] The processing document was optimized for small molecule detection. Six mercapturic acid transitions were selected from the generated mass spectral library using HCD activation at 30% normalized collision energy. Precursor and transition ion traces were detected using a resolution setting of 120,000 and 30,000, respectively. Putative mercapturic acid detection was determined from cosine dot product similarity. A minimum 70% similarity, the community standard, was selected as the discovery threshold. All discoveries were analyzed manually, and transition alignments were exported directly from Skyline.

4.3 Results and Discussion

4.3.1 Overview of Mercapturic Acid Mass Spectral Library

To build the basis of a comprehensive mercapturic acid conjugate mass spectral library, a collection of 20 commercially available compounds was analyzed using HRMS (**Figure 4.1**). These compounds represent a diverse collection of mercapturic acids associated with commonly encountered reactive electrophiles and other toxic substances. Of these compounds, only two (PhMA and BzMA) possess HRMS tandem mass spectra entries in the NIST mass spectra library. One compound, 5AcAHPhMA, possesses only an electron ionization spectrum in the NIST mass spectra library. Otherwise, there is no evidence that the spectra of the remaining compounds are curated in any open-source

mass spectral library. Additionally, none of these spectra possess multistage activation MS^n spectra of any kind.

We collected tandem MS^2 spectra and multistage activation MS^n for all the listed compounds in positive and negative ionization, using both CID and HCD activation with 14 distinct collision energies. We aimed to collect MS^n spectra on fragment ions that exhibited a diagnostic neutral loss of the N-acetylcysteine moiety in positive ion mode (-131.0582 Da), the N-acetylcysteine moiety in negative mode (-129.0426 Da), or the combined loss of the N-acetyl and carboxyl functional groups in positive mode (-105.0426 Da), respectively; however, not all the analyzed compounds exhibited these neutral losses. Cumulatively, we collected 2,548 unique spectra for the 20 compounds, of which 1,120 consisted of positive and negative MS^2 spectra and 1,428 positive and negative MS^3 spectra. These spectra were diligently curated using a parameterized algorithm to remove non-reproducible ions not repeatably detected in replicate spectra collection and noise ions detected at a notable intensity in the blank background reference. Following this curation, 203 spectra were excluded because no ions satisfied our detection criteria following filtering. All these were multistage MS^3 spectra, of which 72% were acquired in negative ion mode, with a mean collision energy of 60%. In general, we observed that fewer reliable fragment ion species were detectable for many of the compounds at high collision energies, particularly in negative mode.

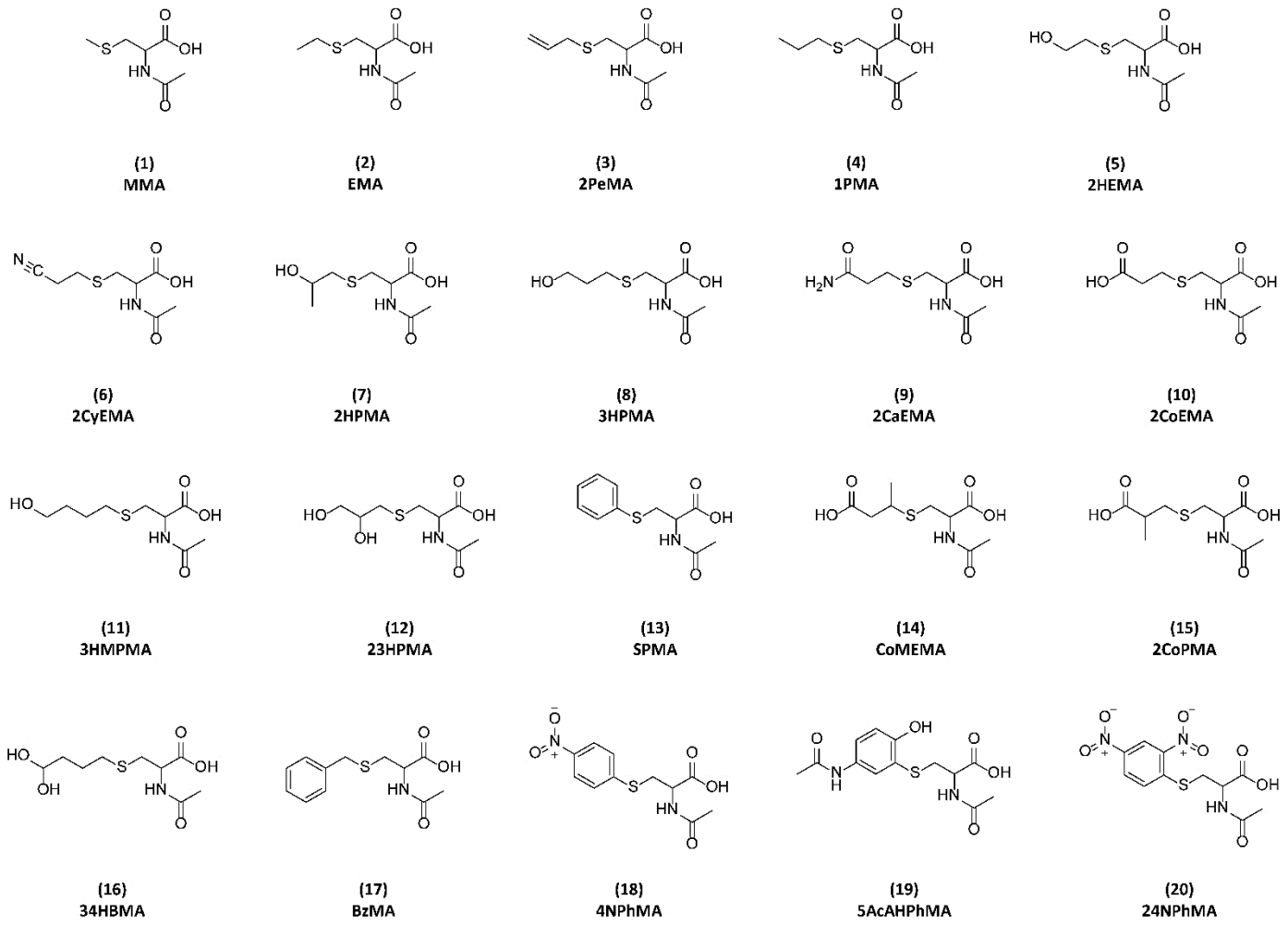


Figure 4.1 - Chemical structures and abbreviations of the mercapturic acid conjugate references.

To maximize the information content of our mass spectral library, we annotated all observed fragment ions to ensure that most detected masses were ascribed to a valid chemical formula. We aimed for 95% of the total ion intensity to be annotated following the formula assignment. The average explained intensity for the acquired MS^2 spectra was 99.45% for positive ion mode and 99.50% for negative ion mode. Generally, the observed fragments that could not be annotated were low-intensity ions measured at less than 10% of the base peak intensity. All positive mode MS^2 spectra satisfied the 95% explained intensity threshold. Two negative ion mode spectra failed to meet this threshold for 2CyEMA at 75% and 90% collision energy, respectively. These spectra were dominated by two annotated ions of m/z 154.9456 and 192.9575 that could not be assigned a chemical formula and were not removed by any noise detection filters. To test the fidelity of our spectral library, we compared the similarity of PhMA and BzMA of our spectra to that of the NIST mass spectral library (**Figure 4.2**). For our spectral results, we observed a mean cosine dot product similarity of 99.3% and 99.4% for PhMA and BzMA in positive ion mode, respectively, and 99.9% and 99.9% similarity in negative ion mode, respectively, among all 14 collision energies. Our proposed chemical annotations matched the NIST assignments precisely. We failed to annotate one ion NIST assigned at *m/z* 105.0445 for PhMA. This ion was annotated as a $C_6H_5N_2$ produced from a structural rearrangement with the N_2 collision gas that our annotation approach did not consider. These results affirm that our spectral library generation and structural assignments are accurate and can be trusted for the remaining undescribed mercapturic acid conjugates.

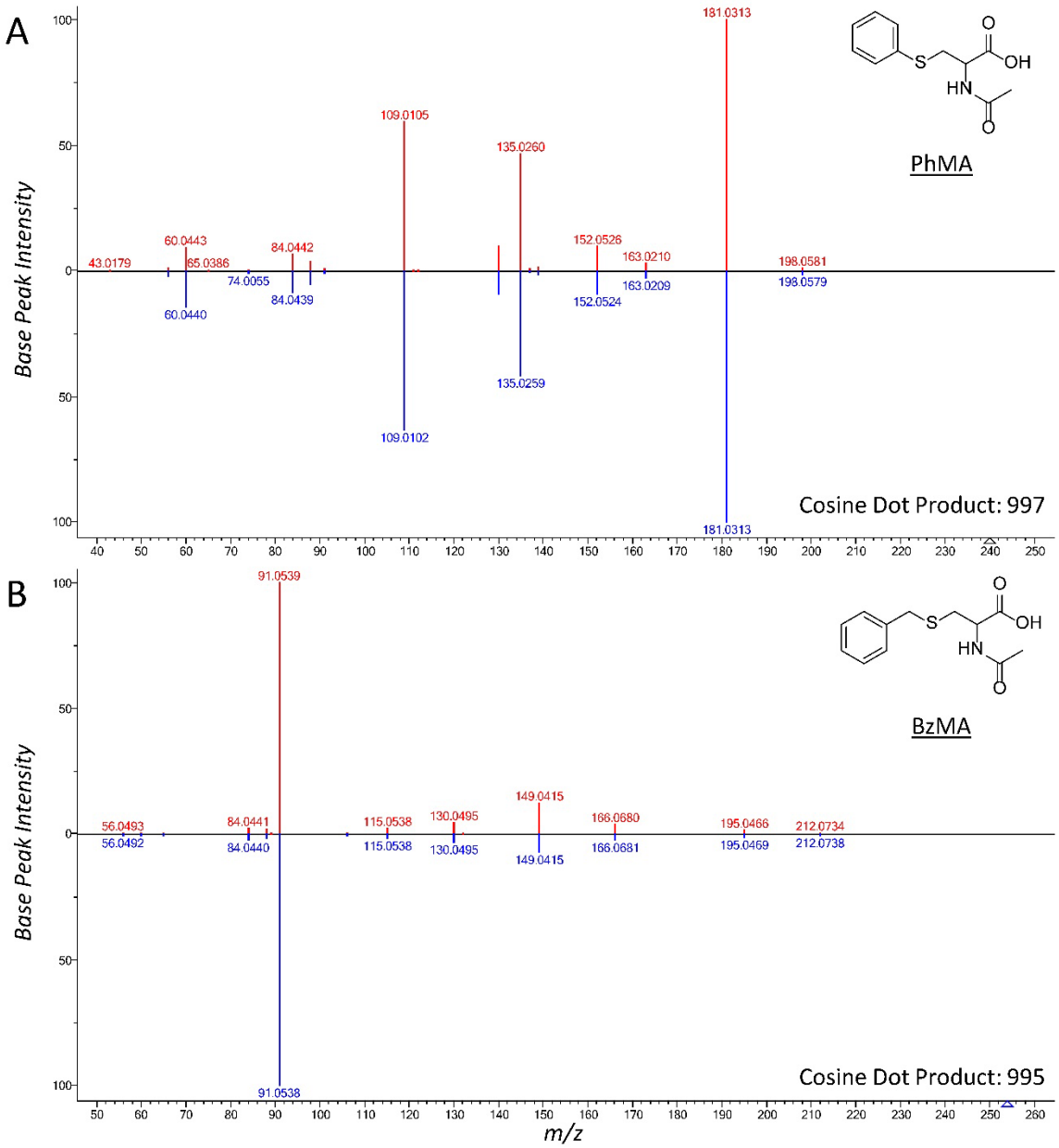


Figure 4.2 - Comparison of generated mass spectral library results (red) to NIST reference spectra (blue) using HCD activation with 30% normalized collision energy. (A) Comparison of PhMA. (B) Comparison of BzMA.

Annotation of the MS^3 spectra proved more challenging. The average explained intensity for the acquired MS^3 spectra was 90.21% for positive ion mode and 98.97% for negative ion mode. While almost all (96%) of the negative ion mode MS^3 spectra satisfied the 95% annotation threshold, only 75% of the MS^3 collected in positive ion mode passed this threshold. Evaluation of these poorly annotated results suggested the occurrence of chimera spectra, in which more than one unique ion species is isolated for fragmentation simultaneously. Using pure, authentic reference material, observing chimeric MS^2 spectra is nearly impossible. However, the fragmentation of these compounds produced a wide variety of fragment ions of varying masses. Using 2HPMA as an example (**Figure 4.3**), the diagnostic neutral loss of $C_3H_7NO_3$ produces a fragment ion at *m/z* 117.0369 (C_5H_9OS+). With an instrument default MS^n precursor isolation width of 2 Da, the neighboring ion at *m/z* 116.0161 ($C_4H_6NOS^+$) is co-isolated for multistage activation. Many of the fragment ions cannot be properly annotated as they correspond to this co-isolating ion. While the isolation width can be lowered below the instrument vendor recommended setting to minimize this phenomenon, this will not entirely remove the chance of co-isolation and will limit the transmission of low abundant ions. As negative ion mode produced fewer fragment ions overall, the occurrence of chimeric spectra was not an issue. As multistage MS^n spectra represent a small proportion of common mass spectral libraries, there are few guidelines on how best to annotate chimeric annotation. As chimeric spectra are often unavoidable, the community will need to develop clear annotation guidelines to account for these fringe cases.

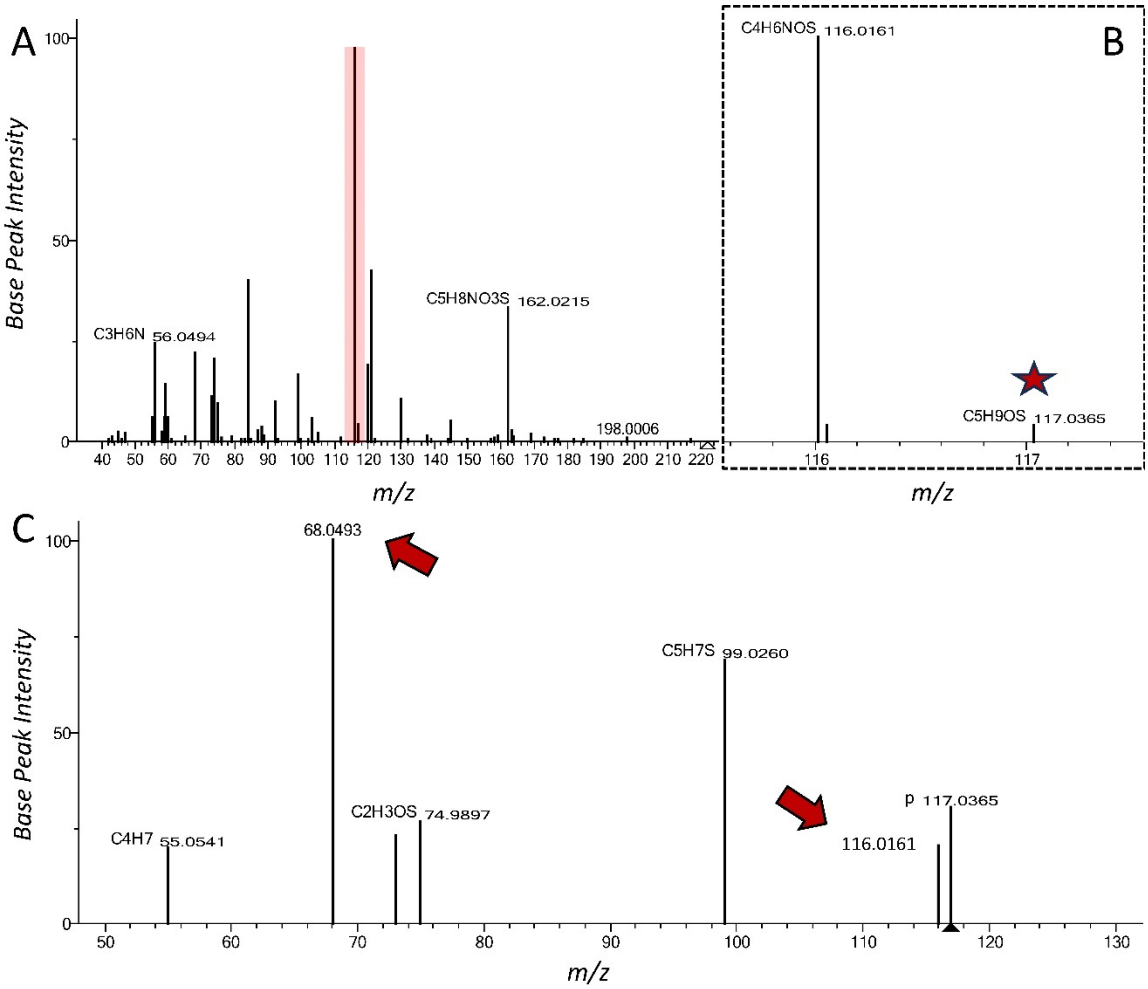


Figure 4.3 - Evaluation of positive ion mode chimeric multistage activation MS^n of 2HPMA. (A) Full MS^2 spectra of 2HPMA using HCD activation at 50% normalized collision energy. The highlighted region represents the 2 Da isolation window for the diagnostic fragment at *m/z* 117.0369 following the neutral loss of $C_3H_7NO_3$. (B) Zoom-in on the fragment ions in the isolation window. The star marks the diagnostic fragment ion. (C) The multistage activation MS^n of 2HPMA. The red arrows indicate chimeric fragments that cannot be annotated to the intended source ion.

4.3.2 Common Fragmentation Pathways of Mercapturic Acids

Using the assembled mercapturic acid mass spectral library, we evaluated common fragmentation pathways descriptive of this class of molecule. As HCD-based activation is the more common HRMS fragmentation method for small molecule discovery experiments, we elected to focus our analysis on only HCD MS^2 and MS^3 spectra in positive and negative ion mode. The most immediate observation was the stark difference in the information content between positive and negative ion mode MS^2 spectra (**Figure 4.4**). As expected, positive ion mode produces more fragmentation patterns than negative ion mode. This phenomenon improves the performance of the spectral matching algorithm and would increase the annotation rate of unknown mercapturic acids encountered during a discovery experiment.

We evaluated reproducible patterns observed between the 20 independent mercapturic acid compounds to identify common fragmentation pathways. For this assessment, we identified fragment ions or neutral losses detected in at least 80% (N = 16) of the standards in at least one collision energy using HCD activation. We then monitored the average intensity of these signals over the acquired collision energies to categorize fragmentation pathway patterns. For MS^2 spectra collected in positive ion mode, ten unique fragment ions (**Figure 4.5**) and seven unique neutral losses (**Figure 4.6**) were shared between the mercapturic acid references. These patterns exhibited a wide range of masses and intensities, supporting our empirical observations on the rich fragmentation potential of positive ion mode acquisition. We found that expected neutral losses of functional groups and structural moiety were the prominent elements of mercapturic acids in positive ion mode. Importantly, we reaffirmed that the neutral losses of $C_5H_9NO_3$ (-131.0582 Da) and $C_3H_7NO_3$ (-105.0426 Da) were diagnostic fragmentation

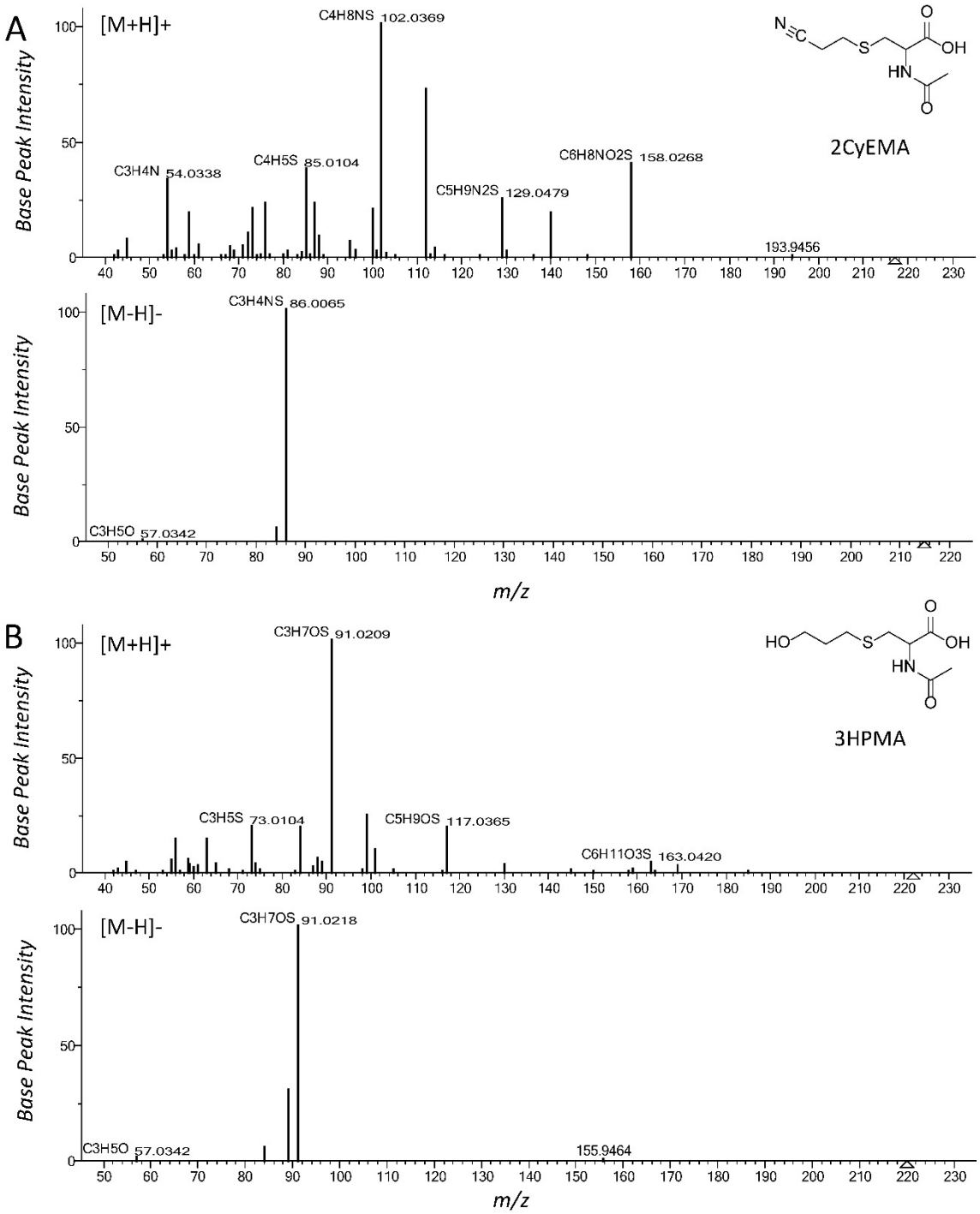


Figure 4.4 - Comparison of mercapturic acid conjugate fragmentation pathways in positive and negative ion mode. All spectra were generated using HCD activation and 50% normalized collusion energy. (A) Comparison of 2CyEMA spectra. (B) Comparison of 3HPMA spectra.

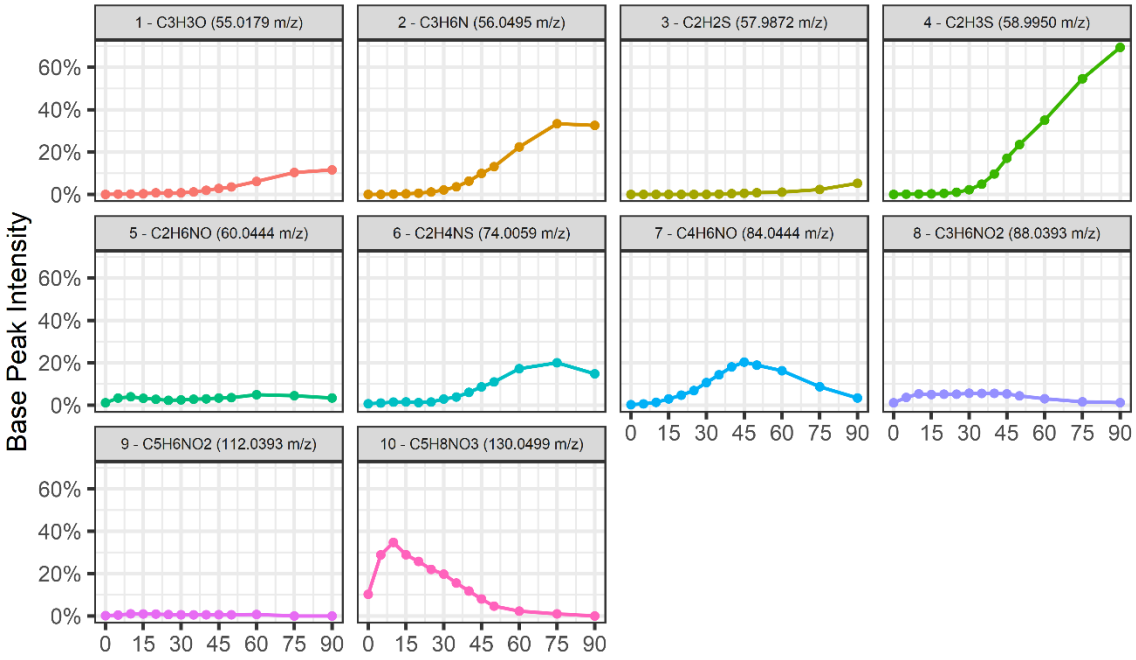


Figure 4.5 - Intensity profiles of shared fragment ions of mercapturic acid conjugates in positive ion mode over multiple collision energies. Shared fragmentation patterns are defined as detectable in at least 80% of the reference standards in at least one nominal collision energy.

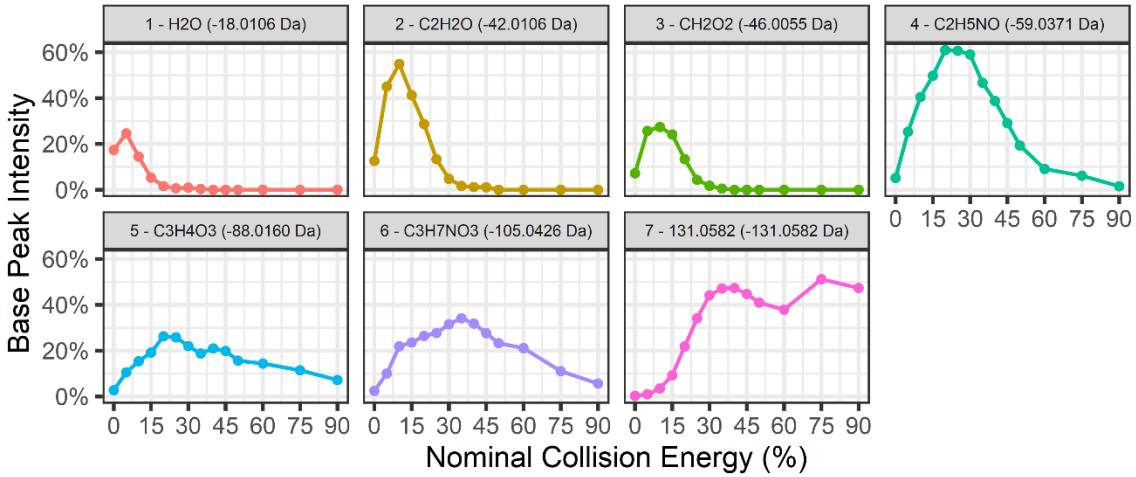


Figure 4.6 - Intensity profiles of shared neutral loss patterns of mercapturic acid conjugates in positive ion mode over multiple collision energies. Shared fragmentation patterns are defined as detectable in at least 80% of the reference standards in at least one nominal collision energy.

patterns of this compound class. Evaluation of the MS^3 spectra produced from the multistage activation of their corresponding fragment ions indicated few shared fragmentation pathways. Only one product ion was shared among these spectra over the multiple collision energies at *m/z* 44.9794 (CHS^+). As expected, the MS^3 spectra were descriptive of the adducted electrophile moiety, and there was little shared between these spectra.

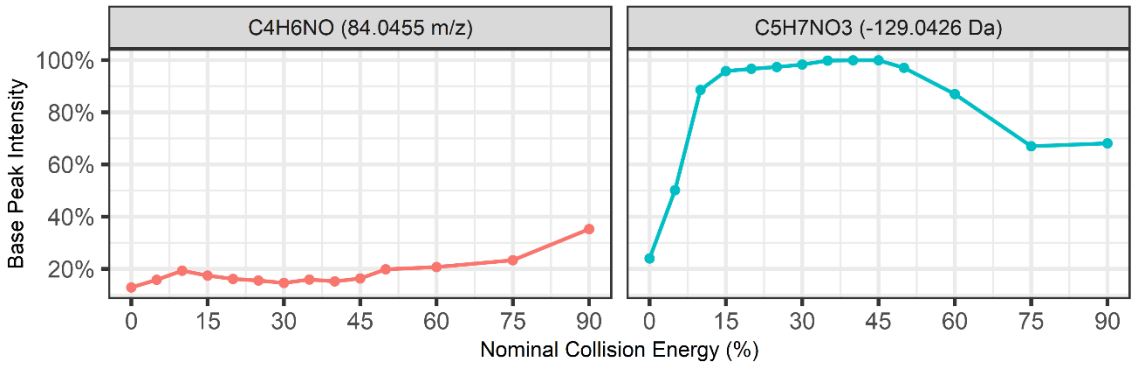


Figure 4.7 - Intensity profiles of shared fragment ion (left) and neutral loss (right) patterns of mercapturic acid conjugates in negative ion mode over multiple collision energies. Shared fragmentation patterns are defined as detectable in at least 80% of the reference standards in at least one nominal collision energy.

Evaluation of the shared fragmentation pathways in negative ion mode further highlighted the difference in information potential between positive and negative modes. We only observed one product ion at *m/z* 84.0450 ($C_4H_6NO^-$) and one neutral loss of 129.0426 Da ($C_5H_7NO_3$) shared between 80% of the mercapturic acid references at any collision energy (**Figure 4.7**). The neutral loss of the N-acetylcysteine moiety is the only neutral loss commonly reported for mercapturic acid conjugates in negative mode.[30,31,52,53] The common occurrence of a fragment ion at *m/z* 84.0455 has been previously reported;[31] however, other losses were reported for mercapturic acid conjugates in negative mode,

including fragment ions at *m/z* 162.0225 ($C_5H_8NO_3S^-$), 128.0348 ($C_5H_6NO_3^-$), and 74.0242 ($C_2H_4NO_2^-$). These patterns were only individually observed for a select few reference standards. There were no shared fragmentation pathways for the multistage activation MS^3 spectra following the neutral loss of the N-acetylcysteine moiety. In general, very few fragment ions were observed in any of the MS^3 spectra in negative ion mode. Acquiring these spectra would provide minimal additional structural insights in a discovery experiment.

As a final evaluation, we re-evaluated the optimal collision energy for the unbiased detection of mercapturic acids in positive ion mode using these new spectral insights. Using the common fragmentation profiles, we determined that the neutral loss of the N-acetylcysteine moiety (-131.0582 Da) was most formed at 40% fixed collision energy, and the joint loss of the N-acetyl and carboxyl functional groups (-105.0426 Da) occurred at 35% fixed collision energies. For an unbiased discovery analysis, detecting one of these neutral losses at a minimum of 10% base peak intensity and an instrument-specific absolute noise threshold would provide adequate specificity to ensure a novel discovery is not dependent on a potential noise peak. Examining the mercapturic acid reference spectra at a 40% fixed collision energy with HCD activation, we observed that 85% of compounds (N = 17) satisfied this detection threshold (**Figure 4.8**). While many of the compounds could be detected using only the neutral loss of N-acetylcysteine moiety, supplementing this approach with the combined N-acetyl and carboxyl neutral loss will improve the breadth of discovery.

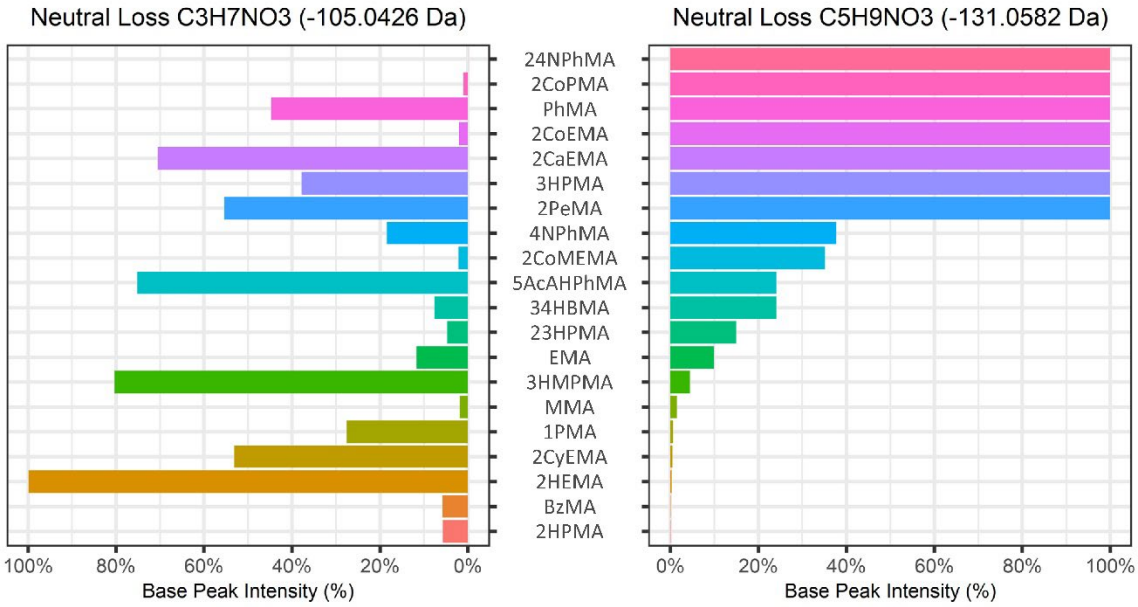


Figure 4.8 - Evaluation of diagnostic neutral loss coverage for the reference mercapturic acid conjugates in positive ion mode. All spectra collected with HCD activation with 40% normalized collision energy.

Only three compounds failed to exhibit one of these neutral losses at the desired minimum intensity: MMA, BzMA and 2HPMA. In contrast to the other standards evaluated in this study, the spectra for BzMA were dominated by a fragment produced by the neutral loss of 163.0303 Da ($C_5H_9NO_3S$) (**Figure 4.2B**). This loss represents the complete N-acetylcysteine moiety loss, including the sulfur atom. This fragmentation pattern was observed for a subset of the other mercapturic acids but was not as commonly formed as the 131.0582 Da ($C_5H_9NO_3$) loss. This pattern could be included as a diagnostic neutral loss of mercapturic acid conjugates in positive ion mode. However, more compounds will need to be analyzed to determine the prevalence of this neutral loss. Evaluation of the MMA and 2HPMA spectra indicated that this compound was dominated by the fragment ions descriptive of the N-Acetyl-L-Cysteine moiety of the conjugate structure (**Figure 4.9**). This included prominent ions at *m/z* 162.0220 ($C_5H_8NO_3S^+$), 130.0499 ($C_5H_8NO_3^+$), and

84.0444 ($C_4H_6NO^+$), the positive ion mode equivalents of the characteristic fragment ions reported for mercapturic acid conjugates in negative ion mode. While the work presented in Chapter 3 of this thesis focused on detecting mercapturic acid conjugates by neutral loss screening, subsidizing these discovery efforts with fragment ion-based detection screening could prove beneficial for these fringe cases.

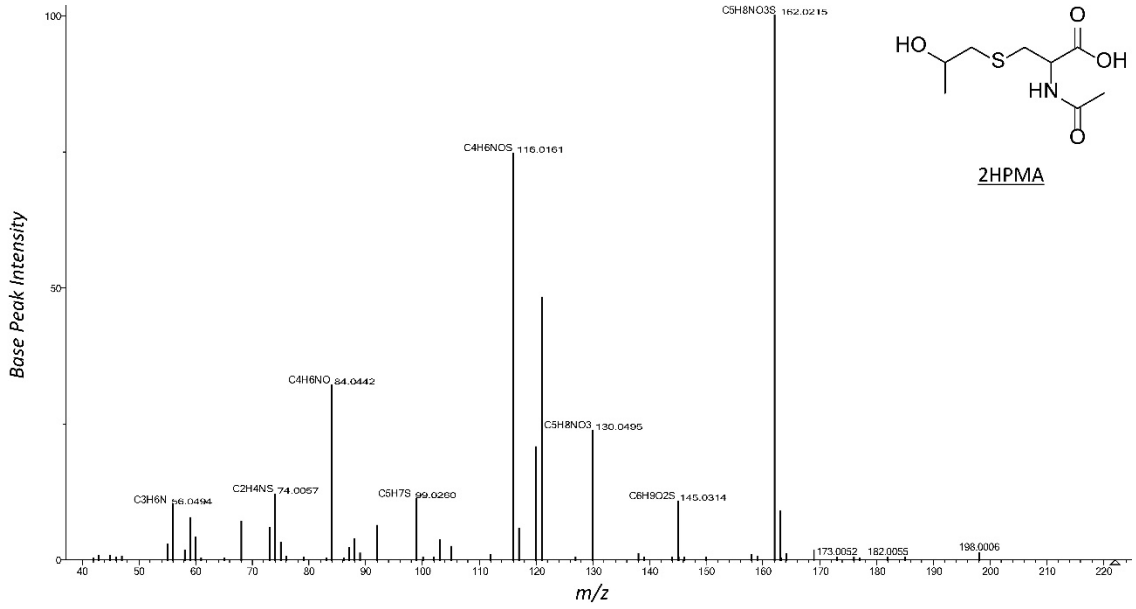


Figure 4.9 - Fragmentation spectra of 2HPMA in positive ion mode using HCD activation with 40% normalized collision energy.

4.3.3 Detection of Mercapturic Acids in Human Urine

In Chapter 3 of this thesis, we presented an LC-DDA-MS/MS method for discovering mercapturic acids using nanoflow chromatography and positive ion mode detection. While our method discovered hundreds of putative mercapturic acid signatures and a small collection of potential biomarkers of tobacco cigarette usage, we could not confidently identify most of our discoveries. We could only verify a small collection of our

discoveries using a set of heavy-labeled mercapturic acids standards of previously reported cigarette exposure biomarkers. To evaluate the longitudinal performance of our LC-DDA-MS/MS experiment, we analyzed pooled quality control samples periodically throughout the experiment. These samples were analyzed using a DIA-based acquisition approach to subsidize the discovery efforts *post hoc* if we obtained new spectral information to corroborate any of our discoveries. Here, we aimed to evaluate if any of the mercapturic acid conjugates compiled in our mass spectral library could be detected in a human urine matrix.

The ten pooled quality control samples were analyzed using the open-source software Skyline and a DIA-based workflow. We queried these results using spectra for all 20 mercapturic acid standards in positive ion mode generated using HCD activation with 30% fixed collision energy. We monitored the alignment of the extracted ion chromatogram of the precursor ion trace with its respective fragment ion patterns. We verified the detection of 11 of 20 mercapturic acid references among pooled quality control samples with a minimum cosine dot product of 700, i.e., 70% similarity (**Figure 4.10**). Five of these identifications are supported by the co-elution of the heavy-labeled mercapturic acid spike-in analyzed in the experiment, namely, 2CyEMA, 2HPMA, 3HPMA, 2CaEMA, and 3HMPMA. The additional six spectral library-driven identifications support the capacity of our LC-MS/MS workflow to perform an unbiased characterization of mercapturic acids in positive ion mode. Using these results, we supported the identification of 2CoEMA as a putative biomarker of cigarette usage. Strangely, while a putative mercapturic acid signature at *m/z* 204.0689 was detected as a biomarker of cigarette usage, this compound was not identified as 2PeMA. This unknown molecule eluted three minutes before the 2PeMA chromatographic peak, and its fragments did not align with the reference mass spectra for this compound. 2PeMA was the top candidate for this molecule following

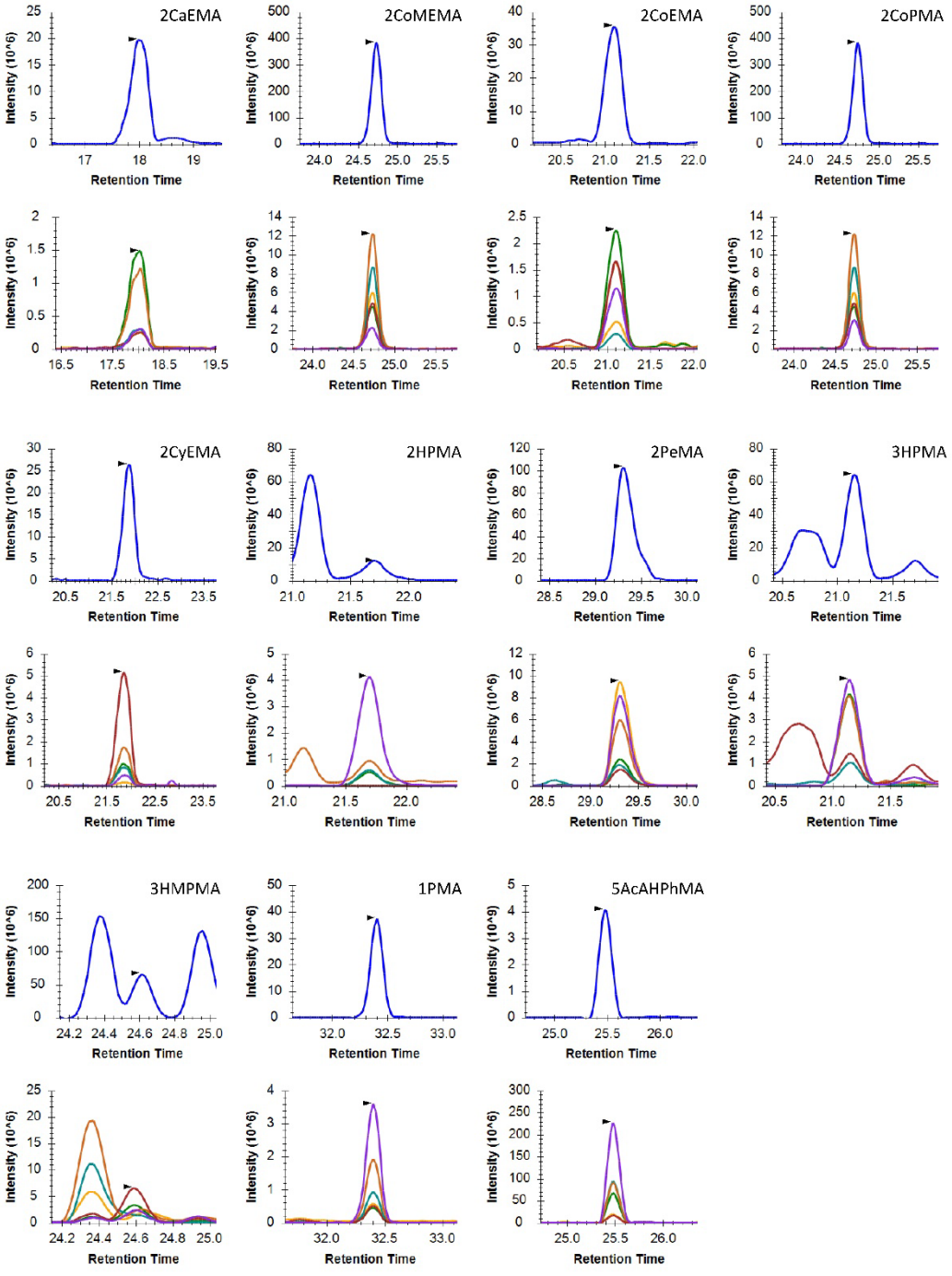


Figure 4.10 - DIA analysis of precursor and fragment ion traces in positive ion mode of detected mercapturic acid conjugates in human urine. The top trace is the precursor ion trace, ie. Extracted ion chromatogram. The bottom traces are the corresponding fragment ions from the generated mass spectral library.

SIRIUS structural prediction. However, other mercapturic acid structures were computed, including N-Acetyl-S-[(E/Z)-propen-1-yl)]-L-Cysteine and N-Acetyl-S-cyclopropyl-L-Cysteine. This observation shows the limitation of structural prediction-based annotation, and all assignments, regardless of confidence score and substructure coverage, represent equally plausible candidate identities.

4.4 Conclusion

We generated and curated a mass spectral library of 20 mercapturic acid conjugates using positive and negative ion modes, with HCD and CID activation, over multiple normalized collision energies. This library consisted of over 2,000 independent mass spectra and included both tandem MS^2 and multistage activation MS^n spectra. To ensure the integrity of this library, we performed a robust chemical formula annotation of all detected fragment ions. All acquired MS^2 spectra could explain 95% of the total ion intensity in positive and negative ion modes. We were able to explain the majority of ion intensity in the acquired MS3 spectra, but the frequent occurrence of chimeric spectra limited complete annotation. The analyzed mercapturic acid conjugates shared many fragmentation patterns. There were 17 combined product ion and neutral losses shared between at least 80% of these compounds in positive ion mode and 2 in negative ion mode. We determined that 17 of the 20 analyzed compounds exhibited one of the diagnostic neutral losses proposed to monitor mercapturic acids in conjugates in positive ion mode, and there are many other supplementary patterns available to detect fringe non-conforming molecules. As a final examination, we determined which mercapturic acids were detectable in human urine using our proposed LC-MS/MS monitoring approach. We identified 11 of the 20 standards in our global profiling experiment and corroborated that 2CoEMA is a putative biomarker of tobacco cigarette use. This mass spectral library represents only the beginning of our efforts to improve resources available to the field of exposomics and systems toxicology. We aim to continue growing this library and develop new methods to assist in annotating unknown mercapturic acid signatures. We plan to submit this library to NIST and open-source mass spectral library repositories.

Chapter 5: Conclusion and Future Directions

With a paradigm shift towards a mechanistic, pathway-oriented understanding of chemical exposure, new methods for toxicology profiling are being developed rapidly. Nowhere are these expansions more pronounced than in the development of unbiased, global profiling methodology. To better understand the molecular mechanisms of exposure and develop sophisticated computational models, the toxicology community aims to generate universal datasets representative of all xenobiotics and their by-products in a biological system. While there are no all-encompassing systems toxicology profiling methods, numerous analytical approaches have been developed to characterize discreet classes of molecular biomarkers of exposure. Critical advancements in LC-MS/MS analytical technology have enabled the simultaneous characterization of hundreds of covalently modified DNA, RNA, and protein species in a single experiment. However, rapid developments in experimental technology have revealed new obstacles impeding the application of system toxicology approaches to emerging problems. The computational workflows to support these emerging methods have yet to be developed. Many researchers are generating experimental results that need more resources to get processed in a timely and reproducible manner. The simultaneous development of new computational workflows and software is paramount to the continued growth of these branches of systems biology. The work presented in this thesis aims to address these limitations and provide much-needed computational support to researchers throughout the toxicology sciences.

In Chapter 1, multiple contemporary LC-MS/MS methods for the unbiased discovery of molecular biomarkers of exposure are reviewed. These methods capitalize on the shared structural similarity between discreet compound classes to facilitate global

discovery methods in complex matrices. While each reviewed method has distinct strengths and limitations, hundreds of toxicologically relevant compounds may be discovered in a singular experiment. A critical insight from this review was the lack of computational resources for processing and interpreting these structure-oriented LC-MS/MS methods. Many research applications depended on manual interpretation or utilized private, in-house software solutions. The few open-source software available were discordant and offered limited integration with emerging chemical annotation software. These observations highlighted the pressing need for new computational resources in this field of analytical science.

Chapter 2 of this thesis presents a novel MZmine module to process LC-MS/MS results by fragmentation filtering. This module enables users to input masses of diagnostic neutral loss, fragment ions, or both to detect characteristic patterns of a discreet class of chemical compounds. As part of a complete computational workflow, the DFBuilder module may be used to distinguish prospective chemical ions of interest, detect and integrate their corresponding chromatographic peaks, and perform retention alignment between multiple runs. Users of the MZmine software ecosystem can access various spectral library searching algorithms, molecular networking functionality, and other emerging chemical annotation tools. The capacity of these tools to detect covalently modified DNA nucleosides is presented, and a novel colibactin-derived DNA adduct is discovered. Continued application of this workflow has demonstrated its strengths but also revealed some weaknesses. MZmine is a popular metabolomics software, but optimizing processing parameters requires time and experience. In the future, robust documentation and tutorials will need to be developed to increase the accessibility of this computational workflow to a general toxicology audience.

Chapter 3 of this thesis extends the application of the DFBuilder module and structure-oriented LC-MS/MS profiling to a new class of molecule. Mercapturic acid conjugates represent an informative and accessible biomarker of exposure commonly detected in urine matrices. Historically, this compound class has been characterized by low-resolution targeted approaches or CNL-based methods that offer limited capacity to annotate unexpected mercapturic acid signatures. A novel HRMS LC-MS/MS method using nanoflow chromatography and positive ion mode detection was developed to expand the characterization efforts for this class of compound. This work demonstrates the potential of positive ion mode to unbiasedly detect known and undescribed mercapturic acids associated with tobacco cigarette exposure. As a complex chemical matrix, additional filtering stringency is applied to minimize the risk of false detection. In addition to elemental formula restrictions, chemical structural annotation software is explored for classifying compounds an N-acyl alpha amino acids and predicting candidate mercapturic acid structures. While this work exhibited solid results for detecting mercapturic acids, there were key limitations. The available resources for identifying and verifying mercapturic acid signatures need to be improved. Most of our results could not be identified to existing commercial and open-source reference libraries. Additionally, evaluating our results against contemporary methods using positive ion mode is difficult as no published reports were immediately available for review.

Chapter 4 of this thesis aimed to expand the available resources for identifying mercapturic acid conjugates. Twenty authentic mercapturic acid conjugate standards were obtained, and a mass spectral library was constructed. To maximize the information content of this library and ensure its generalizability, the results were generated with positive and negative ion polarity using multiple fragmentation strategies. In parallel with the tandem MS^2 spectra library, a multistage activation MS^n library was generated to

expand prospective structural annotation efforts on unknown conjugate structures. In total, 2,458 spectra were annotated and collected into the final mass spectral library. Using the collected spectra, multiple insights regarding reproducible fragmentation patterns could improve future discovery efforts. This library highlighted the discrepancy in fragmentation pathways between positive and negative ion modes for mercapturic acids. Analysis of reproducible fragmentation patterns indicated ten reproducible fragment ions and eight reproducible neutral losses in positive ion mode. At the same time, in negative ion mode, there was only one fragment ion and neutral loss, respectively. These results were used to verify the detection of 11 mercapturic acid conjugates in a human urine matrix analyzed using a DIA variant of our approach presented in Chapter 3. This work represents the basis for improving the resources available to analytical researchers in the toxicology sciences. However, considerable developments will be required to ensure the community's needs are adequately met. Hundreds of mercapturic acid conjugates are reported in the PubChem repository, and almost none possess meaningful background summaries or spectral results. Enhanced efforts will be required to grow and expand existing chemical libraries.

In conclusion, this thesis aims to expand the computational workflows and resources available for the unbiased detection of molecular biomarkers of exposure. Developing new computational solutions enhances the application of systems toxicology methods, and improved experimental insights can be obtained. Continuing resources and innovations will be required to ensure that the computational needs of future research efforts are met.

Bibliography

(1) Sturla, S. J.; Boobis, A. R.; FitzGerald, R. E.; Hoeng, J.; Kavlock, R. J.; Schirmer, K.; Whelan, M.; Wilks, M. F.; Peitsch, M. C. Systems Toxicology: From Basic Research to Risk Assessment. *Chem. Res. Toxicol.* **2014**, *27* (3), 314–329. https://doi.org/10.1021/tx400410s.

(2) Pristner, M.; Warth, B. Drug–Exposome Interactions: The Next Frontier in Precision Medicine. *Trends Pharmacol. Sci.* **2020**, *41* (12), 994–1005. https://doi.org/10.1016/j.tips.2020.09.012.

(3) Hoffman, J. M.; Soltow, Q. A.; Li, S.; Sidik, A.; Jones, D. P.; Promislow, D. E. L. Effects of Age, Sex, and Genotype on High-Sensitivity Metabolomic Profiles in the Fruit Fly, Drosophila Melanogaster. *Aging Cell* **2014**, *13* (4), 596–604. https://doi.org/10.1111/acel.12215.

(4) Dator, R.; Villalta, P. W.; Thomson, N.; Jensen, J.; Hatsukami, D. K.; Stepanov, I.; Warth, B.; Balbo, S. Metabolomics Profiles of Smokers from Two Ethnic Groups with Differing Lung Cancer Risk. *Chem. Res. Toxicol.* **2020**, *33* (8), 2087–2098. https://doi.org/10.1021/acs.chemrestox.0c00064.

(5) *Toxicity Testing in the 21st Century: A Vision and a Strategy*; National Academies Press: Washington, D.C., 2007. https://doi.org/10.17226/11970.

(6) Hartung, T.; FitzGerald, R. E.; Jennings, P.; Mirams, G. R.; Peitsch, M. C.; Rostami-Hodjegan, A.; Shah, I.; Wilks, M. F.; Sturla, S. J. Systems Toxicology: Real World Applications and Opportunities. *Chem. Res. Toxicol.* **2017**, *30* (4), 870–882. https://doi.org/10.1021/acs.chemrestox.7b00003.

(7) Folger, O.; Jerby, L.; Frezza, C.; Gottlieb, E.; Ruppin, E.; Shlomi, T. Predicting Selective Drug Targets in Cancer through Metabolic Networks. *Mol. Syst. Biol.* **2011**, *7* (1), 501. https://doi.org/10.1038/msb.2011.35.

(8) Blais, E. M.; Rawls, K. D.; Dougherty, B. V.; Li, Z. I.; Kolling, G. L.; Ye, P.; Wallqvist, A.; Papin, J. A. Reconciled Rat and Human Metabolic Networks for Comparative Toxicogenomics and Biomarker Predictions. *Nat. Commun.* **2017**, *8* (1), 14250. https://doi.org/10.1038/ncomms14250.

(9) Haswell, L. E.; Baxter, A.; Banerjee, A.; Verrastro, I.; Mushonganono, J.; Adamson, J.; Thorne, D.; Gaça, M.; Minet, E. Reduced Biological Effect of E-Cigarette Aerosol Compared to Cigarette Smoke Evaluated in Vitro Using Normalized Nicotine Dose and RNA-Seq-Based Toxicogenomics. *Sci. Rep.* **2017**, *7* (1), 888. https://doi.org/10.1038/s41598-017-00852-y.

(10) Husari, A.; Shihadeh, A.; Talih, S.; Hashem, Y.; El Sabban, M.; Zaatari, G. Acute Exposure to Electronic and Combustible Cigarette Aerosols: Effects in an Animal Model and in Human Alveolar Cells. *Nicotine Tob. Res.* **2016**, *18* (5), 613–619. https://doi.org/10.1093/ntr/ntv169.

(11) Azzopardi, D.; Patel, K.; Jaunky, T.; Santopietro, S.; Camacho, O. M.; McAughey, J.; Gaça, M. Electronic Cigarette Aerosol Induces Significantly Less Cytotoxicity than Tobacco Smoke. *Toxicol. Mech. Methods* **2016**, *26* (6), 477–491. https://doi.org/10.1080/15376516.2016.1217112.

(12) Wang, M.; Yang, S.; Cai, J.; Yan, R.; Meng, L.; Long, M.; Zhang, Y. Proteomic Analysis Using iTRAQ Technology Reveals the Toxic Effects of Zearalenone on the Leydig Cells of Rats. *Food Chem. Toxicol.* **2020**, *141*, 111405. https://doi.org/10.1016/j.fct.2020.111405.

(13) Cimbalo, A.; Frangiamone, M.; Font, G.; Manyes, L. The Importance of Transcriptomics and Proteomics for Studying Molecular Mechanisms of Mycotoxin Exposure: A Review. *Food Chem. Toxicol.* **2022**, *169*, 113396. https://doi.org/10.1016/j.fct.2022.113396.

(14) Li, Y.; Zhang, B.; Huang, K.; He, X.; Luo, Y.; Liang, R.; Luo, H.; Shen, X. L.; Xu, W. Mitochondrial Proteomic Analysis Reveals the Molecular Mechanisms Underlying Reproductive Toxicity of Zearalenone in MLTC-1 Cells. *Toxicology* **2014**, *324*, 55–67. https://doi.org/10.1016/j.tox.2014.07.007.

(15) Gebreab, K. Y.; Eeza, M. N. H.; Bai, T.; Zuberi, Z.; Matysik, J.; O'Shea, K. E.; Alia, A.; Berry, J. P. Comparative Toxicometabolomics of Perfluorooctanoic Acid (PFOA) and next-Generation Perfluoroalkyl Substances. *Environ. Pollut.* **2020**, *265*, 114928. https://doi.org/10.1016/j.envpol.2020.114928.

(16) Beale, D. J.; Sinclair, G. M.; Shah, R.; Paten, A. M.; Kumar, A.; Long, S. M.; Vardy, S.; Jones, O. A. H. A Review of Omics-Based PFAS Exposure Studies Reveals Common Biochemical Response Pathways. *Sci. Total Environ.* **2022**, *845*, 157255. https://doi.org/10.1016/j.scitotenv.2022.157255.

(17) Zhang, P.; Carlsten, C.; Chaleckis, R.; Hanhineva, K.; Huang, M.; Isobe, T.; Koistinen, V. M.; Meister, I.; Papazian, S.; Sdougkou, K.; Xie, H.; Martin, J. W.; Rappaport, S. M.; Tsugawa, H.; Walker, D. I.; Woodruff, T. J.; Wright, R. O.; Wheelock, C. E. Defining the Scope of Exposome Studies and Research Needs from a Multidisciplinary Perspective. *Environ. Sci. Technol. Lett.* **2021**, *8* (10), 839–852. https://doi.org/10.1021/acs.estlett.1c00648.

(18) Hecht, S. S.; Carmella, S. G.; Foiles, P. G.; Murphy, S. E. Biomarkers for Human Uptake and Metabolic Activation of Tobacco-Specific Nitrosamines. *Cancer Res.* **1994**, *54* (7 Suppl), 1912s–1917s.

(19) Hecht, S. S.; Stepanov, I.; Carmella, S. G. Exposure and Metabolic Activation Biomarkers of Carcinogenic Tobacco-Specific Nitrosamines. *Acc. Chem. Res.* **2016**, *49* (1), 106–114. https://doi.org/10.1021/acs.accounts.5b00472.

(20) Dator, R. P.; Murray, K. J.; Luedtke, M. W.; Jacobs, F. C.; Kassie, F.; Nguyen, H. D.; Villalta, P. W.; Balbo, S. Identification of Formaldehyde-Induced DNA–RNA Cross-Links in the A/J Mouse Lung Tumorigenesis Model. *Chem. Res. Toxicol.* **2022**, *35* (11), 2025–2036. https://doi.org/10.1021/acs.chemrestox.2c00206.

(21) Carpenter, D. O.; Arcaro, K.; Spink, D. C. Understanding the Human Health Effects of Chemical Mixtures. *Environ. Health Perspect.* **2002**, *110* (Suppl 1), 25–42.

(22) Peterson, L. A.; Oram, M. K.; Flavin, M.; Seabloom, D.; Smith, W. E.; O'Sullivan, M. G.; Vevang, K. R.; Upadhyaya, P.; Stornetta, A.; Floeder, A. C.; Ho, Y.-Y.; Zhang, L.; Hecht, S. S.; Balbo, S.; Wiedmann, T. S. Coexposure to Inhaled Aldehydes or Carbon Dioxide Enhances the Carcinogenic Properties of the Tobacco-Specific Nitrosamine 4-Methylnitrosamino-1-(3-Pyridyl)-1-Butanone in the A/J Mouse Lung. *Chem. Res. Toxicol.* **2021**, *34* (3), 723–732. https://doi.org/10.1021/acs.chemrestox.0c00350.

(23) Balbo, S.; Turesky, R. J.; Villalta, P. W. DNA Adductomics. *Chem. Res. Toxicol.* **2014**, *27* (3), 356–366. https://doi.org/10.1021/tx4004352.

(24) Carrà, A.; Guidolin, V.; Dator, R. P.; Upadhyaya, P.; Kassie, F.; Villalta, P. W.; Balbo, S. Targeted High Resolution LC/MS3 Adductomics Method for the Characterization of Endogenous DNA Damage. *Front. Chem.* **2019**, *7*, 658. https://doi.org/10.3389/fchem.2019.00658.

(25) Guo, J.; Turesky, R. J. Emerging Technologies in Mass Spectrometry-Based DNA Adductomics. *High-Throughput* **2019**, *8* (2), 13. https://doi.org/10.3390/ht8020013.

(26) Takeshita, T.; Kanaly, R. A. In Vitro DNA/RNA Adductomics to Confirm DNA Damage Caused by Benzo[a]Pyrene in the Hep G2 Cell Line. *Front. Chem.* **2019**, *7*, 491. https://doi.org/10.3389/fchem.2019.00491.

(27) Martella, G.; Motwani, N. H.; Khan, Z.; Sousa, P. F. M.; Gorokhova, E.; Motwani, H. V. Simultaneous RNA and DNA Adductomics Using Single Data-Independent Acquisition Mass Spectrometry Analysis. *Chem. Res. Toxicol.* **2023**. https://doi.org/10.1021/acs.chemrestox.3c00041.

(28) Carlsson, H.; Rappaport, S. M.; Törnqvist, M. Protein Adductomics: Methodologies for Untargeted Screening of Adducts to Serum Albumin and Hemoglobin in Human Blood Samples. *High-Throughput* **2019**, *8* (1), 6. https://doi.org/10.3390/ht8010006.

(29) Rajczewski, A. T.; Ndreu, L.; Pujari, S. S.; Griffin, T. J.; Törnqvist, M. Å.; Karlsson, I.; Tretyakova, N. Y. Novel 4-Hydroxybenzyl Adducts in Human Hemoglobin: Structures and Mechanisms of Formation. *Chem. Res. Toxicol.* **2021**, *34* (7), 1769–1781. https://doi.org/10.1021/acs.chemrestox.1c00111.

(30) Bloch, R.; Schütze, S.-E.; Müller, E.; Röder, S.; Lehmann, I.; Brack, W.; Krauss, M. Non-Targeted Mercapturic Acid Screening in Urine Using LC-MS/MS with Matrix Effect Compensation by Postcolumn Infusion of Internal Standard (PCI-IS). *Anal. Bioanal. Chem.* **2019**, *411* (29), 7771–7781. https://doi.org/10.1007/s00216-019-02166-6.

(31) Xie, Z.; Chen, J. Y.; Gao, H.; Keith, R. J.; Bhatnagar, A.; Lorkiewicz, P.; Srivastava, S. Global Profiling of Urinary Mercapturic Acids Using Integrated Library-Guided Analysis. *Environ. Sci. Technol.* **2023**. https://doi.org/10.1021/acs.est.2c09554.

(32) Mervant, L.; Tremblay-Franco, M.; Olier, M.; Jamin, E.; Martin, J.; Trouilh, L.; Buisson, C.; Naud, N.; Maslo, C.; Héliès-Toussaint, C.; Fouché, E.; Kesse-Guyot, E.; Hercberg, S.; Galan, P.; Deschasaux-Tanguy, M.; Touvier, M.; Pierre, F. H. F.; Debrauwer, L.; Guéraud, F. Urinary Metabolome Analysis Reveals Potential Microbiota Alteration and Electrophilic Burden Induced by High Red Meat Diet: Results from the French NutriNet-Santé Cohort and an in Vivo Intervention Study in Rats. *Mol. Nutr. Food Res.* **2023**, 2200432. https://doi.org/10.1002/mnfr.202200432.

(33) Correia, M. S. P.; Lin, W.; Aria, A. J.; Jain, A.; Globisch, D. Rapid Preparation of a Large Sulfated Metabolite Library for Structure Validation in Human Samples. *Metabolites* **2020**, *10* (10), 415. https://doi.org/10.3390/metabo10100415.

(34) Correia, M. S. P.; Thapa, B.; Vujasinovic, M.; Löhr, J.-M.; Globisch, D. Investigation of the Individual Human Sulfatome in Plasma and Urine Samples Reveals an Age-Dependency. *RSC Adv.* **2021**, *11* (55), 34788–34794. https://doi.org/10.1039/D1RA05994G.

(35) Perez de Souza, L.; Alseekh, S.; Scossa, F.; Fernie, A. R. Ultra-High-Performance Liquid Chromatography High-Resolution Mass Spectrometry Variants for Metabolomics Research. *Nat. Methods* **2021**, *18* (7), 733–746. https://doi.org/10.1038/s41592-021-01116-4.

(36) Gil-Solsona, R.; Sancho, J. V.; Gassner, A.; Weyermann, C.; Hernández, F.; Delémont, O.; Bijlsma, L. Use of Ion Mobility-high Resolution Mass Spectrometry in Metabolomics Studies to Provide near MS/MS Quality Data in a Single Injection. *J. Mass Spectrom.* **2021**, *56* (5). https://doi.org/10.1002/jms.4718.

(37) Shi, X.; Yang, W.; Huang, Y.; Hou, J.; Qiu, S.; Yao, C.; Feng, Z.; Wei, W.; Wu, W.; Guo, D. Direct Screening of Malonylginsenosides from Nine Ginseng Extracts by an Untargeted Profiling Strategy Incorporating In-Source Collision-Induced Dissociation, Mass Tag, and Neutral Loss Scan on a Hybrid Linear Ion-Trap/Orbitrap Mass Spectrometer Coupled to Ultra-High Performance Liquid Chromatography. *J. Chromatogr. A* **2018**, *1571*, 213–222. https://doi.org/10.1016/j.chroma.2018.08.026.

(38) Goldansaz, S. A.; Guo, A. C.; Sajed, T.; Steele, M. A.; Plastow, G. S.; Wishart, D. S. Livestock Metabolomics and the Livestock Metabolome: A Systematic Review. *PLoS ONE* **2017**, *12* (5), e0177675. https://doi.org/10.1371/journal.pone.0177675.

(39) Perchalski, R. J.; Yost, R. A.; Wilder, B. J. Structural Elucidation of Drug Metabolites by Triple-Quadrupole Mass Spectrometry. *Anal. Chem.* **1982**, *54* (9), 1466–1471. https://doi.org/10.1021/ac00246a006.

(40) Ballard, K. D.; Raftery, M. J.; Jaeschke, H.; Gaskell, S. J. Multiple Scan Modes in the Hybrid Tandem Mass Spectrometric Screening and Characterization of the Glutathione Conjugate of 2-Furamide. *J. Am. Soc. Mass Spectrom.* **1991**, *2* (1), 55–68. https://doi.org/10.1016/1044-0305(91)80061-B.

(41) Houjou, T.; Yamatani, K.; Nakanishi, H.; Imagawa, M.; Shimizu, T.; Taguchi, R. Rapid and Selective Identification of Molecular Species in Phosphatidylcholine and Sphingomyelin by Conditional Neutral Loss Scanning and MS3. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 3123–3130. https://doi.org/10.1002/rcm.1737.

(42) Scholz, K.; Dekant, W.; Völkel, W.; Pähler, A. Rapid Detection and Identification of *N* -Acetyl- *L* -Cysteine Thioethers Using Constant Neutral Loss and Theoretical Multiple Reaction Monitoring Combined with Enhanced Product-Ion Scans on a Linear Ion Trap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 1976–1984. https://doi.org/10.1016/j.jasms.2005.08.003.

(43) Lafaye, A.; Junot, C.; Ramounet-Le Gall, B.; Fritsch, P.; Ezan, E.; Tabet, J.-C. Profiling of Sulfoconjugates in Urine by Using Precursor Ion and Neutral Loss Scans in Tandem Mass Spectrometry. Application to the Investigation of Heavy Metal Toxicity in Rats. *J. Mass Spectrom.* **2004**, *39* (6), 655–664. https://doi.org/10.1002/jms.635.

(44) Chen, Y.-C.; Wu, H.-Y.; Chang, C.-W.; Liao, P.-C. Post-Deconvolution MS/MS Spectra Extraction with Data-Independent Acquisition for Comprehensive Profiling of Urinary Glucuronide-Conjugated Metabolome. *Anal. Chem.* **2022**, *94* (6), 2740–2748. https://doi.org/10.1021/acs.analchem.1c03557.

(45) Zhu, X.; Kalyanaraman, N.; Subramanian, R. Enhanced Screening of Glutathione-Trapped Reactive Metabolites by In-Source Collision-Induced Dissociation and Extraction

of Product Ion Using UHPLC-High Resolution Mass Spectrometry. *Anal. Chem.* **2011**, *83* (24), 9516–9523. https://doi.org/10.1021/ac202280f.

(46) LeBlanc, A.; Sleno, L. Atrazine Metabolite Screening in Human Microsomes: Detection of Novel Reactive Metabolites and Glutathione Adducts by LC-MS. *Chem. Res. Toxicol.* **2011**, *24* (3), 329–339. https://doi.org/10.1021/tx200008f.

(47) Castro-Perez, J.; Plumb, R.; Liang, L.; Yang, E. A High-Throughput Liquid Chromatography/Tandem Mass Spectrometry Method for Screening Glutathione Conjugates Using Exact Mass Neutral Loss Acquisition. *Rapid Commun. Mass Spectrom.* **2005**, *19* (6), 798–804. https://doi.org/10.1002/rcm.1855.

(48) Haroldsen, P. E.; Reilly, M. H.; Hughes, H.; Gaskell, S. J.; Porter, C. J. Characterization of Glutathione Conjugates by Fast Atom Bombardment/Tandem Mass Spectrometry. *Biol. Mass Spectrom.* **1988**, *15* (11), 615–621. https://doi.org/10.1002/bms.1200151107.

(49) Pearson, P. G.; Threadgill, M. D.; Howald, W. N.; Baillie, T. A. Applications of Tandem Mass Spectrometry to the Characterization of Derivatized Glutathione Conjugates. Studies withS-(N-Methylcarbamoyl)-Glutathione, a Metabolite of the Antineoplastic agentN-Methylformamide. *Biol. Mass Spectrom.* **1988**, *16* (1–12), 51–56. https://doi.org/10.1002/bms.1200160110.

(50) Prakash, C.; Shaffer, C. L.; Nedderman, A. Analytical Strategies for Identifying Drug Metabolites. *Mass Spectrom. Rev.* **2007**, *26* (3), 340–369. https://doi.org/10.1002/mas.20128.

(51) Tchoumtchoua, J.; Njamen, D.; Mbanya, J. C.; Skaltsounis, A.-L.; Halabalaki, M. Structure-Oriented UHPLC-LTQ Orbitrap-Based Approach as a Dereplication Strategy for the Identification of Isoflavonoids from *Amphimas Pterocarpoides* Crude Extract: Identification of Isoflavonoids by UHPLC-Orbitrap. *J. Mass Spectrom.* **2013**, *48* (5), 561–575. https://doi.org/10.1002/jms.3167.

(52) Wagner, S.; Scholz, K.; Donegan, M.; Burton, L.; Wingate, J.; Völkel, W. Metabonomics and Biomarker Discovery: LC−MS Metabolic Profiling and Constant Neutral Loss Scanning Combined with Multivariate Data Analysis for Mercapturic Acid Analysis. *Anal. Chem.* **2006**, *78* (4), 1296–1305. https://doi.org/10.1021/ac051705s.

(53) Jamin, E. L.; Costantino, R.; Mervant, L.; Martin, J.-F.; Jouanin, I.; Blas-Y-Estrada, F.; Guéraud, F.; Debrauwer, L. Global Profiling of Toxicologically Relevant Metabolites in Urine: Case Study of Reactive Aldehydes. *Anal. Chem.* **2020**, *92* (2), 1746–1754. https://doi.org/10.1021/acs.analchem.9b03146.

(54) McLeod, M. D.; Waller, C. C.; Esquivel, A.; Balcells, G.; Ventura, R.; Segura, J.; Pozo, Ó. J. Constant Ion Loss Method for the Untargeted Detection of Bis-Sulfate Metabolites. *Anal. Chem.* **2017**, *89* (3), 1602–1609. https://doi.org/10.1021/acs.analchem.6b03671.

(55) Sánchez-Guijo, A.; Oji, V.; Hartmann, M. F.; Schuppe, H.-C.; Traupe, H.; Wudy, S. A. High Levels of Oxysterol Sulfates in Serum of Patients with Steroid Sulfatase Deficiency. *J. Lipid Res.* **2015**, *56* (2), 403–412. https://doi.org/10.1194/jlr.M055608.

(56) Fabregat, A.; Pozo, O. J.; Marcos, J.; Segura, J.; Ventura, R. Use of LC-MS/MS for the Open Detection of Steroid Metabolites Conjugated with Glucuronic Acid. *Anal. Chem.* **2013**, *85* (10), 5005–5014. https://doi.org/10.1021/ac4001749.

(57) Xia, Y.-Q.; Miller, J. D.; Bakhtiar, R.; Franklin, R. B.; Liu, D. Q. Use of a Quadrupole Linear Ion Trap Mass Spectrometer in Metabolite Identification and Bioanalysis: Quadrupole Linear Traps in Bioanalysis. *Rapid Commun. Mass Spectrom.* **2003**, *17* (11), 1137–1145. https://doi.org/10.1002/rcm.1037.

(58) Thevis, M.; Geyer, H.; Mareck, U.; Schänzer, W. Screening for Unknown Synthetic Steroids in Human Urine by Liquid Chromatography-Tandem Mass Spectrometry. *J. Mass Spectrom.* **2005**, *40* (7), 955–962. https://doi.org/10.1002/jms.873.

(59) Pozo, O. J.; Deventer, K.; Van Eenoo, P.; Delbeke, F. T. Efficient Approach for the Comprehensive Detection of Unknown Anabolic Steroids and Metabolites in Human Urine by Liquid Chromatography−Electrospray-Tandem Mass Spectrometry. *Anal. Chem.* **2008**, *80* (5), 1709–1720. https://doi.org/10.1021/ac7020757.

(60) Polet, M.; Van Gansbeke, W.; Van Eenoo, P.; Deventer, K. Efficient Approach for the Detection and Identification of New Androgenic Metabolites by Applying SRM GC-CI-MS/MS: A Methandienone Case Study: Detection of New Metabolites by SRM GC-CI-MS/MS. *J. Mass Spectrom.* **2016**, *51* (7), 524–534. https://doi.org/10.1002/jms.3781.

(61) Bessette, E. E.; Goodenough, A. K.; Langouët, S.; Yasa, I.; Kozekov, I. D.; Spivack, S. D.; Turesky, R. J. Screening for DNA Adducts by Data-Dependent Constant Neutral Loss-Triple Stage Mass Spectrometry with a Linear Quadrupole Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (2), 809–819. https://doi.org/10.1021/ac802096p.

(62) Teichert, F.; Winkler, S.; Keun, H. C.; Steward, W. P.; Gescher, A. J.; Farmer, P. B.; Singh, R. Evaluation of Urinary Ribonucleoside Profiling for Clinical Biomarker Discovery Using Constant Neutral Loss Scanning Liquid Chromatography/Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2011**, *25* (14), 2071–2082. https://doi.org/10.1002/rcm.5086.

(63) Murray, K. J.; Carlson, E. S.; Stornetta, A.; Balskus, E. P.; Villalta, P. W.; Balbo, S. Extension of Diagnostic Fragmentation Filtering for Automated Discovery in DNA Adductomics. *Anal. Chem.* **2021**, *93* (14), 5754–5762. https://doi.org/10.1021/acs.analchem.0c04895.

(64) Hsiao, H.-H.; Urlaub, H. Pseudo-Neutral-Loss Scan for Selective Detection of Phosphopeptides and N-Glycopeptides Using Liquid Chromatography Coupled with a Hybrid Linear Ion-Trap/Orbitrap Mass Spectrometer. *PROTEOMICS* **2010**, *10* (21), 3916–3921. https://doi.org/10.1002/pmic.201000290.

(65) Mathieu-Rivet, E.; Scholz, M.; Arias, C.; Dardelle, F.; Schulze, S.; Mauff, F. L.; Teo, G.; Hochmal, A. K.; Blanco-Rivero, A.; Loutelier-Bourhis, C.; Kiefer-Meyer, M.-C.; Fufezan, C.; Burel, C.; Lerouge, P.; Martinez, F.; Bardor, M.; Hippler, M. Exploring the N-Glycosylation Pathway in Chlamydomonas Reinhardtii Unravels Novel Complex Structures *. *Mol. Cell. Proteomics* **2013**, *12* (11), 3160–3183. https://doi.org/10.1074/mcp.M113.028191.

(66) Dai, W.; Tan, J.; Lu, M.; Xie, D.; Li, P.; Lv, H.; Zhu, Y.; Guo, L.; Zhang, Y.; Peng, Q.; Lin, Z. Nontargeted Modification-Specific Metabolomics Investigation of Glycosylated Secondary Metabolites in Tea (*Camellia Sinensis* L.) Based on Liquid Chromatography–High-Resolution Mass Spectrometry. *J. Agric. Food Chem.* **2016**, *64* (35), 6783–6790. https://doi.org/10.1021/acs.jafc.6b02411.

(67) Qin, L.; Wang, Y.; Gong, Y.; Chen, J.; Xu, B.; Tang, L.; Guo, L.; Xie, J. Capsaicin Metabolites and GSH-Associated Detoxification and Biotransformation Pathways in Human Liver Microsomes Revealed by LC-HRMS/MS with Data-Mining Tools. *J. Chromatogr. B* **2019**, *1133*, 121843. https://doi.org/10.1016/j.jchromb.2019.121843.

(68) Zhang, H.; Zhang, D.; Ray, K.; Zhu, M. Mass Defect Filter Technique and Its Applications to Drug Metabolite Identification by High-Resolution Mass Spectrometry. *J. Mass Spectrom.* **2009**, *44* (7), 999–1016. https://doi.org/10.1002/jms.1610.

(69) Pan, H.; Yang, W.; Yao, C.; Shen, Y.; Zhang, Y.; Shi, X.; Yao, S.; Wu, W.; Guo, D. Mass Defect Filtering-Oriented Classification and Precursor Ions List-Triggered High-Resolution Mass Spectrometry Analysis for the Discovery of Indole Alkaloids from Uncaria Sinensis. *J. Chromatogr. A* **2017**, *1516*, 102–113. https://doi.org/10.1016/j.chroma.2017.08.035.

(70) Zhu, P.; Tong, W.; Alton, K.; Chowdhury, S. An Accurate-Mass-Based Spectral-Averaging Isotope-Pattern-Filtering Algorithm for Extraction of Drug Metabolites Possessing a Distinct Isotope Pattern from LC-MS Data. *Anal. Chem.* **2009**, *81* (14), 5910–5917. https://doi.org/10.1021/ac900626d.

(71) Gao, Y.; Zhang, R.; Bai, J.; Xia, X.; Chen, Y.; Luo, Z.; Xu, J.; Gao, Y.; Liu, Y.; He, J.; Abliz, Z. Targeted Data-Independent Acquisition and Mining Strategy for Trace Drug Metabolite Identification Using Liquid Chromatography Coupled with Tandem Mass Spectrometry. *Anal. Chem.* **2015**, *87* (15), 7535–7539. https://doi.org/10.1021/acs.analchem.5b01205.

(72) Ruan, Q.; Peterman, S.; Szewc, M. A.; Ma, L.; Cui, D.; Humphreys, W. G.; Zhu, M. An Integrated Method for Metabolite Detection and Identification Using a Linear Ion Trap/Orbitrap Mass Spectrometer and Multiple Data Processing Techniques: Application to Indinavir Metabolite Detection. *J. Mass Spectrom.* **2008**, *43* (2), 251–261. https://doi.org/10.1002/jms.1311.

(73) Zhang, H.; Zhang, D.; Ray, K. A Software Filter to Remove Interference Ions from Drug Metabolites in Accurate Mass Liquid Chromatography/Mass Spectrometric Analyses. *J. Mass Spectrom.* **2003**, *38* (10), 1110–1112. https://doi.org/10.1002/jms.521.

(74) Liang, W.; Zheng, F.; Chen, T.; Zhang, X.; Xia, Y.; Li, Z.; Lu, X.; Zhao, C.; Xu, G. Nontargeted Screening Method for Veterinary Drugs and Their Metabolites Based on Fragmentation Characteristics from Ultrahigh-Performance Liquid Chromatography-High-Resolution Mass Spectrometry. *Food Chem.* **2022**, *369*, 130928. https://doi.org/10.1016/j.foodchem.2021.130928.

(75) Zhang, M.; Liu, Y.; Chen, J.; Liu, H.; Lu, X.; Wu, J.; Zhang, Y.; Lin, Y.; Liu, Q.; Wang, H.; Guo, L.; Gao, R.; Xu, B.; Xie, J. Sensitive Untargeted Screening of Nerve Agents and Their Degradation Products Using Liquid Chromatography–High Resolution Mass Spectrometry. *Anal. Chem.* **2020**, *92* (15), 10578–10587. https://doi.org/10.1021/acs.analchem.0c01508.

(76) Drug, E.; Gershonov, E.; Ashkenazi, N.; Zafrani, Y.; Chen, R.; Dagan, S. Software-Assisted Automated Detection and Identification of “Unknown” Analogues: Implementation on V-Type Nerve Agents. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (8), 1541–1547. https://doi.org/10.1021/jasms.2c00114.

(77) Mayer, P. M.; Poon, C. The Mechanisms of Collisional Activation of Ions in Mass Spectrometry. *Mass Spectrom. Rev.* **2009**, *28* (4), 608–639. https://doi.org/10.1002/mas.20225.

(78) Ichou, F.; Schwarzenberg, A.; Lesage, D.; Alves, S.; Junot, C.; Machuron-Mandard, X.; Tabet, J.-C. Comparison of the Activation Time Effects and the Internal Energy Distributions for the CID, PQD and HCD Excitation Modes. *J. Mass Spectrom.* **2014**, *49* (6), 498–508. https://doi.org/10.1002/jms.3365.

(79) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; Wolan,

D. W.; Spilker, M. E.; Benton, H. P.; Siuzdak, G. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* **2018**, *90* (5), 3156–3164. https://doi.org/10.1021/acs.analchem.7b04424.

(80) Zheng, C.; Hao, H.; Wang, X.; Wu, X.; Wang, G.; Sang, G.; Liang, Y.; Xie, L.; Xia, C.; Yao, X. Diagnostic Fragment-Ion-Based Extension Strategy for Rapid Screening and Identification of Serial Components of Homologous Families Contained in Traditional Chinese Medicine Prescription Using High-Resolution LC-ESI- IT-TOF/MS: Shengmai Injection as an Example. *J. Mass Spectrom.* **2009**, *44* (2), 230–244. https://doi.org/10.1002/jms.1502.

(81) Dai, W.; Yin, P.; Zeng, Z.; Kong, H.; Tong, H.; Xu, Z.; Lu, X.; Lehmann, R.; Xu, G. Nontargeted Modification-Specific Metabolomics Study Based on Liquid Chromatography–High-Resolution Mass Spectrometry. *Anal. Chem.* **2014**, *86* (18), 9146–9153. https://doi.org/10.1021/ac502045j.

(82) Reis, A.; Domingues, P.; Ferrer-Correia, A. J. V.; Domingues, M. R. M. Tandem Mass Spectrometry of Intact Oxidation Products of Diacylphosphatidylcholines: Evidence for the Occurrence of the Oxidation of the Phosphocholine Head and Differentiation of Isomers. *J. Mass Spectrom.* **2004**, *39* (12), 1513–1522. https://doi.org/10.1002/jms.751.

(83) Zemski Berry, K. A.; Murphy, R. C. Electrospray Ionization Tandem Mass Spectrometry of Glycerophosphoethanolamine Plasmalogen Phospholipids. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (10), 1499–1508. https://doi.org/10.1016/j.jasms.2004.07.009.

(84) Kuuranne, T.; Vahermo, M.; Leinonen, A.; Kostiainen, R. Electrospray and Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometric Behavior of Eight Anabolic Steroid Glucuronides. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (8), 722–730. https://doi.org/10.1016/S1044-0305(00)00135-5.

(85) Habicht, S. C.; Vinueza, N. R.; Duan, P.; Fu, M.; Kenttämaa, H. I. Data-Dependent Neutral Gain MS3: Toward Automated Identification of the N-Oxide Functional Group in Drug Metabolites. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (4), 559–563. https://doi.org/10.1016/j.jasms.2009.12.015.

(86) Gomez-Gomez, A.; Rodríguez-Morató, J.; Haro, N.; Marín-Corral, J.; Masclans, J. R.; Pozo, O. J. Untargeted Detection of the Carbonyl Metabolome by Chemical Derivatization and Liquid Chromatography-Tandem Mass Spectrometry in Precursor Ion Scan Mode: Elucidation of COVID-19 Severity Biomarkers. *Anal. Chim. Acta* **2022**, *1196*, 339405. https://doi.org/10.1016/j.aca.2021.339405.

(87) Mahajan, M. K.; Evans, C. A. Dual Negative Precursor Ion Scan Approach for Rapid Detection of Glutathione Conjugates Using Liquid Chromatography/Tandem Mass

Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2008**, *22* (7), 1032–1040. https://doi.org/10.1002/rcm.3458.

(88) Stornetta, A.; Villalta, P. W.; Hecht, S. S.; Sturla, S. J.; Balbo, S. Screening for DNA Alkylation Mono and Cross-Linked Adducts with a Comprehensive LC-MS3 Adductomic Approach. *Anal. Chem.* **2015**, *87* (23), 11706–11713. https://doi.org/10.1021/acs.analchem.5b02759.

(89) Liebler, D. C.; Guengerich, F. P. Elucidating Mechanisms of Drug-Induced Toxicity. *Nat. Rev. Drug Discov.* **2005**, *4* (5), 410–420. https://doi.org/10.1038/nrd1720.

(90) Thompson, R. A.; Isin, E. M.; Ogese, M. O.; Mettetal, J. T.; Williams, D. P. Reactive Metabolites: Current and Emerging Risk and Hazard Assessments. *Chem. Res. Toxicol.* **2016**, *29* (4), 505–533. https://doi.org/10.1021/acs.chemrestox.5b00410.

(91) Wild, C. P. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **2005**, *14* (8), 1847–1850. https://doi.org/10.1158/1055-9965.EPI-05-0456.

(92) Perbellini, L.; Veronese, N.; Princivalle, A. Mercapturic Acids in the Biological Monitoring of Occupational Exposure to Chemicals. *J. Chromatogr. B* **2002**, *781* (1), 269–290. https://doi.org/10.1016/S1570-0232(02)00501-9.

(93) Ma, S.; Subramanian, R. Detecting and Characterizing Reactive Metabolites by Liquid Chromatography/Tandem Mass Spectrometry. *J. Mass Spectrom.* **2006**, *41* (9), 1121–1139. https://doi.org/10.1002/jms.1098.

(94) Dieckhaus, C. M.; Fernández-Metzler, C. L.; King, R.; Krolikowski, P. H.; Baillie, T. A. Negative Ion Tandem Mass Spectrometry for the Detection of Glutathione Conjugates. *Chem. Res. Toxicol.* **2005**, *18* (4), 630–638. https://doi.org/10.1021/tx049741u.

(95) Wen, B.; Ma, L.; Nelson, S. D.; Zhu, M. High-Throughput Screening and Characterization of Reactive Metabolites Using Polarity Switching of Hybrid Triple Quadrupole Linear Ion Trap Mass Spectrometry. *Anal. Chem.* **2008**, *80* (5), 1788–1799. https://doi.org/10.1021/ac702232r.

(96) Jin, L.; Baillie, T. A. Metabolism of the Chemoprotective Agent Diallyl Sulfide to Glutathione Conjugates in Rats. *Chem. Res. Toxicol.* **1997**, *10* (3), 318–327. https://doi.org/10.1021/tx9601768.

(97) Lu, D.; Peterson, L. A. Identification of Furan Metabolites Derived from Cysteine−cis-2-Butene-1,4-dial−Lysine Cross-Links. *Chem. Res. Toxicol.* **2010**, *23* (1), 142–151. https://doi.org/10.1021/tx9003215.

(98) Parmar, K. R.; Jhajra, S.; Singh, S. Detection of Glutathione Conjugates of Amiodarone and Its Reactive Diquinone Metabolites in Rat Bile Using Mass Spectrometry Tools. *Rapid Commun. Mass Spectrom.* **2016**, *30* (10), 1242–1248. https://doi.org/10.1002/rcm.7545.

(99) Dekant, R.; Langer, M.; Lupp, M.; Adaku Chilaka, C.; Mally, A. In Vitro and In Vivo Analysis of Ochratoxin A-Derived Glucuronides and Mercapturic Acids as Biomarkers of Exposure. *Toxins* **2021**, *13* (8), 587. https://doi.org/10.3390/toxins13080587.

(100) Rincon Nigro, M. E.; Du, T.; Gao, S.; Kaur, M.; Xie, H.; Olaleye, O. A.; Liang, D. Metabolite Identification of a Novel Anti-Leishmanial Agent OJT007 in Rat Liver Microsomes Using LC-MS/MS. *Molecules* **2022**, *27* (9), 2854. https://doi.org/10.3390/molecules27092854.

(101) Rousu, T.; Pelkonen, O.; Tolonen, A. Rapid Detection and Characterization of Reactive Drug Metabolites in Vitro Using Several Isotope-Labeled Trapping Agents and Ultra-Performance Liquid Chromatography/Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, *23* (6), 843–855. https://doi.org/10.1002/rcm.3953.

(102) Walsh, J. P.; Renaud, J. B.; Hoogstra, S.; McMullin, D. R.; Ibrahim, A.; Visagie, C. M.; Tanney, J. B.; Yeung, K. K. -C.; Sumarah, M. W. Diagnostic Fragmentation Filtering for the Discovery of New Chaetoglobosins and Cytochalasins. *Rapid Commun. Mass Spectrom.* **2019**, *33* (1), 133–139. https://doi.org/10.1002/rcm.8306.

(103) Bi, Q.; Hou, J.; Yang, M.; Shen, Y.; Qi, P.; Feng, R.; Dai, Z.; Yan, B.; Wang, J.; Shi, X.; Wu, W.; Guo, D. A Strategy Combining Higher Energy C-Trap Dissociation with Neutral Loss- and Product Ion-Based MS ^{n} Acquisition for Global Profiling and Structure Annotation of Fatty Acids Conjugates. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (3), 443–451. https://doi.org/10.1007/s13361-016-1558-y.

(104) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; Bushuiev, R.; Murray, K. J.; Hoffmann, N.; Lu, M.; Sarvepalli, A.; Zhang, Z.; Fleischauer, M.; Dührkop, K.; Wesner, M.; Hoogstra, S. J.; Rudt, E.; Mokshyna, O.; Brungs, C.; Ponomarov, K.; Mutabdžija, L.; Damiani, T.; Pudney, C. J.; Earll, M.; Helmer, P. O.; Fallon, T. R.; Schulze, T.; Rivas-Ubach, A.; Bilbao, A.; Richter, H.; Nothias, L.-F.; Wang, M.; Orešič, M.; Weng, J.-K.; Böcker, S.; Jeibmann, A.; Hayen, H.; Karst, U.; Dorrestein, P. C.; Petras, D.; Du, X.; Pluskal, T. Integrative Analysis of Multimodal Mass Spectrometry Data in MZmine 3. *Nat. Biotechnol.* **2023**, *41* (4), 447–449. https://doi.org/10.1038/s41587-023-01690-2.

(105) Renaud, J. B.; Kelman, M. J.; McMullin, D. R.; Yeung, K. K.-C.; Sumarah, M. W. Application of C8 Liquid Chromatography-Tandem Mass Spectrometry for the Analysis of Enniatins and Bassianolides. *J. Chromatogr. A* **2017**, *1508*, 65–72. https://doi.org/10.1016/j.chroma.2017.05.070.

(106) Zhang, J.-Y.; Zhang, Q.; Li, N.; Wang, Z.-J.; Lu, J.-Q.; Qiao, Y.-J. Diagnostic Fragment-Ion-Based and Extension Strategy Coupled to DFIs Intensity Analysis for Identification of Chlorogenic Acids Isomers in Flos Lonicerae Japonicae by HPLC-ESI-MSn. *Talanta* **2013**, *104*, 1–9. https://doi.org/10.1016/j.talanta.2012.11.012.

(107) Yang, W.; Si, W.; Zhang, J.; Yang, M.; Pan, H.; Wu, J.; Qiu, S.; Yao, C.; Hou, J.; Wu, W.; Guo, D. Selective and Comprehensive Characterization of the Quinochalcone C-Glycoside Homologs in Carthamus Tinctorius L. by Offline Comprehensive Two-Dimensional Liquid Chromatography/LTQ-Orbitrap MS Coupled with Versatile Data Mining Strategies. *RSC Adv.* **2016**, *6* (1), 495–506. https://doi.org/10.1039/C5RA23744K.

(108) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics MCP* **2012**, *11* (6), O111.016717. https://doi.org/10.1074/mcp.O111.016717.

(109) Walmsley, S. J.; Guo, J.; Murugan, P.; Weight, C. J.; Wang, J.; Villalta, P. W.; Turesky, R. J. Comprehensive Analysis of DNA Adducts Using Data-Independent wSIM/MS2 Acquisition and wSIM-City. *Anal. Chem.* **2021**, *93* (16), 6491–6500. https://doi.org/10.1021/acs.analchem.1c00362.

(110) Sousa, P. F. M.; Martella, G.; Åberg, K. M.; Esfahani, B.; Motwani, H. V. nLossFinder—A Graphical User Interface Program for the Nontargeted Detection of DNA Adducts. *Toxics* **2021**, *9* (4), 78. https://doi.org/10.3390/toxics9040078.

(111) Guo, J.; Villalta, P. W.; Turesky, R. J. Data-Independent Mass Spectrometry Approach for Screening and Identification of DNA Adducts. *Anal. Chem.* **2017**, *89* (21), 11728–11736. https://doi.org/10.1021/acs.analchem.7b03208.

(112) Guo, J.; Huan, T. Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography–Mass Spectrometry Based Untargeted Metabolomics. *Anal. Chem.* **2020**, *92* (12), 8072–8080. https://doi.org/10.1021/acs.analchem.9b05135.

(113) Davidson, J. T.; Sasiene, Z. J.; Jackson, G. P. Comparison of In-source Collision-induced Dissociation and Beam-type Collision-induced Dissociation of Emerging Synthetic Drugs Using a High-resolution Quadrupole Time-of-flight Mass Spectrometer. *J. Mass Spectrom.* **2021**, *56* (2). https://doi.org/10.1002/jms.4679.

(114) Gabelica, V.; Pauw, E. D. Internal Energy and Fragmentation of Ions Produced in Electrospray Sources. *Mass Spectrom. Rev.* **2005**, *24* (4), 566–587. https://doi.org/10.1002/mas.20027.

(115) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Benton, H. P.; Siuzdak, G. Annotation: A Computational Solution for Streamlining Metabolomics Analysis. *Anal. Chem.* **2018**, *90* (1), 480–489. https://doi.org/10.1021/acs.analchem.7b03929.

(116) Lee, S. H.; Choi, D. W. Comparison between Source-Induced Dissociation and Collision-Induced Dissociation of Ampicillin, Chloramphenicol, Ciprofloxacin, and Oxytetracycline via Mass Spectrometry. *Toxicol. Res.* **2013**, *29* (2), 107–114. https://doi.org/10.5487/TR.2013.29.2.107.

(117) An, N.; Zhu, Q.-F.; Wang, Y.-Z.; Xiong, C.-F.; Hu, Y.-N.; Feng, Y.-Q. Integration of Chemical Derivatization and In-Source Fragmentation Mass Spectrometry for High-Coverage Profiling of Submetabolomes. *Anal. Chem.* **2021**, *93* (32), 11321–11328. https://doi.org/10.1021/acs.analchem.1c02673.

(118) Parcher, J. F.; Wang, M.; Chittiboyina, A. G.; Khan, I. A. In-Source Collision-Induced Dissociation (IS-CID): Applications, Issues and Structure Elucidation with Single-Stage Mass Analyzers. *Drug Test. Anal.* **2018**, *10* (1), 28–36. https://doi.org/10.1002/dta.2249.

(119) Kolmonen, M.; Leinonen, A.; Kuuranne, T.; Pelander, A.; Ojanperä, I. Hydrophilic Interaction Liquid Chromatography and Accurate Mass Measurement for Quantification and Confirmation of Morphine, Codeine and Their Glucuronide Conjugates in Human Urine. *J. Chromatogr. B* **2010**, *878* (29), 2959–2966. https://doi.org/10.1016/j.jchromb.2010.08.047.

(120) Guddat, S.; Thevis, M.; Thomas, A.; Schänzer, W. Rapid Screening of Polysaccharide-Based Plasma Volume Expanders Dextran and Hydroxyethyl Starch in Human Urine by Liquid Chromatography–Tandem Mass Spectrometry. *Biomed. Chromatogr.* **2008**, *22* (7), 695–701. https://doi.org/10.1002/bmc.986.

(121) Wang, H.; Chen, H.; Geng, J.; Zheng, Y.; Zhang, Z.; Sun, L.; Tai, G.; Zhou, Y. Quantitative Analysis of Dextran in Rat Plasma Using Q-Orbitrap Mass Spectrometry Based on All Ion Fragmentation Strategy. *J. Chromatogr. B* **2018**, *1095*, 24–31. https://doi.org/10.1016/j.jchromb.2018.07.015.

(122) Humbert, L.; Grisel, F.; Richeval, C.; Lhermitte, M. Screening of Xenobiotics by Ultra-Performance Liquid Chromatography-Mass Spectrometry Using In-Source Fragmentation at Increasing Cone Voltages: Library Constitution and an Evaluation of Spectral Stability. *J. Anal. Toxicol.* **2010**, *34* (9), 571–580. https://doi.org/10.1093/jat/34.9.571.

(123) Marquet, P.; Venisse, N.; Lacassie, É.; Lachâtre, G. In-Source CID Mass Spectral Libraries for the “General Unknown” Screening of Drugs and Toxicants. *Analusis* **2000**, *28* (10), 925–934. https://doi.org/10.1051/analusis:2000280925.

(124) de Castro, A.; Gergov, M.; Östman, P.; Ojanperä, I.; Pelander, A. Combined Drug Screening and Confirmation by Liquid Chromatography Time-of-Flight Mass Spectrometry with Reverse Database Search. *Anal. Bioanal. Chem.* **2012**, *403* (5), 1265–1278. https://doi.org/10.1007/s00216-012-5889-7.

(125) Yang, W.; Shi, X.; Yao, C.; Huang, Y.; Hou, J.; Han, S.; Feng, Z.; Wei, W.; Wu, W.; Guo, D. A Novel Neutral Loss/Product Ion Scan-Incorporated Integral Approach for the Untargeted Characterization and Comparison of the Carboxyl-Free Ginsenosides from Panax Ginseng, Panax Quinquefolius, and Panax Notoginseng. *J. Pharm. Biomed. Anal.* **2020**, *177*, 112813. https://doi.org/10.1016/j.jpba.2019.112813.

(126) Ramabulana, A.-T.; Steenkamp, P.; Madala, N.; Dubery, I. A. Profiling of Chlorogenic Acids from Bidens Pilosa and Differentiation of Closely Related Positional Isomers with the Aid of UHPLC-QTOF-MS/MS-Based In-Source Collision-Induced Dissociation. *Metabolites* **2020**, *10* (5), 178. https://doi.org/10.3390/metabo10050178.

(127) Xue, J.; Domingo-Almenara, X.; Guijas, C.; Palermo, A.; Rinschen, M. M.; Isbell, J.; Benton, H. P.; Siuzdak, G. Enhanced In-Source Fragmentation Annotation Enables Novel Data Independent Acquisition and Autonomous METLIN Molecular Identification. *Anal. Chem.* **2020**, *92* (8), 6051–6059. https://doi.org/10.1021/acs.analchem.0c00409.

(128) Cai, W.; Zhang, J.-Y.; Dong, L.-Y.; Yin, P.-H.; Wang, C.-G.; Lu, J.-Q.; Zhang, H.-G. Identification of the Metabolites of Ixerin Z from Ixeris Sonchifolia Hance in Rats by HPLC–LTQ-Orbitrap Mass Spectrometry. *J. Pharm. Biomed. Anal.* **2015**, *107*, 290–297. https://doi.org/10.1016/j.jpba.2015.01.004.

(129) Sleno, L. The Use of Mass Defect in Modern Mass Spectrometry. *J. Mass Spectrom.* **2012**, *47* (2), 226–236. https://doi.org/10.1002/jms.2953.

(130) Su, C.-Y.; Wang, J.-H.; Chang, T.-Y.; Shih, C.-L. Mass Defect Filter Technique Combined with Stable Isotope Tracing for Drug Metabolite Identification Using High-Resolution Mass Spectrometry. *Anal. Chim. Acta* **2022**, *1208*, 339814. https://doi.org/10.1016/j.aca.2022.339814.

(131) Liu, M.; Zhao, S.; Wang, Z.; Wang, Y.; Liu, T.; Li, S.; Wang, C.; Wang, H.; Tu, P. Identification of Metabolites of Deoxyschizandrin in Rats by UPLC–Q-TOF-MS/MS Based on Multiple Mass Defect Filter Data Acquisition and Multiple Data Processing Techniques. *J. Chromatogr. B* **2014**, *949–950*, 115–126. https://doi.org/10.1016/j.jchromb.2013.12.022.

(132) Liang, Y.; Xiao, W.; Dai, C.; Xie, L.; Ding, G.; Wang, G.; Meng, Z.; Zhang, J.; Kang, A.; Xie, T.; Liu, Y.; Zhou, Y.; Liu, W.; Zhao, L.; Xu, J. Structural Identification of the Metabolites for Strictosamide in Rats Bile by an Ion Trap-TOF Mass Spectrometer and

Mass Defect Filter Technique. *J. Chromatogr. B* **2011**, *879* (20), 1819–1822. https://doi.org/10.1016/j.jchromb.2011.04.015.

(133) Zhu, M.; Ma, L.; Zhang, H.; Humphreys, W. G. Detection and Structural Characterization of Glutathione-Trapped Reactive Metabolites Using Liquid Chromatography−High-Resolution Mass Spectrometry and Mass Defect Filtering. *Anal. Chem.* **2007**, *79* (21), 8333–8341. https://doi.org/10.1021/ac071119u.

(134) Zhu, C.; Cai, T.; Jin, Y.; Chen, J.; Liu, G.; Xu, N.; Shen, R.; Chen, Y.; Han, L.; Wang, S.; Wu, C.; Zhu, M. Artificial Intelligence and Network Pharmacology Based Investigation of Pharmacological Mechanism and Substance Basis of Xiaokewan in Treating Diabetes. *Pharmacol. Res.* **2020**, *159*, 104935. https://doi.org/10.1016/j.phrs.2020.104935.

(135) Xiao, Y.; Wang, Y.-K.; Xiao, X.-R.; Zhao, Q.; Huang, J.-F.; Zhu, W.-F.; Li, F. Metabolic Profiling of Coumarins by the Combination of UPLC-MS-Based Metabolomics and Multiple Mass Defect Filter. *Xenobiotica* **2020**, *50* (9), 1076–1089. https://doi.org/10.1080/00498254.2020.1744047.

(136) Geng, J.; Dai, Y.; Yao, Z.; Qin, Z.; Wang, X.; Qin, L.; Yao, X. Metabolites Profile of Xian-Ling-Gu-Bao Capsule, a Traditional Chinese Medicine Prescription, in Rats by Ultra Performance Liquid Chromatography Coupled with Quadrupole Time-of-Flight Tandem Mass Spectrometry Analysis. *J. Pharm. Biomed. Anal.* **2014**, *96*, 90–103. https://doi.org/10.1016/j.jpba.2014.03.024.

(137) Shi, T.; Yao, Z.; Qin, Z.; Ding, B.; Dai, Y.; Yao, X. Identification of Absorbed Constituents and Metabolites in Rat Plasma after Oral Administration of Shen-Song-Yang-Xin Using Ultra-Performance Liquid Chromatography Combined with Quadrupole Time-of-Flight Mass Spectrometry. *Biomed. Chromatogr.* **2015**, *29* (9), 1440–1452. https://doi.org/10.1002/bmc.3443.

(138) Xu, M.; Yi, L.; Ren, D.; Li, B. Mass Defect Filtering Combined with Molecular Networking to Profile Flavonoids in Citrus Fruit Based on Liquid Chromatography-High Resolution Mass Spectrometry Platform: Citrus Sinensis (L.) Osbeck as a Case Study. *J. Chromatogr. A* **2022**, *1685*, 463640. https://doi.org/10.1016/j.chroma.2022.463640.

(139) Di Paola-Naranjo, R. D.; Baroni, M. V.; Podio, N. S.; Rubinstein, H. R.; Fabani, M. P.; Badini, R. G.; Inga, M.; Ostera, H. A.; Cagnoni, M.; Gallegos, E.; Gautier, E.; Peral-García, P.; Hoogewerff, J.; Wunderlin, D. A. Fingerprints for Main Varieties of Argentinean Wines: Terroir Differentiation by Inorganic, Organic, and Stable Isotopic Analyses Coupled to Chemometrics. *J. Agric. Food Chem.* **2011**, *59* (14), 7854–7865. https://doi.org/10.1021/jf2007419.

(140) Tsednee, M.; Huang, Y.-C.; Chen, Y.-R.; Yeh, K.-C. Identification of Metal Species by ESI-MS/MS through Release of Free Metals from the Corresponding Metal-Ligand Complexes. *Sci. Rep.* **2016**, *6* (1), 26785. https://doi.org/10.1038/srep26785.

(141) Roullier, C.; Guitton, Y.; Valery, M.; Amand, S.; Prado, S.; Robiou du Pont, T.; Grovel, O.; Pouchus, Y. F. Automated Detection of Natural Halogenated Compounds from LC-MS Profiles–Application to the Isolation of Bioactive Chlorinated Compounds from Marine-Derived Fungi. *Anal. Chem.* **2016**, *88* (18), 9143–9150. https://doi.org/10.1021/acs.analchem.6b02128.

(142) LeBlanc, A.; Shiao, T. C.; Roy, R.; Sleno, L. Improved Detection of Reactive Metabolites with a Bromine-Containing Glutathione Analog Using Mass Defect and Isotope Pattern Matching. *Rapid Commun. Mass Spectrom.* **2010**, *24* (9), 1241–1250. https://doi.org/10.1002/rcm.4507.

(143) Baars, O.; Morel, F. M. M.; Perlman, D. H. ChelomEx: Isotope-Assisted Discovery of Metal Chelates in Complex Media Using High-Resolution LC-MS. *Anal. Chem.* **2014**, *86* (22), 11298–11305. https://doi.org/10.1021/ac503000e.

(144) Ma, L.; Wen, B.; Ruan, Q.; Zhu, M. Rapid Screening of Glutathione-Trapped Reactive Metabolites by Linear Ion Trap Mass Spectrometry with Isotope Pattern-Dependent Scanning and Postacquisition Data Mining. *Chem. Res. Toxicol.* **2008**, *21* (7), 1477–1483. https://doi.org/10.1021/tx8001115.

(145) Lim, H.-K.; Chen, J.; Cook, K.; Sensenhauser, C.; Silva, J.; Evans, D. C. A Generic Method to Detect Electrophilic Intermediates Using Isotopic Pattern Triggered Data-Dependent High-Resolution Accurate Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22* (8), 1295–1311. https://doi.org/10.1002/rcm.3504.

(146) Defoy, D.; Dansette, P. M.; Neugebauer, W.; Wagner, J. R.; Klarskov, K. Evaluation of Deuterium Labeled and Unlabeled Bis-Methyl Glutathione Combined with Nanoliquid Chromatography−Mass Spectrometry to Screen and Characterize Reactive Drug Metabolites. *Chem. Res. Toxicol.* **2011**, *24* (3), 412–417. https://doi.org/10.1021/tx1003694.

(147) Mutlib, A.; Lam, W.; Atherton, J.; Chen, H.; Galatsis, P.; Stolle, W. Application of Stable Isotope Labeled Glutathione and Rapid Scanning Mass Spectrometers in Detecting and Characterizing Reactive Metabolites. *Rapid Commun. Mass Spectrom.* **2005**, *19* (23), 3482–3492. https://doi.org/10.1002/rcm.2223.

(148) Mezine, I.; Bode, C.; Raughley, B.; Bhoopathy, S.; Roberts, K. J.; Owen, A. J.; Hidalgo, I. J. Application of Exogenous Mixture of Glutathione and Stable Isotope Labeled Glutathione for Trapping Reactive Metabolites in Cryopreserved Human Hepatocytes. Detection of the Glutathione Conjugates Using High Resolution Accurate Mass

Spectrometry. *Chem. Biol. Interact.* **2013**, *204* (3), 173–184. https://doi.org/10.1016/j.cbi.2013.05.011.

(149) Baumeister, T. U. H.; Ueberschaar, N.; Schmidt-Heck, W.; Mohr, J. F.; Deicke, M.; Wichard, T.; Guthke, R.; Pohnert, G. DeltaMS: A Tool to Track Isotopologues in GC- and LC-MS Data. *Metabolomics* **2018**, *14* (4), 41. https://doi.org/10.1007/s11306-018-1336-x.

(150) Castro-Falcón, G.; Hahn, D.; Reimer, D.; Hughes, C. C. Thiol Probes To Detect Electrophilic Natural Products Based on Their Mechanism of Action. *ACS Chem. Biol.* **2016**, *11* (8), 2328–2336. https://doi.org/10.1021/acschembio.5b00924.

(151) Wichard, T. Identification of Metallophores and Organic Ligands in the Chemosphere of the Marine Macroalga Ulva (Chlorophyta) and at Land-Sea Interfaces. *Front. Mar. Sci.* **2016**, *3*.

(152) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nat. Methods* **2019**, *16* (4), 299–302. https://doi.org/10.1038/s41592-019-0344-8.

(153) Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; van der Hooft, J. J. J.; Ernst, M.; Kang, K. B.; Aceves, C. M.; Caraballo-Rodríguez, A. M.; Koester, I.; Weldon, K. C.; Bertrand, S.; Roullier, C.; Sun, K.; Tehan, R. M.; Boya P., C. A.; Christian, M. H.; Gutiérrez, M.; Ulloa, A. M.; Tejeda Mora, J. A.; Mojica-Flores, R.; Lakey-Beitia, J.; Vásquez-Chaves, V.; Zhang, Y.; Calderón, A. I.; Tayler, N.; Keyzers, R. A.; Tugizimana, F.; Ndlovu, N.; Aksenov, A. A.; Jarmusch, A. K.; Schmid, R.; Truman, A. W.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Reproducible Molecular Networking of Untargeted Mass Spectrometry Data Using GNPS. *Nat. Protoc.* **2020**, *15* (6), 1954–1991. https://doi.org/10.1038/s41596-020-0317-5.

(154) Li, Y.; Kind, T.; Folz, J.; Vaniya, A.; Mehta, S. S.; Fiehn, O. Spectral Entropy Outperforms MS/MS Dot Product Similarity for Small-Molecule Compound Identification. *Nat. Methods* **2021**, *18* (12), 1524–1531. https://doi.org/10.1038/s41592-021-01331-z.

(155) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for

Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3* (1), 160018. https://doi.org/10.1038/sdata.2016.18.

(156) Dipple, A. DNA Adducts of Chemical Carcinogens. *Carcinogenesis* **1995**, *16* (3), 437–441. https://doi.org/10.1093/carcin/16.3.437.

(157) Wogan, G. N.; Hecht, S. S.; Felton, J. S.; Conney, A. H.; Loeb, L. A. Environmental and Chemical Carcinogenesis. *Semin. Cancer Biol.* **2004**, *14* (6), 473–486. https://doi.org/10.1016/j.semcancer.2004.06.010.

(158) Wiencke, J. K. DNA Adduct Burden and Tobacco Carcinogenesis. *Oncogene* **2002**, *21* (48), 7376–7391. https://doi.org/10.1038/sj.onc.1205799.

(159) Xiao, S.; Guo, J.; Yun, B. H.; Villalta, P. W.; Krishna, S.; Tejpaul, R.; Murugan, P.; Weight, C. J.; Turesky, R. J. Biomonitoring DNA Adducts of Cooked Meat Carcinogens in Human Prostate by Nano Liquid Chromatography–High Resolution Tandem Mass Spectrometry: Identification of 2-Amino-1-Methyl-6-Phenylimidazo[4,5-b]Pyridine DNA Adduct. *Anal. Chem.* **2016**, *88* (24), 12508–12515. https://doi.org/10.1021/acs.analchem.6b04157.

(160) Singer, B.; Grunberger, D. *Molecular Biology of Mutagens and Carcinogens*; Springer US: Boston, MA, 1983. https://doi.org/10.1007/978-1-4613-3772-0.

(161) Balbo, S.; Hecht, S. S.; Upadhyaya, P.; Villalta, P. W. Application of a High-Resolution Mass-Spectrometry-Based DNA Adductomics Approach for Identification of DNA Adducts in Complex Mixtures. *Anal. Chem.* **2014**, *86* (3), 1744–1752. https://doi.org/10.1021/ac403565m.

(162) Farmer, P. B.; Singh, R. Use of DNA Adducts to Identify Human Health Risk from Exposure to Hazardous Environmental Pollutants: The Increasing Role of Mass Spectrometry in Assessing Biologically Effective Doses of Genotoxic Carcinogens. *Mutat. Res. Mutat. Res.* **2008**, *659* (1), 68–76. https://doi.org/10.1016/j.mrrev.2008.03.006.

(163) Gangl, E. T.; Turesky, R. J.; Vouros, P. Determination of in Vitro- and in Vivo-Formed DNA Adducts of 2-Amino-3-Methylimidazo[4,5-f]Quinoline by Capillary Liquid Chromatography/Microelectrospray Mass Spectrometry. *Chem. Res. Toxicol.* **1999**, *12* (10), 1019–1027. https://doi.org/10.1021/tx990060m.

(164) Wolf, S. M.; Vouros, P. Application of Capillary Liquid Chromatography Coupled with Tandem Mass Spectrometric Methods to the Rapid Screening of Adducts Formed by the Reaction of N-Acetoxy-N-Acetyl-2-Aminofluorene with Calf DNA. *Chem. Res. Toxicol.* **1994**, *7* (1), 82–88. https://doi.org/10.1021/tx00037a013.

(165) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat. Methods* **2015**, *12* (6), 523–526. https://doi.org/10.1038/nmeth.3393.

(166) Roushan, A.; Wilson, G. M.; Kletter, D.; Sen, K. I.; Tang, W.; Kil, Y. J.; Carlson, E.; Bern, M. Peak Filtering, Peak Annotation, and Wildcard Search for Glycoproteomics. *Mol. Cell. Proteomics* **2021**, *20*, 100011. https://doi.org/10.1074/mcp.RA120.002260.

(167) Medzihradszky, K. F.; Kaasik, K.; Chalkley, R. J. Characterizing Sialic Acid Variants at the Glycopeptide Level. *Anal. Chem.* **2015**, *87* (5), 3064–3071. https://doi.org/10.1021/ac504725r.

(168) T, P.; S, C.; A, V.-B.; M, O. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* **2010**, *11*. https://doi.org/10.1186/1471-2105-11-395.

(169) Lao, Y.; Villalta, P. W.; Sturla, S. J.; Wang, M.; Hecht, S. S. Quantitation of Pyridyloxobutyl DNA Adducts of Tobacco-Specific Nitrosamines in Rat Tissue DNA by High-Performance Liquid Chromatography−Electrospray Ionization−Tandem Mass Spectrometry. *Chem. Res. Toxicol.* **2006**, *19* (5), 674–682. https://doi.org/10.1021/tx050351x.

(170) Upadhyaya, P.; Kalscheuer, S.; Hochalter, J. B.; Villalta, P. W.; Hecht, S. S. Quantitation of Pyridylhydroxybutyl-DNA Adducts in Liver and Lung of F-344 Rats Treated with 4-(Methylnitrosamino)-1-(3-Pyridyl)-1-Butanone and Enantiomers of Its Metabolite 4-(Methylnitrosamino)-1-(3-Pyridyl)-1-Butanol. *Chem. Res. Toxicol.* **2008**, *21* (7), 1468–1476. https://doi.org/10.1021/tx8001109.

(171) Zhang, S.; Villalta, P. W.; Wang, M.; Hecht, S. S. Detection and Quantitation of Acrolein-Derived 1,N2-Propanodeoxyguanosine Adducts in Human Lung by Liquid Chromatography-Electrospray Ionization-Tandem Mass Spectrometry. *Chem. Res. Toxicol.* **2007**, *20* (4), 565–571. https://doi.org/10.1021/tx700023z.

(172) Guidolin, V.; Carrà, A.; Carlson, E. S.; Villalta, P. W.; Balbo, S. Molecular Characterization of Alcohol-Induced DNA Damage for Cancer Prevention, 2018.

(173) Wilson, M. R.; Jiang, Y.; Villalta, P. W.; Stornetta, A.; Boudreau, P. D.; Carrá, A.; Brennan, C. A.; Chun, E.; Ngo, L.; Samson, L. D.; Engelward, B. P.; Garrett, W. S.; Balbo, S.; Balskus, E. P. The Human Gut Bacterial Genotoxin Colibactin Alkylates DNA. *Science* **2019**, *363* (6428), eaar7785. https://doi.org/10.1126/science.aar7785.

(174) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.;

Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920. https://doi.org/10.1038/nbt.2377.

(175) Bryant, M. S.; Lay, J. O.; Chiarelli, M. P. Development of Fast Atom Bombardment Mass Spectral Methods for the Identification of Carcinogen-Nucleoside Adducts. *J. Am. Soc. Mass Spectrom.* **1992**, *3* (4), 360–371. https://doi.org/10.1016/1044-0305(92)87064-6.

(176) Cuevas-Ramos, G.; Petit, C. R.; Marcq, I.; Boury, M.; Oswald, E.; Nougayrède, J.-P. Escherichia Coli Induces DNA Damage in Vivo and Triggers Genomic Instability in Mammalian Cells. *Proc. Natl. Acad. Sci.* **2010**, *107* (25), 11537–11542. https://doi.org/10.1073/pnas.1001261107.

(177) Buc, E.; Dubois, D.; Sauvanet, P.; Raisch, J.; Delmas, J.; Darfeuille-Michaud, A.; Pezet, D.; Bonnet, R. High Prevalence of Mucosa-Associated E. Coli Producing Cyclomodulin and Genotoxin in Colon Cancer. *PloS One* **2013**, *8* (2), e56964. https://doi.org/10.1371/journal.pone.0056964.

(178) Pluskal, T.; Uehara, T.; Yanagida, M. Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Anal. Chem.* **2012**, *84* (10), 4396–4403. https://doi.org/10.1021/ac3000418.

(179) Faïs, T.; Delmas, J.; Barnich, N.; Bonnet, R.; Dalmasso, G. Colibactin: More Than a New Bacterial Toxin. *Toxins* **2018**, *10* (4), 151. https://doi.org/10.3390/toxins10040151.

(180) Jiang, Y.; Stornetta, A.; Villalta, P. W.; Wilson, M. R.; Boudreau, P. D.; Zha, L.; Balbo, S.; Balskus, E. P. Reactivity of an Unusual Amidase May Explain Colibactin's DNA Cross-Linking Activity. *J. Am. Chem. Soc.* **2019**, *141* (29), 11489–11496. https://doi.org/10.1021/jacs.9b02453.

(181) Xue, M.; Kim, C. S.; Healy, A. R.; Wernke, K. M.; Wang, Z.; Frischling, M. C.; Shine, E. E.; Wang, W.; Herzon, S. B.; Crawford, J. M. Structure Elucidation of Colibactin and Its DNA Cross-Links. *Science* **2019**, *365* (6457), eaax2685. https://doi.org/10.1126/science.aax2685.

(182) Altmann, L.; Böttger, A.; Wiegand, H. Neurophysiological and Psychophysical Measurements Reveal Effects of Acute Low-Level Organic Solvent Exposure in Humans. *Int. Arch. Occup. Environ. Health* **1990**, *62* (7), 493–499. https://doi.org/10.1007/BF00381179.

(183) Sarigiannis, D. A.; Papaioannou, N.; Handakas, E.; Anesti, O.; Polanska, K.; Hanke, W.; Salifoglou, A.; Gabriel, C.; Karakitsios, S. Neurodevelopmental Exposome: The Effect of in Utero Co-Exposure to Heavy Metals and Phthalates on Child Neurodevelopment. *Environ. Res.* **2021**, *197*, 110949. https://doi.org/10.1016/j.envres.2021.110949.

(184) Beulens, J. W. J.; Pinho, M. G. M.; Abreu, T. C.; den Braver, N. R.; Lam, T. M.; Huss, A.; Vlaanderen, J.; Sonnenschein, T.; Siddiqui, N. Z.; Yuan, Z.; Kerckhoffs, J.; Zhernakova, A.; Brandao Gois, M. F.; Vermeulen, R. C. H. Environmental Risk Factors of Type 2 Diabetes—an Exposome Approach. *Diabetologia* **2022**, *65* (2), 263–274. https://doi.org/10.1007/s00125-021-05618-w.

(185) Orešič, M.; McGlinchey, A.; Wheelock, C. E.; Hyötyläinen, T. Metabolic Signatures of the Exposome—Quantifying the Impact of Exposure to Environmental Chemicals on Human Health. *Metabolites* **2020**, *10* (11), 454. https://doi.org/10.3390/metabo10110454.

(186) Wild, C. P. Environmental Exposure Measurement in Cancer Epidemiology. *Mutagenesis* **2009**, *24* (2), 117–125. https://doi.org/10.1093/mutage/gen061.

(187) Papp, E. M.; Miller, A. S. Screening and Surveillance. OSHA's Medical Surveillance Provisions. *AAOHN J. Off. J. Am. Assoc. Occup. Health Nurses* **2000**, *48* (2), 59–72.

(188) Husten, C. G.; Deyton, L. R. Understanding the Tobacco Control Act: Efforts by the US Food and Drug Administration to Make Tobacco-Related Morbidity and Mortality Part of the USA's Past, Not Its Future. *Lancet Lond. Engl.* **2013**, *381* (9877), 1570–1580. https://doi.org/10.1016/S0140-6736(13)60735-7.

(189) Carmella, S. G.; Chen, M.; Zhang, Y.; Zhang, S.; Hatsukami, D. K.; Hecht, S. S. Quantitation of Acrolein-Derived (3-Hydroxypropyl)Mercapturic Acid in Human Urine by Liquid Chromatography-Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry: Effects of Cigarette Smoking. *Chem. Res. Toxicol.* **2007**, *20* (7), 986–990. https://doi.org/10.1021/tx700075y.

(190) Park, S. L.; Carmella, S. G.; Chen, M.; Patel, Y.; Stram, D. O.; Haiman, C. A.; Marchand, L. L.; Hecht, S. S. Mercapturic Acids Derived from the Toxicants Acrolein and Crotonaldehyde in the Urine of Cigarette Smokers from Five Ethnic Groups with Differing Risks for Lung Cancer. *PLOS ONE* **2015**, *10* (6), e0124841. https://doi.org/10.1371/journal.pone.0124841.

(191) Hanna, P. E.; Anders, M. W. The Mercapturic Acid Pathway. *Crit. Rev. Toxicol.* **2019**, *49* (10), 819–929. https://doi.org/10.1080/10408444.2019.1692191.

(192) Chen, M.; Carmella, S. G.; Lindgren, B. R.; Luo, X.; Ikuemonisan, J.; Niesen, B.; Thomson, N. M.; Murphy, S. E.; Hatsukami, D. K.; Hecht, S. S. Increased Levels of the Acrolein Metabolite 3-Hydroxypropyl Mercapturic Acid in the Urine of e-Cigarette Users. *Chem. Res. Toxicol.* **2023**, *36* (4), 583–588. https://doi.org/10.1021/acs.chemrestox.2c00145.

(193) Products, C. for T. Harmful and Potentially Harmful Constituents in Tobacco Products and Tobacco Smoke: Established List. *FDA* **2020**.

(194) Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X. One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Anal. Chem.* **2017**, *89* (17), 8696–8703. https://doi.org/10.1021/acs.analchem.7b00947.

(195) R Core Team. R: A Language and Environment for Statistical Computing, 2023. https://www.R-project.org/.

(196) Trevor Hastie; Robert Tibshirani; Balasubramanian Narasimhan; Gilbert Chu. Impute: Imputation for Microarray Data, 2023.

(197) Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using Lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

(198) Maximilian Helf. MassTools: R Package for Mass and Molecular Formula Calculations, 2023.

(199) Kind, T.; Fiehn, O. Seven Golden Rules for Heuristic Filtering of Molecular Formulas Obtained by Accurate Mass Spectrometry. *BMC Bioinformatics* **2007**, *8* (1), 105. https://doi.org/10.1186/1471-2105-8-105.

(200) Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D. A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.-É.; Li, S.; Xia, J. MetaboAnalyst 5.0: Narrowing the Gap between Raw Spectra and Functional Insights. *Nucleic Acids Res.* **2021**, *49* (W1), W388–W396. https://doi.org/10.1093/nar/gkab382.

(201) Thévenot, E. A.; Roux, A.; Xu, Y.; Ezan, E.; Junot, C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J. Proteome Res.* **2015**, *14* (8), 3322–3335. https://doi.org/10.1021/acs.jproteome.5b00354.

(202) Benjamini, Yoav; Hochberg, Yosef. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57* (1), 289–300.

(203) Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M. A.; Petras, D.; Gerwick, W. H.; Rousu, J.; Dorrestein, P. C.; Böcker, S. Systematic Classification of Unknown Metabolites Using High-Resolution Fragmentation Mass Spectra. *Nat. Biotechnol.* **2021**, *39* (4), 462–471. https://doi.org/10.1038/s41587-020-0740-8.

(204) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminformatics* **2016**, *8*, 61. https://doi.org/10.1186/s13321-016-0174-y.

(205) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), 12580–12585. https://doi.org/10.1073/pnas.1509788112.

(206) Hoffmann, M. A.; Nothias, L.-F.; Ludwig, M.; Fleischauer, M.; Gentry, E. C.; Witting, M.; Dorrestein, P. C.; Dührkop, K.; Böcker, S. High-Confidence Structural Annotation of Metabolites Absent from Spectral Libraries. *Nat. Biotechnol.* **2022**, *40* (3), 411–421. https://doi.org/10.1038/s41587-021-01045-9.

(207) Chen, M.; Carmella, S. G.; Li, Y.; Zhao, Y.; Hecht, S. S. Resolution and Quantitation of Mercapturic Acids Derived from Crotonaldehyde, Methacrolein, and Methyl Vinyl Ketone in the Urine of Smokers and Nonsmokers. *Chem. Res. Toxicol.* **2020**, *33* (2), 669–677. https://doi.org/10.1021/acs.chemrestox.9b00491.

(208) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.

(209) Johnson, C. H.; Patterson, A. D.; Idle, J. R.; Gonzalez, F. J. Xenobiotic Metabolomics: Major Impact on the Metabolome. *Annu. Rev. Pharmacol. Toxicol.* **2012**, *52*, 37–56. https://doi.org/10.1146/annurev-pharmtox-010611-134748.

(210) Lamplugh, A.; Harries, M.; Xiang, F.; Trinh, J.; Hecobian, A.; Montoya, L. D. Occupational Exposure to Volatile Organic Compounds and Health Risks in Colorado Nail Salons. *Environ. Pollut.* **2019**, *249*, 518–526. https://doi.org/10.1016/j.envpol.2019.03.086.

(211) Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Rathman, J. F.; Richarz, A.-N.; Yang, C. A Review of in Silico Toxicology Approaches to Support the Safety Assessment of Cosmetics-Related Materials. *Comput. Toxicol.* **2022**, *21*, 100213. https://doi.org/10.1016/j.comtox.2022.100213.

(212) Baute-Pérez, D.; Santana-Mayor, Á.; Herrera-Herrera, A. V.; Socas-Rodríguez, B.; Rodríguez-Delgado, M. Á. Analysis of Alkylphenols, Bisphenols and Alkylphenol Ethoxylates in Microbial-Fermented Functional Beverages and Bottled Water: Optimization of a Dispersive Liquid-Liquid Microextraction Protocol Based on Natural Hydrophobic Deep Eutectic Solvents. *Food Chem.* **2022**, *377*, 131921. https://doi.org/10.1016/j.foodchem.2021.131921.

(213) Kobylewski, S.; Jacobson, M. F. Toxicology of Food Dyes. *Int. J. Occup. Environ. Health* **2012**, *18* (3), 220–246. https://doi.org/10.1179/1077352512Z.00000000034.

(214) Valerio, L. G. In Silico Toxicology for the Pharmaceutical Sciences. *Toxicol. Appl. Pharmacol.* **2009**, *241* (3), 356–370. https://doi.org/10.1016/j.taap.2009.08.022.

(215) Brouwer, D. H.; de Pater, N. A. J.; Zomer, C.; Lurvink, M. W. M.; van Hemmen, J. J. An Experimental Study to Investigate the Feasibility to Classify Paints According to Neurotoxicological Risks: Occupational Air Requirement (OAR) and Indoor Use of Alkyd Paints. *Ann. Occup. Hyg.* **2005**, *49* (5), 443–451. https://doi.org/10.1093/annhyg/mei007.

(216) Schulz, M.; Schmoldt, A.; Andresen-Streichert, H.; Iwersen-Bergmann, S. Revisited: Therapeutic and Toxic Blood Concentrations of More than 1100 Drugs and Other Xenobiotics. *Crit. Care* **2020**, *24* (1), 195. https://doi.org/10.1186/s13054-020-02915-5.

(217) Nakayama Wong, L. S.; Aung, H. H.; Lamé, M. W.; Wegesser, T. C.; Wilson, D. W. Fine Particulate Matter from Urban Ambient and Wildfire Sources from California's San Joaquin Valley Initiate Differential Inflammatory, Oxidative Stress, and Xenobiotic Responses in Human Bronchial Epithelial Cells. *Toxicol. In Vitro* **2011**, *25* (8), 1895–1905. https://doi.org/10.1016/j.tiv.2011.06.001.

(218) Idle, J. R.; Gonzalez, F. J. Metabolomics. *Cell Metab.* **2007**, *6* (5), 348–351. https://doi.org/10.1016/j.cmet.2007.10.005.

(219) Alwis, K. U.; Blount, B. C.; Britt, A. S.; Patel, D.; Ashley, D. L. Simultaneous Analysis of 28 Urinary VOC Metabolites Using Ultra High Performance Liquid Chromatography Coupled with Electrospray Ionization Tandem Mass Spectrometry (UPLC-ESI/MSMS). *Anal. Chim. Acta* **2012**, *750*, 152–160. https://doi.org/10.1016/j.aca.2012.04.009.

(220) Carmella, S. G.; Chen, M.; Han, S.; Briggs, A.; Jensen, J.; Hatsukami, D. K.; Hecht, S. S. Effects of Smoking Cessation on Eight Urinary Tobacco Carcinogen and Toxicant Biomarkers. *Chem. Res. Toxicol.* **2009**, *22* (4), 734–741. https://doi.org/10.1021/tx800479s.

(221) Luo, X.; Carmella, S. G.; Chen, M.; Jensen, J. A.; Wilkens, L. R.; Le Marchand, L.; Hatsukami, D. K.; Murphy, S. E.; Hecht, S. S. Urinary Cyanoethyl Mercapturic Acid, a Biomarker of the Smoke Toxicant Acrylonitrile, Clearly Distinguishes Smokers From Nonsmokers. *Nicotine Tob. Res.* **2020**, *22* (10), 1744–1747. https://doi.org/10.1093/ntr/ntaa080.

(222) Keith, R. J.; Fetterman, J. L.; Orimoloye, O. A.; Dardari, Z.; Lorkiewicz, P. K.; Hamburg, N. M.; DeFilippis, A. P.; Blaha, M. J.; Bhatnagar, A. Characterization of Volatile Organic Compound Metabolites in Cigarette Smokers, Electronic Nicotine Device Users, Dual Users, and Nonusers of Tobacco. *Nicotine Tob. Res.* **2019**, *22* (2), 264–272. https://doi.org/10.1093/ntr/ntz021.

(223) Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50* (D1), D622–D631. https://doi.org/10.1093/nar/gkab1062.

(224) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry

Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(225) Guha, R. Chemical Informatics Functionality in R. *J. Stat. Softw.* **2007**, *18*, 1–16. https://doi.org/10.18637/jss.v018.i05.

(226) Adams, K. J.; Pratt, B.; Bose, N.; Dubois, L. G.; St John-Williams, L.; Perrott, K. M.; Ky, K.; Kapahi, P.; Sharma, V.; MacCoss, M. J.; Moseley, M. A.; Colton, C. A.; MacLean, B. X.; Schilling, B.; Thompson, J. W.; Alzheimer's Disease Metabolomics Consortium. Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *J. Proteome Res.* **2020**, *19* (4), 1447–1458. https://doi.org/10.1021/acs.jproteome.9b00640.