Improving the Analysis of Nitrogen-Containing Organics in Source Water

by

Kristin Carroll

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

Master of Science

in

Analytical and Environmental Toxicology

Department of Laboratory Medicine and Pathology

University of Alberta

Abstract

Drinking water treatment is a necessary practice for the elimination of waterborne disease. The study of drinking water treatment can be challenging as large portions of the composition of the source water used by treatment plants remains unknown. In **Chapter 1**, I review the importance of drinking water treatment, and potential challenges and concerns it faces. I then review current methods used for the analysis of water. While analytical methods are becoming well developed, a remaining challenge is in the data analysis for the identification of new compounds in water. I draw attention to available platforms and some limitations that remain. I emphasize the importance of the careful application of available data analysis programs and important considerations when they are developed.

In **Chapter 2**, a stable isotopic labelling, high-performance liquid chromatography (HPLC) – high resolution mass spectrometry (HRMS) method is applied for the analysis of amine containing species in source water. Previously, data analysis for this method had been done manually. Here, I compare the manual data analysis with a pilot program we developed, HDPairFinder. Two source water samples were analyzed and the various parameters to select peaks on considered. I reviewed the benefits of each type of data analysis, differences between the pairs selected, and the possibility for further application of this program to source water samples.

In **Chapter 3**, I first review the final development of HDPairFinder and the merits it can provide to the data analysis procedure. Following the development of HDPairFinder, I applied it to analyze various groups of samples. Using HDPairFinder, I was able to analyze the composition of the samples over the course of spring runoff. Differences between sampling

locations, sampling years, and stage of the water disinfection process were observed. Additionally, I investigated the potential use of pooled samples for the ability to be used as quality controls in the future. Overall, the application of HDPairFinder was able to help in the analysis of temporal trends in amine containing compounds in source water.

This thesis worked to improve the complete analysis procedure for amine containing compounds in source water. Through the development of HDPairFinder many improvements to data analysis were made. These improvements included automated data processing, reduced time for data analysis, and additional data cleaning steps. Additionally, multiple samples can be processed and therefore HDPairFinder is applied to study changes over time. In the application, HDPairFinder helped to provide insight into the composition and changes in composition over the course of spring runoff. This work will help to lay a foundation for future identification of amine containing compounds in the future. In the identification of these compounds, we hope to provide insight into how to improve water treatment in the future.

Preface

Parts of **Chapter 2** and **Chapter 3** of this thesis were published in Zhao, T.; Carroll, K.; Craven, C. B.; Wawryk, N. J. P.; Xing, S.; Guo, J.; Li, X.; Huan, T. HDPairFinder : A Data Processing Platform for Hydrogen / Deuterium Isotopic Labeling-Based Nontargeted Analysis of Trace-Level Amino-Containing Chemicals in Environmental Water. J. Environ. Sci. 2024, 136, 583–593. Reprinted with permission. Copyright 2023 Elsevier. Available from: https://doi.org/10.1016/j.jes.2023.02.033. As co-first author on this publication ran the stable isotopic labelling on all the source water samples used for the development. Tingting Zhao did the development and coding of the program. I provided feedback and tested the program. We jointly wrote this manuscript. Caley B. Craven and Nicholas J.P. Wawryk assisted in running and data collection of the source water samples. Shipei Xing and Jian Guo provided technical assistance during the development of the program. As supervisory authors, Tao Huan and Xing-Fang Li oversaw the concept development, manuscript composition, and editing.

Acknowledgements

To my supervisor Dr. Xing-Fang Li, thank you for all of your support and guidance throughout my graduate program. You have provided opportunities and challenged me to grow as an individual. I will always be thankful for everything you have provided for me. To my committee members Dr. Hongquan Zhang and Dr. Andrei Drabovich, thank you for your guidance and support in finishing my graduate program. I would also like to thank Dr. Liang Li for serving on my examining committee and Dr. Jelena Holovati for chairing and supporting graduate students. I am thankful for Dr. Tao Huan and his group from UBC for their collaboration to build and complete HDPairFinder and for all the advice about data analysis you provided. To Katerina Carastathis and Barb Thomson for their help with everything administrative and making sure I was on track and had everything I needed throughout my program. And to the rest of my analytical and Environmental Toxicology Division and to the Department of Laboratory Medicine and Pathology, thank you for all your support.

To all of the friends I have made throughout my program, I could not have done it without you. Jenny Chau, Dr. Caley Craven, and (almost Dr.) Nicholas Wawryk, I could not imagine the long lab days without you. Thank you to Teresa Kumblathan for always making sure I was taking care of myself. Thank you for all of the training, support, and lab memories you have given me. To Tingting Zhao, I am grateful I had the opportunity to get to know you and collaborate with you. My thesis would not be possible without all of your hard work. Qiming Shen and Karen Hoy, thank you for your help with editing my thesis. To the rest of my research group, Yanming, Mary, and Di, I appreciate your support and advice.

Much of this work would not be possible without the generous support from the EPCOR team in Edmonton. The samples and advice they provided were invaluable. To Wendell James and Trevor Shu, thank you for your support as a part of the collaboration we have built with EPCOR. I am also grateful to the University of Alberta and Alberta Innovates for their financial support during my graduate program.

Finally, I would like to thank the family and friends for their ongoing support throughout my graduate program journey. To Deven Goett and Rebecca Clark, I could not have done this without the support from calling you every week. To my Mom, Dad, and sisters Katie and Kameryn, I appreciate all of your encouragement of my academic endeavors and advice throughout. To the rest of the Carroll, Telang, Raugust, and Heizer family, I am grateful for all of the support you have given me. To Seth Heizer, thank you for your love, support, and encouragement. I could not have made it through without having you behind me.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

ACN	Acetonitrile
BAC	Biologically Activated Carbon
CCL-5	Contaminant Candidate List 5
C-DBPs	Carbonaceous Disinfection By-products
ClarE	Clarifier Effluent
C18	Carbon-18
DBP	Disinfection By-product
DDA	Data Dependent Acquisition
DIA	Data Independent Acquisition
DLLME	Dispersive Liquid-Liquid Microextraction
EI	Electron Impact
ELS	E.L. Smith
ESI	Electrospray Ionization
FA	Formic Acid
FDR	False Discovery Rate
GAC	Granular Activated Carbon
GC	Gas Chromatography
HAN	Haloacetonitrile
HAM	Haloacetoamide
H/D	Hydrogen/Deuterium
HNM	Halonitromethane

HILIC	Hydrophilic Interaction Chromatography
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
HSS	High Strength Silica
IDA	Information Dependent Acquisition
LLE	Liquid-liquid extraction
MAC	Maximum Acceptable Concentration
MS	Mass Spectrometry
m/z	Mass-to-charge Ratio
N-DBPs	Nitrogenous Disinfection-Byproducts
NDMA	N-Nitrosodimethylamine
NH3-N	Ammonia as Nitrogen
NOM	Natural Organic Matter
NP	Normal Phase
PAC	Powdered Activated Carbon
PFAS	Perfluoroalkyl and Polyfluoroalkyl Substances
PMTs	Persistent, Mobile, and Toxic Substances
PPCPs	Pharmaceuticals and Personal Care Products
ROS	Rossdale
RP	Reverse Phase
SPE	Solid Phase Extraction
SPME	Solid Phase Microextraction
TKN	Total Kjeldahl Nitrogen

TOC	Total Organic Carbon
TON	Total Organic Nitrogen
UN	United Nations
USEPA	United States Environmental Protection Agency
WHO	World Health Organization
WTP	Water Treatment Plant

Chapter 1

Introduction

1.1: Motivation

The United Nations (UN) declared that access to safe drinking water is a basic human need. As such, under UN sustainable goal number 6.1, it states "By 2030, achieve universal and equitable access to safe and affordable drinking water for all".[1] A major component of this goal is to provide drinking water free from microbial pathogens. The World Health Organization (WHO) states infectious disease from these microbial pathogens is the most common and widespread health concern regarding drinking water.[2] In 2020, it was estimated that 74% of the world's population had access to clean drinking water. However, 2 billion people globally still rely on water that is contaminated with microbial pathogens.[2] To control pathogens in drinking water, cost effective water treatment is a necessity.

In Canada, 88% of households receive drinking water provided by water utilities that treat source water from rivers or lakes.[3] An important area of drinking water research is the natural organic matter (NOM) present in this source water. During drinking water treatment, NOM reacts with disinfectants to form disinfection by-products (DBPs), which are potentially harmful to human health. Thus, DBPs require regulations to limit or remove their presence in drinking water. One strategy to control DBP formation is the removal of NOM prior to disinfection.[4] However, a challenge in NOM removal is determining the removal efficiency in complex matrices due to lacking data regarding the composition of water. Additionally, NOM composition will vary depending on the environment and geological location.[5] To characterize NOM, sensitive methods to detect trace compounds are required. Literature shows various

methods that have been well developed, but these methods require further improvements to the data analysis. Thus, it is critical that data analysis is improved for understanding NOM in source water. By improving data analysis, we can better understand water composition and potentially lead to better DBP controls. Further, we can compare how these NOM may change over time.

1.2: Water Treatment

1.2.1: General Water Treatment Practices

The practice of disinfection has virtually eliminated severe waterborne disease in developed countries. However, the complete disinfection of microbial pathogens is ongoing and requires water utilities and regulatory agencies to constantly set and monitor regulatory limits.[6] For example, Health Canada sets the minimum removal level of pathogens that must be reached during water treatment. Regulations state that the pathogens *Giardia* and *Cryptosporidium* must have a minimum of three log removal. Additionally, a common strain of *E.coli,* O157:H7, should not be detected in a 100 mL sample.[7] To reach these regulations, various disinfectants like chlorine, chloramine, ozone, and UV disinfection are used. Another important consideration for pathogen inactivation is water safety throughout the distribution system. To achieve inactivation within the distribution system, residual disinfectant levels need to be maintained, generally at 0.04 to 2.0 mg/L, in Canada.[8,9] Thus, disinfectants help to minimize the acute risks posed by waterborne disease to humans. However, these disinfectants can react with NOM and form DBPs, which pose potential chronic risks to human health. This chronic risk can be reduced and controlled in numerous ways.

One strategy to reduce chronic risks to a persons' health, is to minimize the formation of DBPs through the removal of NOM prior to disinfection. A common practice for NOM removal is coagulation and flocculation.[4] This process encourages the binding of dirt particles, causing

them to settle out of the water. The settling of these particles can carry NOM, which removes larger or higher molecular weight NOM. One strategy for specifically removing smaller molecular weight NOM is the use of activated carbon.[4] Activated carbon can come in various forms including powdered activated carbon (PAC), granular activated carbon (GAC), and biologically activated carbon (BAC).[10–12] However, many factors, such as the pH and water quality, can impact the efficiency of NOM removal. Additionally, as emerging contaminants and water quality events occur, there are corresponding changes to the composition of NOM, which causes difficulties in effectively designing processes that can improve the removal efficiency. Therefore, it is important to understand challenges posed to water treatment facilities and their impacts on the composition of source water.

1.2.2: Challenges Posed to Water Treatment Processes

Numerous challenges can be posed to water treatment facilities, with most challenges being centered around changes in NOM. For example, NOM can have increases in emerging contaminants like pharmaceuticals and personal care products (PPCPs).[13,14] Additionally, persistent chemicals such as per- and polyfluoroalkyl substances (PFAS) have been increasing in the environment.[15] Many of these emerging contaminants lack studies on the concentrations in the environment and their removal during the water treatment process is unknown.[16] Therefore, the extent of the impact they may be having on DBP formation and human health is unknown.

Further challenges arise as sources for drinking water begin changing. One increasingly used source is portable water reuse, which takes wastewater, performs normal treatment, and then delivers the treated water to the drinking water treatment facility. The challenge with portable water reuse is that the composition of water has been shown to have higher concentrations of nitrogen-containing groups, such as urea.[17,18] These nitrogen containing groups

can react to form different DBPs during disinfection. Another challenge of water treatment is global warming, which can cause increased rainfall and forest fires. Increased rainfall can introduce more NOM through greater runoff entering source water, whereas forest fires can increase particulates entering source water.[19–21] These challenges cause difficulties for water treatment plants to completely remove changing NOM. In turn, this changes the formation of DBPs in drinking water. To prepare for these challenges, the ability to quickly screen and identify new compounds in water is required. Studies regarding the fate of NOM during water treatment are necessary to ensure the continued safety of drinking water.

1.3: Disinfection By-products

1.3.1: Formation of DBPs and Precursors

Various compounds from natural and anthropogenic sources make up the NOM entering water treatment plants. More varied NOM means a wider range of compounds acting as precursors to DBPs. For example, amino acids have been shown to form various DBPs, including haloacetonitriles (HANs) and N-chloraldamines.[23,24] Additionally, pharmaceuticals including sulfonamide antibiotics and indole-derivative non-steroidal anti-inflammatories have been shown to be chlorinated during water treatment.[25,26] Other studies have shown the formation of halobenzoquinones from common contaminants bisphenol A and 4-nonylphenol, as well as the common artificial sweetener aspartame.[27,28] A recommended strategy for DBP control is the removal of precursors, or NOM, due to the lack of information and identification on DBPs and their formation.[29] However, a challenge in effectively removing precursors is the incomplete understanding of the composition of source water. To both improve water treatment and the understanding of DBPs, increasing the knowledge regarding the composition of source water and potential DBP precursors is required.

1.3.2: DBP Risks and Regulation

While waterborne diseases pose an acute risk to human health, DBPs may pose a chronic risk. Epidemiological studies have shown a potential association between chronic exposure to DBPs and an increased risk of bladder cancer.[30] To reduce risks, regulation of DBPs is necessary. In Canada, a total of 13 DBPs are regulated using maximum acceptable concentration values (MAC). The inorganic DBPs that are regulated include bromate (0.0 1mg/L), chlorate (1 mg/L), and chlorite (1 mg/L). Regulated organic DBPs include five haloacetic acids (monochloroacetic acid, dichloroacetic acid, trichloroacetic acid, monobromoacetic acid, and dibromoacetic acid) at a total concentration of 0.08 mg/L, four trihalomethanes (chloroform, bromoform, chlorodibromoform, and bromodichlorofom) at a total concentration of 0.1 mg/L, and N-nitrosodimethylamine (NDMA) at 0.04 μg/L.[7] The United States Environmental Protection Agency (USEPA) regulates DBPs in a similar way, including trihalomethanes at 0.08 mg/L, haloacetic acids at 0.06 mg/L, bromate at 0.01 mg/L, and chlorite at 1 mg/L.[31] Additionally, the United States includes a contaminant candidate list (CCL-5), that contains additional DBPs that should be watched and potentially regulated in the future.[32] The DBPs on the CCL-5 mainly include nitrogenous containing DBPs, including HANs, nitrosamines, and halonitromethanes (HNMs). Previous studies have demonstrated that currently regulated DBPs are not the toxicity drivers in water, while nitrogen containing DBPs including HANs, HNMs, and nitrosamines, have been shown to be more toxic then currently regulated DBPs.[33] Thus, the unregulated and unknown DBPs can provide a higher risk to human health.[34] However, gaps in knowledge still remain as many DBPs and their precursors are unknown.

1.3.3: Gaps in DBP Research

A challenge in reducing risks posed by DBPs is the diversity that exists in water treatment. For example, different treatment plants use different disinfectants (ex. chlorine and chloramine). The water entering the treatment plant also shows high degrees of variation in NOM depending on the location, time of year, and environmental factors.[5] Variation in water treatment and a lack of comprehensive understanding makes it difficult to make regulations to control DBP formation. Therefore, ongoing development for quantification and identification of DBPs is required to work towards regulation. Various analytical techniques should be applied to be able to detect and identify new classes of compounds at low concentrations. While numerous analytical techniques are being developed, there is still many challenges with data analysis procedures. A lack of standardization in data analysis reduces the ability to overlap data accurately, which limits the ability to determine the unknown DBPs and precursors. Therefore, development of sensitive and clear methodology for big data analysis is necessary for the identification of unknown DBPs.

1.4: Introduction to Analytical Instrumentation and Methodology

Reviews done by Yang *et al*. and Wawryk *et al*. have shown various analysis methods for DBPs, which could also be applied to many of their precursors.[36,37] Generally, analysis methods for water research require three major steps. First, sample cleanup and removal of matrix effects is required. This step can also include concentrating the analytes of interest, as they are generally at low concentrations in water. In many cases, this step is solid-phase extraction (SPE), solid-phase microextraction (SPME), or liquid-liquid extraction (LLE). The second component of analysis is separation of analytes. In many cases, this is done with high performance liquid chromatography (HPLC) or gas chromatography (GC). Finally, the last step uses detection to

quantify and identify analytes, in many cases with mass spectrometry (MS). Here, the focus will be on SPE-HPLC-MS, as this was the method of choice for analysis in this thesis.

1.4.1: Solid Phase Extraction

To identify unknown DBPs and their precursors, they must first be efficiently extracted (concentrated) from water. Preconcentration is required, as DBPs and their precursors are typically present in μg/L to ng/L ranges in complex matrices. Although many different techniques currently exist for the extraction of small compounds from water, SPE is one of the most common. A variety of solvents and sorbents allows for a diverse range of functional groups to be efficiently extracted. These techniques are widely applied due to their well-known natures and wide scope of analysis.

SPE can cover a large range of applications including environmental analysis, industrial applications, and food analysis.[38] The study of DBPs in water samples is an important area under environmental applications. In DBP analysis, SPE is favored over other methods. Compared to LLE, SPE can reduce toxic waste and lower the amounts of solvents required. SPE allows the possibility for storage of sorbents, compared to LLE where the analyte is extracted into another liquid. Additionally, SPE will prevent any emulsions where DBPs could be lost.[39] Other methods such as solid phase microextraction (SPME) and dispersive liquid-liquid microextraction (DLLME), are newer techniques with similar mechanisms to SPE. SPME and DLLME uses smaller amounts of sorbent, and thus extract a smaller load compared to SPE.[39] Therefore, SPME and DLLME should be considered for further method development in the future.

The basic procedure of SPE includes four simple steps, shown in **Figure 1.1**. These steps include conditioning, loading of sample, washing, and elution. Many methods will load large volumes of samples, in the 1-10 L range. Elution uses smaller volumes, such as 10 mL, allowing

for concentration of the sample. This 10 mL can then be evaporated down to even smaller volumes, such as 0.1 mL. This allows for thousand-fold concentration of these low abundant compounds. The type of interaction occurring during extraction will depend on the sorbent type, the solvent, and the desired analyte. For example, a reversed phase mechanism, one of the most common of SPE for DBP analysis, works through hydrophobic interactions. The DBP will interact with the surface of the solid phase, which is hydrophobic in nature, to partition through adsorption. Examples of common sorbent types and some current applications to DBP analysis are seen in **Table 1.1**. Following the efficient extraction of the analyte of interest, it can be further analyzed.

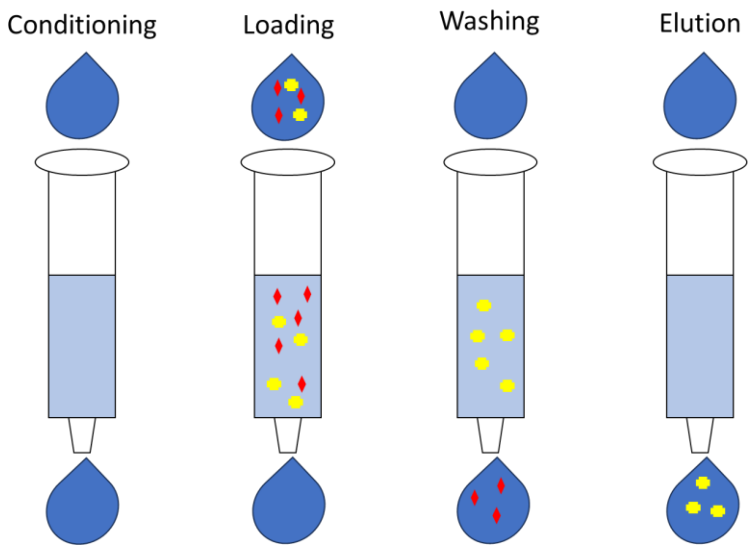
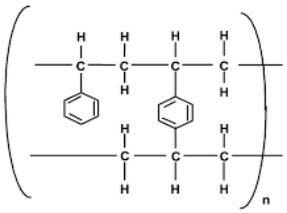
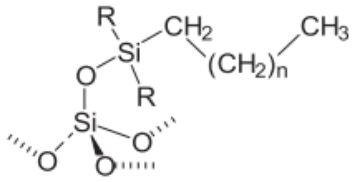
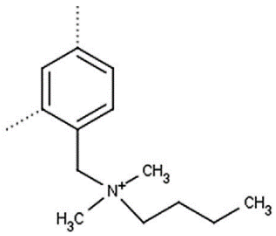
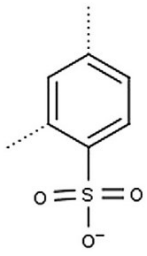
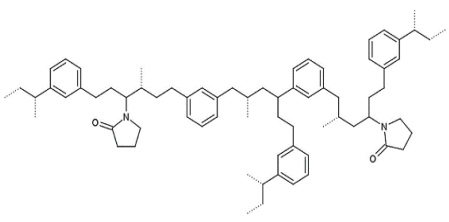


Figure 1.1- Solid Phase Extraction Overview. Interferences are shown in red diamonds and the analyte of interest in yellow crosses.

Table 1.1- Sorbent Types and Associated DBP Classes Extracted.

Cartridge (Sorbent Type)	Sorbent Structure	Mechanism	DBP Classes
XAD Resins		Non-polar hydrophobic copolymer	Total organic halogens (TOX) [40]
C18	 n=16	Reverse Phase Retention	Brominated DBPs, haloacetamides, non-targeted peptides [41–43]
Oasis MAX		Anion Exchange	Iodated DBPs, haloacetamides[44,45]
Oasis MCX		Cation Exchange	Non-targeted amino compounds[46]
Oasis HLB		Hydrophilic-Lipophilic copolymer	Halobenzoquinones, chlorinated amino acids[41,47,48]

1.4.2: High Performance Liquid Chromatography

Separation is a key step in water analysis because thousands of compounds may be detected, thus separating compounds provides better resolution and further information. However, due to the wide range of properties of the compounds, a wide range of separation techniques is required. For example, GC is compatible with volatile DBPs that are easily detected in the gas phase. However, when studying other non-volatile DBPs or their precursors, HPLC would be desirable. While both techniques have been widely applied in the past, my work focuses on the ability to detect water soluble DBP precursors in source water, which requires the use of HPLC.

The main mode or retention, or separation, in liquid chromatography depends on the analyte transferring between two phases. The stationary phase is a solid bonded phase within the column and can have various functional groups bonded to the surface. For example, two commonly used solid phases are Carbon-18 (C18) and biphenyl. The mobile phase is considered the carrier, which moves analytes down the length of the column, and is a combination of water and organic (typically methanol or acetonitrile). The retention and separation of compounds depends on interactions with both the mobile phase and stationary phase. Different types of each, as well as different combinations of mobile phase and stationary phase, lead to various modes of retention mechanisms and allows for the separation of various compounds. Through method development and optimization, efficient separation of all compounds is possible.

Many modes of separation are used in HPLC, including normal phase (NP), reverse phase (RP), and hydrophilic interaction chromatography (HILIC). The different modes of separation will allow for retention of different physiochemical properties depending on the compound of interest. Common RP columns include carbon-based chains like C18. RP is

commonly applied to non-polar and low polarity compounds and uses a more polar mobile phase. On the other hand, NP is applied for the study of polar and hydrophobic compounds. Common NP columns include non-bonded silica or cyano stationary phases and uses a non-polar mobile phase.[49] HILIC is a combination of RP and NP, it uses the stationary phases of NP but with mobile phases of RP. Due to the unique properties of the separation, it can capture polar and hydrophilic analytes.[50] Some examples of the various applications of different columns can be found in **Table 1.2**. As there are a wide range of columns, careful selection is required before application.

Table 1.2- Various column and mobile phase applications to the analysis of DBPs and their precursors.

Stationary Phase	***Compound of Interest***
C18 (RP)	Pharmaceuticals[14], chlorination products of antibiotics[51], chlorinated dipeptides[48]
HSS T3 (RP)	Chlorination products of acetaminophen[52], Pharmaceuticals[14], Persistent mobile and very toxic compounds in water (PMTs)[53]
HILIC	Amino acids[54], peptides and DBPs from peptides[41], PMTs[53]

1.4.3: Mass Spectrometry

While many techniques, such as UV-Vis, are used for the detection of organic compounds, one of the most widely applied for water research is mass spectrometry (MS).[16,55,56] It is generally preferred over other methods, as it allows for sensitive and accurate detection and quantification. When compared to other methods, it can allow for the identification of unknowns.

In the case of water analysis, the application of high-resolution mass spectrometry (HMRS) and non-targeted analysis is required as the composition of source water and many DBPs remain unknown. In particular, tandem mass spectrometry is an important technique for the study of water. This allows for the collection of MS^2 spectra, which creates a fingerprint fragmentation pattern for the compound of interest. This information allows for improved identification of compounds using this technique.

Many types of mass spectrometers exist and can cover a wide range of applications. Lower resolutions instruments, such as triple quadrupoles, allow for the targeted study of a compound of interest. This allows for accurate detection using the specific monitoring of a mass-to-charge ratio (m/z) of interest. Comparatively, higher resolution instruments are necessary for the identification of unknowns, because higher accuracy m/z is required. Additionally, many different ionization methods are available and have been widely applied across environmental research. A commonly applied ionization technique to DBP research is electrospray ionization (ESI). ESI is a soft ionization technique, reducing fragmentation during ionization. It also is compatible with HPLC as it will ionize from the incoming liquid mobile phase. Specifically, in our study, we use HRMS with ESI, as it can efficiently analyze organics in water in which we are interested. However, a challenge with HRMS is the amount of data collected.

HRMS has several types of data collection, which includes data-independent acquisition (DIA), and data-dependent acquisition (DDA). DIA will collect all MS^2 spectra and can be helpful in looking at very low abundance compounds. However, the processing of this data can be difficult. In DDA, also known as information dependent acquisition (IDA), MS^2 spectra will only be collected when the precursor mass in the MS^1 spectra reaches a certain intensity threshold. Therefore, DDA can be useful in reducing the amount of data collected and reducing

the amount of data requiring processing. A recent study systematically comparing the different data collections showed that although DIA can lead to the identification of a higher number of compounds, DDA provides a convenience that eases data analysis.[57] However, even with DDA there is still large amounts of data that require processing. Therefore, additional prioritization strategies are needed.

1.4.4: Isotopic Labelling

When paired with MS, stable isotopic labelling is an effective technique for identification of a specific compound(s) from the collected data. To achieve this, stable isotopic labelling uses non-radioactive isotopes to look for characteristic distributions in MS data. For example, carbon-12 and 13 ($^{12}C/^{13}C$) have been applied to metabolomics studies. Studies have used ^{13}C labelled glucose to identify the metabolites from cells.[58,59] In the case of water research, oxygen isotopes ($^{16}O/^{18}O$) have been used to study ozonation products from water disinfection.[60] While methods such as these look at a single atom, reactions to introduce a functional group with specific isotopic patterns can also exist. For the study of nitrogen-containing species in source water, a methyl-based labelling method has been developed by our research group.[46] In this method, formaldehyde or deuterated formaldehyde is reacted with amine-containing compounds to attach methyl groups on the nitrogen. Thus, the specific isotopic pattern between hydrogen and deuterium (H/D) would be identified in the MS data, allowing for the prioritization of amine-containing species.

Many advantages of stable isotopic labelling exist. By searching the MS data for a specific pattern, data analysis can be improved to focus on one class of compounds of interest. For studying compounds through a water treatment process or metabolism, it allows for accurate tracking of their fate. For methyl-based labelling it has also been shown to improve extraction

efficiencies and therefore improve the sensitivity of the method.[46] Additionally, methyl-based labelling of amino compounds requires the replacement of a hydrogen atom on the nitrogen. Thus, it provides insight into structural information by differentiating a primary and secondary amine based on the number of methyl groups present. However, alongside the advantages of stable isotopic labelling, some limitations remain. In the case of water research, certain isotopic labelling cannot be used. For example, $^{12}C/^{13}C$ are naturally abundant in the environment and therefore could not be used where carbon-containing organics are the compounds of interest. While the study of ozonation products has applied $^{16}O/^{18}O$, it does not differentiate between the distinct types of precursors that are reacting. Therefore, a method that incorporates isotopes that are low abundance in the environment and increased specificity towards a class of compounds of interest is required. These two criteria are ones that the H/D methyl-based labelling of nitrogen-containing species meets.

Prior to the application of a stable isotopic labelling method, few considerations are required. For example, the reaction must be paired to a sensitive detection method to differentiate between the isotopic pattern and random peak distribution. Therefore, an instrument with high sensitivity is required, generally HRMS. Additionally, the ability to separate these compounds and not have any loss of an isotope during ionization is required. Method development for application of these types of reactions needs careful testing and investigation. Another consideration is how to process the data to ensure the correct identification of these compounds. While manual searching for isotopic patterns can be done, many have turned to data analysis programs to assist in the data analysis procedures.

1.5: Big Data and Programs Available for Analysis

1.5.1: Available Programs for Non-targeted Data Analysis

A significant challenge in HPLC-HRMS analysis is the processing of the collected data. Massive amounts are produced, and to try and process data manually can take weeks to months. Additional limitations arise, including not being able to create exhaustive lists of peaks, challenges in identifying compounds, and ability to only process a single sample at a time. Therefore, pursuing the use of various programs is necessary to improve overall data analysis. Many programs have previously been developed to improve data analysis. These programs can be divided into four categories: peak picking, compound identification, visualization tools, and full data processing.

Some examples of currently available data analysis programs are presented in **Figure 1.2**. Peak picking programs, such as XCMS takes raw HPLC-HRMS data files to create a list of peaks and can include additional data cleaning. Programs such as CFM-ID, MetFrag, and SIRIUS, use accurate mass and/or fragmentation patterns, to give tentative identification of a feature.[61–63] Few programs, such as Proteowizard and MS-Dial provide peak picking and identification abilities.[64,65] Visualization tools, including Metaboanalyst and BatMass can be helpful in finding important features within your data, and provide a streamlined avenue to create graphs and charts.[66,67] Finally, some programs, such as MZmine and OpenMS have been designed with the purpose of start to finish processing, from raw data input to visualization.[68,69]

A program that is currently available for environmental analysis is patRoon.[70,71] PatRoon has been developed to cover a start to finish data processing with environmental samples in mind. It completes peak picking and includes various filtering parameters including intensity thresholds, direct comparisons to blanks, and comparison of replicates. It allows for compound

annotation with the ability to filter based on properties. In the updated version, the program allows for automated transformation products. However, this program still has limitations in data analysis, including in-source fragmentation removal and applicability to certain isotopic patterns. Another option for environmental samples is to apply programs developed for proteomics, including some of the ones listed above. For example, visualizations with Metaboanalyst can be easily applied to environmental data. However, it is still important to look at further development of programs for environmental samples, as these are different matrices and are largely uncharacterized.

For the data analysis specific to stable isotopic labelling techniques, a few programs have been developed. As carbon-based labelling, using $^{12}C/^{13}C$ is common, programs such as IsoMS have been developed. IsoMS looks for dansyl-chloride, carbon isotopic patterns in MS data.[72] It includes full data-processing, from peak picking, filtering, and alignment of different datasets. Programs such as IsoMS can help improve the prioritization of MS data. Another example is the processing of H/D labelling patterns. A recently developed program, Peak Pair Pruner, has been developed for the searching for the labelling pattern following preprocessing with MS-DIAL.[73] However, available programs such as Peak Pair Pruner that are designed for H/D labelling are unable to account for retention time shifts due to the deuterium isotopic effect. Therefore, novel programs, that can also be automated to exclude additional processing using algorithms such as MS-DIAL, are required.

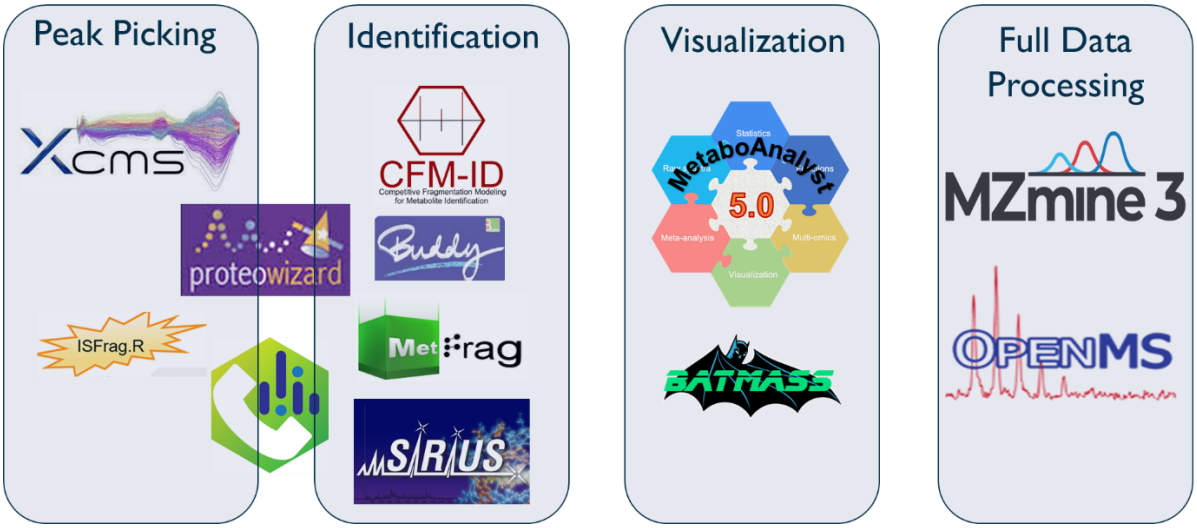


Figure 1.2- Some available programs for mass spectrometry data processing.

1.5.2: Benefits and Limitations with Data Analysis Programs

Many benefits exist from developing programs for data analysis. For example, automation of data analysis, which provides the ability to process more data. Additionally, using data analysis programs can create comprehensive lists of peaks and provide additional data outputs, such as spectral interpretation. If programs are used alongside an analysis method, they can provide consistency in how data is analyzed. A strong benefit is seen in programs that are open access as they provide accessible solutions for any researcher. Many new open access programs are available online and include feedback forms for users. By allowing contribution from users and continuously updating the program, it can adapt to new needs and applications. Open access programs also allow for a user to review the program for their use and understand potential limitations. However, before application of a program, it should be carefully tested for your method and limitations considered.

A remaining challenge in these programs, is false discovery rates (FDRs), which are often increased with the application of a program compared to manual methods. FDRs are when a

feature is incorrectly chosen during data analysis and does not actually exist. FDRs are particularly important during peak picking and in statistical analysis. For example, one study investigated inconsistencies in different peak picking algorithms.[74] Different algorithms have been shown to identify different peaks using the same raw data file and different FDRs. However, studies of peak picking algorithms are not comprehensive as a few programs do not share their algorithm. Challenges can also exist in the ability to apply programs to new problems. As they have not been fully examined, the new application requires careful testing and proof of application over previous data analysis methods. Overall, with careful investigation into new methods before application, programs to improve data analysis can provide solutions to a common bottle neck point in analysis.

A limitation specific to environmental analysis is a lack of databases available to spectral matching. Water can contain a wide range of compounds that continues growing as more are released into the environment. While many databases exist in other fields, a lack of a centralized database for water research limits the ease of searching. This can limit the ability for compound identification with commonly available HRMS data analysis programs. Therefore, it is important to look at creating databases that cover a wide range of compound classes that could capture environmental samples accurately.

1.6: Objectives of Thesis

Nitrogen-containing compounds in source water are largely uncharacterized. To improve removal of organics prior to disinfection treatment, we need to characterize NOM in source water. This is critical to achieve high quality drinking water, by completely inactivating pathogens and minimizing DBPs. N-containing organic compounds in source water are important to this research because they can react to form N-DBPs, some of which are highly

toxic. To characterize N-containing organics, our research group previously developed a stable isotopic labelling method that selectively reacts with amino groups. However, given the huge amount of information generated in HPLC-HRMS analysis, identifying all possible amino chemicals presents a significant challenge in data processing. Combatting challenges with data analysis led to the study described in **Chapter 2**.

Chapter 2 research objectives:

- Apply a stable isotopic labelling method to study nitrogen-containing compounds in source water.
- Develop a data processing program, HDPairFinder, to automate the identification of mass spectral patterns from the stable isotopic labelling method.
- Compare the benefits and limitations of using manual data analysis compared to HDPairFinder.

The development of a program to improve data analysis would allow for the ability to study trends over time. A particular time of interest is spring runoff, where NOM increases in source water. Characterization of NOM during this time would allow for a better understanding of composition of the incoming source water. Thus, it would allow for a better understanding of DBP formation. Specifically studying N-containing organics over the course of the spring would provide insight into potential N-DBP changes over this time. This led to the study described in **Chapter 3.**

Chapter 3 Research Objectives:

- Apply HDPairFinder to characterize changes in source water collected during spring runoff.
- Use data obtained from HDPairFinder to characterize source water samples.

- Investigate pooled samples for sample decomposition.

The anticipated outcomes of my research will improve the data analysis procedure for the detection of N-containing organic in source water. It will allow for a streamlined analysis workflow when compared to manual data analysis. This will improve the ability to detect trends in N-containing organics over the course of the spring. An improved workflow creates a strong foundation for identification of N-containing organics in source water will be provided. This study will provide an increased understanding of the composition of source water, which would direct water treatment processes and regulations in the future.

1.7: References

(1) *United Nations: Goal 6: Ensure access to water and sanitation for all* https://www.un.org/sustainabledevelopment/water-and-sanitation/ (Accessed Jul 26, 2023).

(2) *WHO: Drinking-water* https://www.who.int/news-room/fact-sheets/detail/drinking-water (Accessed Jul 26, 2023).

(3) *Stats Canada: Survey of Drinking Water Plants, 2019*. https://www150.statcan.gc.ca/n1/daily-quotidien/210817/dq210817c-eng.htm (Accessed July 26, 2023).

(4) Jacangelo, J. G.; DeMarco, J.; Owen, D. M.; Stephen, J.; Jacangelo, J. G.; Demarco, J.; Owen, D. M.; Randtke, S. J. Selected Processes for Removing NOM : An Overview. *AWWA*. **1995**, 87 (1), 64–77.

(5) Krasner, S. W. The Formation and Control of Emerging Disinfection By-Products of Health Concern. *Philos. Trans. Math. Phys. Eng. Sci*. **2009**, 367 (1904), 4077–4095. https://doi.org/10.1098/rsta.2009.0108.

(6) Crittenden, J. C.; Trussell, R. R.; Hand, D. W.; Howe, K. J.; Tchbanoglous, G. *MWH's Water Treatment: Principles and Design,* Third.; John Wiley and Sons Inc., **2012**.

(7) Health Canada. *Guidelines for Canadian Drinking Water Quality—Summary Tables*. Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch: Ottawa, ON, 2022; Vol. 24. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/water-quality/guidelines-canadian-drinking-water-quality-summary-table.html. (Accessed July 20th, 2023).

(8) Health Canada. *Guidelines for Canadian Drinking Water Quality: Guideline Technical Document—Chlorine*. Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, 2009. https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-chlorine-guideline-technical-document/page-2-guidelines-canadian-drinking-water-quality-chlorine-guideline-technical-document.html. (Accessed July 20th, 2023).

(9) Hrudey, S. E.; Hrudey, E. J.; Pollard, S. J. T. Risk Management for Assuring Safe Drinking *Water. Environ. Int*. **2006**, 32 (8), 948–957. https://doi.org/10.1016/j.envint.2006.06.004.

(10) Andersson, A.; Lavonen, E.; Harir, M.; Gonsior, M.; Hertkorn, N.; Schmitt-Kopplin, P.; Kylin, H.; Bastviken, D. Selective Removal of Natural Organic Matter during Drinking Water Production Changes the Composition of Disinfection By-Products. *Environ. Sci. Water Res. Technol*. **2020**, 6(3), 779–794. https://doi.org/10.1039/c9ew00931k.

(11) Cuthbertson, A. A.; Kimura, S. Y.; Liberatore, H. K.; Knappe, D. R. U.; Stanford, B.; Summers, R. S.; Dickenson, E. R.; Maness, J. C.; Glover, C.; Selbes, M.; Richardson, S. D. GAC to BAC: Does It Make Chloraminated Drinking Water Safer? *Water Res*. **2020**, 172, 115432. https://doi.org/10.1016/j.watres.2019.115432.

(12) Walker, G. S.; Lee, F. P.; Aieta, E. M. Chlorine Dioxide for Taste and Odor Control. *J. Am. Water Work. Assoc*. **1986**, 78 (3), 84–93. https://doi.org/10.1002/j.1551-8833.1986.tb05719.x.

(13) Bonnefille, B.; Karlsson, O.; Rian, M. B.; Raqib, R.; Parvez, F.; Papazian, S.; Islam, M. S.; Martin, J. W. Nontarget Analysis of Polluted Surface Waters in Bangladesh Using Open Science Workflows. *Environ. Sci. Technol*. **2023**. 57(17), 6808-6824. https://doi.org/10.1021/acs.est.2c08200.

(14) Gros, M.; Rodríguez-Mozaz, S.; Barceló, D. Fast and Comprehensive Multi-Residue Analysis of a Broad Range of Human and Veterinary Pharmaceuticals and Some of Their Metabolites in Surface and Treated Waters by Ultra-High-Performance Liquid Chromatography Coupled to Quadrupole-Linear Ion Trap Tandem. *J. Chromatogr. A* **2012**, 1248, 104–121. https://doi.org/10.1016/j.chroma.2012.05.084.

(15) Charbonnet, J. A.; Mcdonough, C. A.; Xiao, F.; Schwichtenberg, T.; Cao, D.; Kaserzon, S.; Thomas, K. V; Dewapriya, P.; Place, B. J.; Schymanski, E. L.; Field, J. A.; Helbling, D. E.; Higgins, C. P. Communicating Confidence of Per- and Polyfluoroalkyl Substance Identification via High-Resolution Mass Spectrometry. *Environ. Sci. Technol. Lett.* **2022**, 9, 473-481. https://doi.org/10.1021/acs.estlett.2c00206.

(16) Richardson, S. D.; Ternes, T. A. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem*. **2022**, 94 (1), 382–416. https://doi.org/10.1021/acs.analchem.1c04640.

(17) Lau, S. S.; Bokenkamp, K.; Tecza, A.; Wagner, E. D.; Plewa, M. J.; Mitch, W. A. Toxicological Assessment of Potable Reuse and Conventional Drinking Waters. *Nat. Sustain*. **2023**, 6 (1), 39–46. https://doi.org/10.1038/s41893-022-00985-7.

(18) Hu, H. Y.; Du, Y.; Wu, Q. Y.; Zhao, X.; Tang, X.; Chen, Z. Differences in Dissolved Organic Matter between Reclaimed Water Source and Drinking Water Source. *Sci. Total Environ*. **2016**, 551–552, 133–142. https://doi.org/10.1016/j.scitotenv.2015.12.111.

(19) Lepistö, A.; Räike, A.; Sallantaus, T.; Finér, L. Increases in Organic Carbon and Nitrogen Concentrations in Boreal Forested Catchments — Changes Driven by Climate and Deposition. *Sci. Total Environ*. **2021**, 780, 146627. https://doi.org/10.1016/j.scitotenv.2021.146627.

(20) Sebestyen, S. D.; Boyer, E. W.; Shanley, J. B.; Kendall, C.; Doctor, D. H.; Aiken, G. R.; Ohte, N. Sources, Transformations, and Hydrological Processes That Control Stream Nitrate and Dissolved Organic Matter Concentrations during Snowmelt in an Upland Forest. *Water Resources Research* **2008**, 44, 1–14. https://doi.org/10.1029/2008WR006983.

(21) Wallis, P. M.; Hynes, H. B. N.; Telang, S. A. The Importance of Groundwater in the Transportation of Allochthonous Dissolved Organic Matter to the Streams Draining a Small Mountain Basin. *Hydrobiologia* **1981**, 79 (1), 77–90. https://doi.org/10.1007/BF00005821.

(22) Bulloch, D. N.; Nelson, E. D.; Carr, S. A.; Wissman, C. R.; Armstrong, J. L.; Schlenk, D.; Larive, C. K. Occurrence of Halogenated Transformation Products of Selected Pharmaceuticals and Personal Care Products in Secondary and Tertiary Treated Wastewaters from Southern California. *Environ. Sci. Technol*. **2015**, 49 (4), 2044–2051. https://doi.org/10.1021/es504565n.

(23) Tong, Z.; Linge, K. L.; Busetti, F.; Joll, C. A. Formation of Odorous and Hazardous By-Products from the Chlorination of Amino Acids. *Water Res*. **2018**, 146, 10–18. https://doi.org/10.1016/j.watres.2018.08.072.

(24) Linge, K. L.; Kristiana, I.; Liew, D.; Holman, A.; Joll, C. A. Halogenated Semivolatile Acetonitriles as Chloramination Disinfection By-Products in Water Treatment: A New Formation Pathway from Activated Aromatic Compounds. *Environ. Sci. Process. Impacts* **2020**, 22 (3), 653–662. https://doi.org/10.1039/c9em00603f.

(25) Xia, D.; Liu, H.; Lu, Y.; Liu, Y.; Liang, J.; Xie, D.; Lu, G.; Qiu, J.; Wang, R. Utility of a Non-Target Screening Method to Explore the Chlorination of Similar Sulfonamide Antibiotics: Pathways and NCl Intermediates. *Sci. Total Environ*. **2023**, 858, 160042. https://doi.org/10.1016/J.SCITOTENV.2022.160042.

(26) Qiu, J.; Huang, Y.; Wu, Y.; Shi, P.; Xu, B.; Chu, W.; Pan, Y. Detection, Transformation, and Toxicity of Indole-Derivative Nonsteroidal Anti-Inflammatory Drugs during Chlorine Disinfection. *Chemosphere* **2020**, 260, 127579. https://doi.org/10.1016/j.chemosphere.2020.127579.

(27) Wawryk, N. J. P.; Huang, G.; Craven, C.; Qiu, J.; Jmaiff Blackstock, L. K.; Li, X. F. Aspartame-Sweetened Tap Water: Transformation Products and 2,6-Dichloro-1,4-Benzoquinone Formation. *Environ. Sci. Technol*. **2022**, 2–11. https://doi.org/10.1021/acs.est.2c07156.

(28) Kosaka, K.; Nakai, T.; Hishida, Y.; Asami, M.; Ohkubo, K.; Akiba, M. Formation of 2,6-Dichloro-1,4-Benzoquinone from Aromatic Compounds after Chlorination. *Water Res*. **2017**, 110, 48–55. https://doi.org/10.1016/j.watres.2016.12.005.

(29) Diemert, S.; Wang, W.; Andrews, R. C.; Li, X. F. Removal of Halo-Benzoquinone (Emerging Disinfection By-Product) Precursor Material from Three Surface Waters Using Coagulation. *Water Res*. **2013**, 47 (5), 1773–1782. https://doi.org/10.1016/j.watres.2012.12.035.

(30) Hrudey, S. E. Chlorination Disinfection By-Products, Public Health Risk Tradeoffs and Me. *Water Res*. **2009**, 43 (8), 2057–2092. https://doi.org/10.1016/j.watres.2009.02.011.

(31) *USEPA: National Primary Drinking Water Regulations* https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations#Byproducts (Accessed Jul 7, 2023).

(32) USEPA. *Fact Sheet Fifth Contaminant Candidate List (CCL 5)*, 2022. https://www.epa.gov/system/files/documents/2022-10/Fact%20Sheet%20Final%20Fifth%20Contaminant%20Candidate%20List%20%28CCL%205%29.pdf (Accessed July 26, 2023)

(33) Richardson, S. D.; Plewa, M. J.; Wagner, E. D.; Schoeny, R.; Demarini, D. M. Occurrence , Genotoxicity , and Carcinogenicity of Regulated and Emerging Disinfection by-Products in Drinking Water : A Review and Roadmap for Research. *Mutat. Res. - Rev. Mutat.* **2007**. 636, 178–242. https://doi.org/10.1016/j.mrrev.2007.09.001.

(34) Richardson, S. D.; Kimura, S. Y. Emerging Environmental Contaminants: Challenges Facing Our next Generation and Potential Engineering Solutions. *Environ. Technol. Innov*. **2017**, 8, 40–56. https://doi.org/10.1016/j.eti.2017.04.002.

(35) Mitch, W. A.; Richardson, S. D.; Zhang, X.; Gonsior, M. High-Molecular-Weight by-Products of Chlorine Disinfection. *Nat. Water* **2023**, 1, 336-347. https://doi.org/10.1038/s44221-023-00064-x.

(36) Yang, M.; Liberatore, H. K.; Zhang, X. Current Methods for Analyzing Drinking Water Disinfection Byproducts. *Curr. Opin. Environ. Sci. Heal*. **2019**, 7, 98–107. https://doi.org/10.1016/j.coesh.2018.12.006.

(37) Wawryk, N. J. P.; Craven, C. B.; Blackstock, L. K. J.; Li, X. F. New Methods for Identification of Disinfection Byproducts of Toxicological Relevance: Progress and Future Directions. *J. Environ. Sci. (China)* **2021**, 99, 151–159. https://doi.org/10.1016/j.jes.2020.06.020.

(38) Andrade-Eiroa, A.; Canle, M.; Leroy-Cancellieri, V.; Cerdà, V. Solid-Phase Extraction of Organic Compounds: A Critical Review. Part II. *Trends Anal. Chem*. **2016**, 80, 655–667. https://doi.org/10.1016/j.trac.2015.08.014.

(39) Marczak, M.; Wolska, L.; Chrzanowski, W.; Namieśnik, J. Microanalysis of Volatile Organic Compounds (VOCs) in Water Samples - Methods and Instruments. *Microchim. Acta* **2006**, 155, 331–348. https://doi.org/10.1007/s00604-006-0630-x.

(40) Kimura, S.Y.; Zheng, W.; Hipp, T.N.; Allen, J.M.; Richardson, S.D. Total Organic Halogen (TOX) in Human Urine: A Halogen-Specific Method for Human Exposure Studies. *J. Environ. Sci. (China)* **2017**, 58, 285–295. https://doi.org/10.1016/j.jes.2017.04.008.

(41) Tang, Y.; Xu, Y.; Li, F.; Jmaiff, L.; Hrudey, S. E.; Li, X. F. Nontargeted Identification of Peptides and Disinfection Byproducts in Water. *J. Environ. Sci. (China)* **2016,** 42, 259–266. https://doi.org/10.1016/j.jes.2015.08.007.

(42) Zhang, H.; Yang, M. Characterization of Brominated Disinfection Byproducts Formed during Chloramination of Fulvic Acid in the Presence of Bromide. *Sci. Total Environ*. **2018**, 627, 118–124. https://doi.org/10.1016/j.scitotenv.2018.01.215.

(43) Zhou, R.; Xu, Z.; Zhu, J.; Liu, W.; Meng, Y.; Zhu, P.; Zhou, W.; Huang, C.; Ding, X. Determination of 10 Haloacetamides in Drinking Water by Gas Chromatography with Automated Solid Phase Extraction. *J. Chromatogr. B*. **2020**, 1150 (10), 122191. https://doi.org/10.1016/j.jchromb.2020.122191.

(44) Hu, S.; Gong, T.; Ma, J.; Tao, Y.; Xian, Q. Simultaneous Determination of Iodinated Haloacetic Acids and Aromatic Iodinated Disinfection Byproducts in Waters with a New SPE-

HPLC-MS/MS Method. *Chemosphere* **2018**, 198, 147–153. https://doi.org/10.1016/j.chemosphere.2018.01.124.

(45) Yu, Y.; Reckhow, D. A. Formation and Occurrence of N-Chloro-2,2-Dichloroacetamide, a Previously Overlooked Nitrogenous Disinfection Byproduct in Chlorinated Drinking Waters. *Environ. Sci. Technol*. **2017**, 51 (3), 1488–1497. https://doi.org/10.1021/acs.est.6b04218.

(46) Liu, Z.; Craven, C. B.; Huang, G.; Jiang, P.; Wu, D.; Li, X. F. Stable Isotopic Labeling and Nontarget Identification of Nanogram/Liter Amino Contaminants in Water. Anal. Chem. 2019, 91 (20), 13213–13221. https://doi.org/10.1021/acs.analchem.9b03642.

(47) Cuthbertson, A. A.; Bach, C.; Richardson, S. D.; Dauchy, X. A Novel Automated Method for the Quantification of Ten Halobenzoquinones in Drinking Water Using Online Solid-Phase Extraction Coupled with Liquid Chromatography Tandem Mass Spectrometry. *J. Chromatogr. A* **2020**, 1612, 460642. https://doi.org/10.1016/j.chroma.2019.460642.

(48) Huang, G.; Jiang, P.; Li, X. F. Mass Spectrometry Identification of N-Chlorinated Dipeptides in Drinking Water. *Anal. Chem*. **2017**, 89 (7), 4204–4209. https://doi.org/10.1021/acs.analchem.7b00228.

(49) Lloyd R. Snyder, Joseph J. Kirkland, J. W. D. *Introduction to Modern Liquid Chromatography*, Third.; John Wiley & Sons, Inc: Hoboken, NJ, USA, **2011**.

(50) Buszewski, B.; Noga, S. Hydrophilic Interaction Liquid Chromatography (HILIC)-a Powerful Separation Technique. *Anal. Bioanal. Chem*. **2012**, 402 (1), 231–247. https://doi.org/10.1007/s00216-011-5308-5.

(51) Xia, D.; Liu, H.; Lu, Y.; Liu, Y.; Liang, J.; Xie, D.; Lu, G.; Qiu, J.; Wang, R. Utility of a Non-Target Screening Method to Explore the Chlorination of Similar Sulfonamide Antibiotics: Pathways and NCl Intermediates. *Sci. Total Environ*. **2023**, 858, 160042. https://doi.org/10.1016/J.SCITOTENV.2022.160042.

(52) Li, W.; Zhang, X.; Han, J. Formation of Larger Molecular Weight Disinfection Byproducts from Acetaminophen in Chlorine Disinfection. *Environ. Sci. Technol*. **2022**, 56 (23), 16929–16939. https://doi.org/10.1021/acs.est.2c06394.

(53) Neuwald, I. J.; Hübner, D.; Wiegand, H. L.; Valkov, V.; Borchers, U.; Nödler, K.; Scheurer, M.; Hale, S. E.; Arp, H. P. H.; Zahn, D. Occurrence, Distribution, and Environmental Behavior of Persistent, Mobile, and Toxic (PMT) and Very Persistent and Very Mobile (vPvM) Substances in the Sources of German Drinking Water. *Environ. Sci. Technol*. **2022**, 56 (15), 10857–10867. https://doi.org/10.1021/acs.est.2c03659.

(54) Qiu, J.; Craven, C.; Wawryk, N.; Carroll, K.; Li, X.-F. Integration of Solid Phase Extraction with HILIC-MS/MS for Analysis of Free Amino Acids in Source Water. *J. Environ. Sci. (China)* **2022**, 117, 190-196. https://doi.org/10.1016/j.jes.2022.04.025.

(55) Pellicer-Castell, E.; Belenguer-Sapiña, C.; Amorós, P.; El Haskouri, J.; Herrero-Martínez, J. M.; Mauri-Aucejo, A. R. Mesoporous Silica Sorbent with Gold Nanoparticles for Solid-Phase Extraction of Organochlorine Pesticides in Water Samples. *J. Chromatogr. A* **2022**, 1662, 462729. https://doi.org/10.1016/j.chroma.2021.462729.

(56) Domingues, J. T.; Orlando, R. M.; Almeida, M. R.; de Lemos, L. R.; Mageste, A. B.; Rodrigues, G. D. Extraction of Estrogen Hormones from Water Samples Using an Aqueous Two-Phase System: A New Approach for Sample Preparation in the Analysis of Emerging

Contaminants. *Microchem. J.* **2021**, 166 (March), 106231. https://doi.org/10.1016/j.microc.2021.106231.

(57) 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.

(58) Dator, R.; von Weymarn, L. B.; Villalta, P. W.; Hooyman, C. J.; Maertens, L. A.; Upadhyaya, P.; Murphy, S. E.; Balbo, S. In Vivo Stable-Isotope Labeling and Mass-Spectrometry-Based Metabolic Profiling of a Potent Tobacco-Specific Carcinogen in Rats. *Anal. Chem*. **2018**, 90 (20), 11863–11872. https://doi.org/10.1021/acs.analchem.8b01881.

(59) Yuan, M.; Breitkopf, S. B.; Yang, X.; Asara, J. M. A Positive/Negative Ion-Switching, Targeted Mass Spectrometry-Based Metabolomics Platform for Bodily Fluids, Cells, and Fresh and Fixed Tissue. *Nat. Protoc*. **2012**, 7 (5), 872–881. https://doi.org/10.1038/nprot.2012.024.

(60) Jennings, E. K.; Sierra Olea, M.; Kaesler, J. M.; Hübner, U.; Reemtsma, T.; Lechtenfeld, O. J. Stable Isotope Labeling for Detection of Ozonation Byproducts in Effluent Organic Matter with FT-ICR-MS. *Water Res*. **2023**, 229, 119477. https://doi.org/10.1016/j.watres.2022.119477.

(61) Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal. Chem*. **2021**, 93 (34), 11692–11700. https://doi.org/10.1021/acs.analchem.1c01465.

(62) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag Relaunched: Incorporating Strategies beyond in Silico Fragmentation. *J. Cheminform*. **2016**, 8 (3), 1–16. https://doi.org/10.1186/s13321-016-0115-9.

(63) 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.

(64) 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.

(65) 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.

(66) Pang, Z.; Zhou, G.; Ewald, J.; Chang, L.; Hacariz, O.; Basu, N.; Xia, J. Using MetaboAnalyst 5.0 for LC–HRMS Spectra Processing, Multi-Omics Integration and Covariate Adjustment of Global Metabolomics Data. *Nat. Protoc*. **2022**, 17 (8), 1735–1761. https://doi.org/10.1038/s41596-022-00710-w.

(67) Avtonomov, D. M.; Raskind, A.; Nesvizhskii, A. I. BatMass: A Java Software Platform for LC-MS Data Visualization in Proteomics and Metabolomics. *J. Proteome Res*. **2016**, 15 (8), 2500–2509. https://doi.org/10.1021/acs.jproteome.6b00021.

(68) 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.

(69) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H. C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat. Methods* **2016**, 13 (9), 741–748. https://doi.org/10.1038/nmeth.3959.

(70) Helmus, R.; ter Laak, T. L.; van Wezel, A. P.; de Voogt, P.; Schymanski, E. L. patRoon: Open Source Software Platform for Environmental Mass Spectrometry Based Non-Target Screening. *J. Cheminform*. **2021**, 13 (1), 1–25. https://doi.org/10.1186/s13321-020-00477-w.

(71) Helmus, R.; van de Velde, B.; Brunner, A. M.; ter Laak, T. L.; van Wezel, A. P.; Schymanski, E. L. patRoon 2.0: Improved Non-Target Analysis Workflows Including Automated

Transformation Product Screening. *J. Open Source Softw*. **2022**, 7 (71), 4029. https://doi.org/10.21105/joss.04029.

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

(73) Smith, R.A.; Zhang, Q. Peak Pair Pruner: a post-processing software to MS-DIAL for peak pair validation and ratio quantification of isotopic labeling LC-MS(/MS) data. *Adv. Bioinformatics* **2023**. 3 (1), 1-3. https://doi.org/10.1093/bioadv/vbad044.

(74) Guo, J.; Huan, T. Mechanistic Understanding of the Discrepancies between Common Peak Picking Algorithms in Liquid Chromatography-Mass Spectrometry-Based Metabolomics. *Anal. Chem*. **2022**. 95 (14), 5894–5902. https://doi.org/10.1021/acs.analchem.2c04887.

Chapter Two

Non-targeted Analysis of Organics in Water: Challenges of Manual Data Analysis and Development of a New Bioinformatic Platform

2.1: Introduction

Water disinfection is a necessity for the prevention of infection and disease, and this practice has virtually eliminated severe waterborne illness in developed countries.[1] However, a consequence of water disinfection is the formation of disinfection-byproducts (DBPs) from the reaction of natural organic matter (NOM) and disinfectants. Of particular concern are Nitrogenous DBPs (N-DBPs), which have higher toxicity than their corresponding Carbonaceous DBPs (C-DBPs).[2,3] Several studies have shown that N-DBPs are the toxicity drivers in treated water, but are currently not regulated.[2,4,5] Some classes of N-DBPs include haloacetonitriles (HANs), haloacetamides (HAMs), and halonitromethanes (HNMs). However, the majority of N-DBPs remain unidentified, hampering exposure and risk assessments as well as future regulatory improvement. One approach to limit N-DBP formation is to identify their precursors present in source water and improve removal efficiency. By identifying precursors, we may design processes to efficiently remove them during water treatment. DBP precursor removal is an effective strategy for managing DBP formation.

*Parts of Chapter 2 for the development of HDPairFinder and Figure 2.1 were published in Zhao, T.; Carroll, K.; Craven, C. B.; Wawryk, N. J. P.; Xing, S.; Guo, J.; Li, X.; Huan, T. HDPairFinder : A Data Processing Platform for Hydrogen / Deuterium Isotopic Labeling-Based Nontargeted Analysis of Trace-Level Amino-Containing Chemicals in Environmental Water. J. Environ. Sci. 2024, 136, 583–593. Reprinted with permission. Copyright 2023 Elsevier.

N-DBPs are formed from organic compounds that contain reactive nitrogen functional groups. These nitrogen containing precursors can be easily halogenated and transformed during water disinfection treatment (e.g. chlorination and chloramination).[6–10] Additionally, nitrogen containing precursors can be highly abundant and come from various sources. One such example is amino acids. A previous study by our group analyzed individual amino acids in the North Saskatchewan River and showed that they could reach concentrations up to 5.5 μg/L.[11] Other studies have shown total free amino acid concentrations as high as 26 μg/L and 30 μg/L in various source waters.[12,13] Other examples of N-containing compounds include pharmaceuticals and personal care products (PPCPs), many of which can form N-DBPs.[14,15] For example, studies have shown acetaminophen, a common pain-killer, has been detected in source water and can become halogenated during drinking water treatment.[16,17]

A common technique for detection and identification of precursors and DBPs is non-target analysis using mass spectrometry (MS). Because the majority of DBPs and their precursors remain unidentified, non-target analysis is necessary for comprehensive characterization. However, non-targeted analysis of a single source water sample can detect thousands of chemical features.[18] This creates significant challenges when we try to analyze and sort through data manually. To overcome this difficulty, we may apply a prioritization strategy at the beginning of analysis to capture a specific class of compounds. Focusing on a specific class of compounds such as nitrogen-containing precursors, we may achieve comprehensive characterization of these compounds in a sample.

Several different approaches currently exist for non-target data analysis. One such method is suspect screening, which uses a predefined list of compounds to compare to non-target data and looks for matches.[19] While screening can prioritize data, it may be missing important

peaks not on the pre-defined list.[20] The different data processing methods can be classified into either top-down or bottom-up approaches.[21] Top-down methods will create a list of candidates based on MS1 spectra and bottom-up approaches will start with the MS2 spectra to work towards a formula.[21] Each approach comes with limitations. Top-down approaches lack an ability to prioritize data while bottom-up may be missing important features that do not have MS2 library matches. Additionally, different research teams vary when applying various approaches for processing non-targeted data. Schymanski *et al.* compared 17 different research groups on how they used analysis techniques for a single sample. Among all the groups, discrepancies in identification of chemical features were found, which were likely because of a lack of harmonization of instruments and data analysis procedures.[20] With an effort to reduce these discrepancies, it would be helpful to develop a specific data processing procedure alongside a method.

To specifically detect nitrogen containing organics in source water, I applied a previously developed stable isotopic labelling method followed by high-performance liquid chromatography (HPLC)- high resolution mass spectrometry (HRMS) analysis. This stable isotopic labelling method uses two reagents, formaldehyde and deuterated formaldehyde, to methylate reactive amine functional groups of nitrogen-containing compounds. The isotopic pattern between hydrogen and deuterium is used to identify the reactive nitrogen compounds in the HRMS data.[18] However, a challenge of this method is the data analysis and identification of compounds in the chemical features detected using non-targeted analysis. Data processing is complex and time-consuming. Therefore, I aimed to develop a bioinformatic program to improve the processing of non-targeted data of source water samples. Through the development of an automated data processing program, we aimed to provide comprehensive characterization of N-containing

compounds in source water. By identifying N-containing precursors in source water, we can gain a better understanding of what precursors that may form NDBPs during disinfection.

2.2: Experimental

2.2.1: Chemicals and Reagents

Formaldehyde (CH_2O, 37 wt. % in H_2O, contains 10−15 % methanol as stabilizer), deuterated formaldehyde (CD_2O, ~20 wt. % in D_2O, 98 atom % D,), sodium cyanoborohydride ($NaBH_3CN$, 95 %), formic acid (FA, 99 %), and nylon disk filters (0.45 μm) were obtained from Sigma-Aldrich (St. Louis, MO). Optima water, methanol, acetonitrile (I), ammonium hydroxide (30% wt.), and glass microfiber filters (1.5 μm) were purchased from Fisher Scientific (Fair Lawn, NJ). Oasis MCX cartridges (6 mL, 150 mg of sorbent) were obtained from Waters (Milford, MA). Syringe filters (0.45 μm, PVDF) were obtained from Dikma (Markham, ON).

2.2.2: Sample Collection and Stable Isotopic Labelling Reaction

Authentic source water samples were collected from the water treatment plant E.L. Smith (ELS), in Edmonton. Samples were collected on March 21st, 2022 and March 24th, 2022. Bottles were rinsed three times and samples were collected and capped with no headspace. Samples were filtered using 1.5 μm glass microfiber filters, followed by 0.45 μm nylon membrane filters and stored at 4 °C before analysis.

Stable isotopic labelling of the water sample was performed following the previously published method.[18] The procedure of the stable isotopic labelling reaction is presented in **Figure 2.1**. [22] Solutions of CH_2O (1.8 M), CD_2O (1.8 M), and $NaBH_3CN$ (0.6 M) were prepared separately, immediately before each reaction. For the labelling reaction, a source water sample (2 L) is split into two portions of 1 L. To one part, CH_2O is added to reach a final concentration of 3.6 mM and $NaBH_3CN$ to 1.2 mM. To the other part, the same concentrations of CD_2O (3.6

mM) and $NaBH_3CN$ (1.2 mM) were used. The reaction was left stirring for four hours, after which 2 mL of formic acid was added.

SPE was performed using Oasis MCX cartridges on a Supelco vacuum manifold. Cartridges were rinsed with methanol (2 mL) followed by water (4 mL, 0.2 % FA v/v) before sample loading. Samples (1 L) were passed through at a rate ~2-3 mL/min. Following the sample, the cartridge was rinsed with water (2 mL, 0.2 % FA v/v), and eluted with ammonium hydroxide solution (10 mL, 5 % wt. in methanol) to collect the labelled compounds. Eluent was concentrated to 0.1 mL under nitrogen stream. Extracts of light (hydrogen) and heavy (deuterium) labelled samples, each ~0.1 mL, were mixed and filtered using a 0.45 µm PVDF syringe filter. Finally, they were analyzed using HPLC-HRMS.

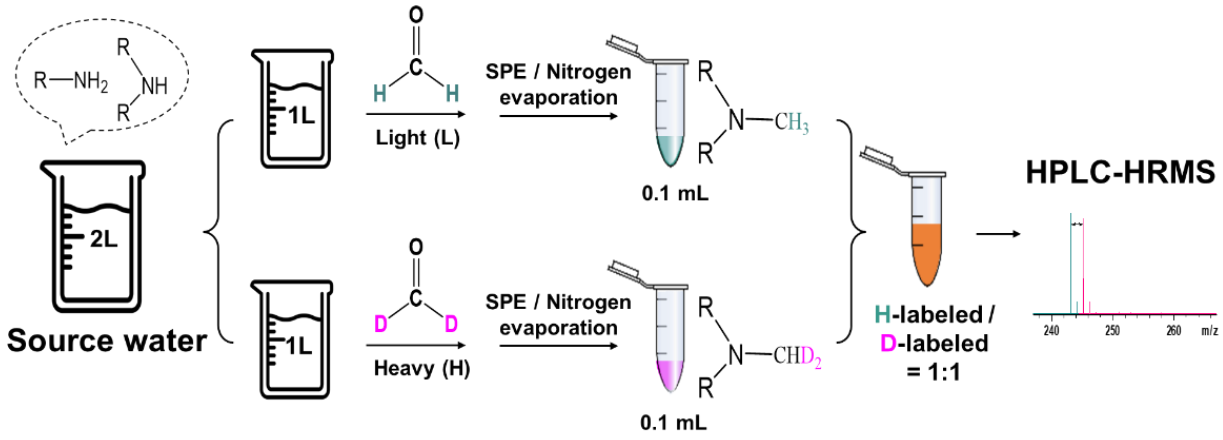


Figure 2.1- Overview of the Labelling Method Procedure for a Source Water Sample.

2.2.3: HPLC Separation Conditions

Separations were performed on an Agilent 1260 Series HPLC system, with a Luna C18 column (100 mm × 2 mm × 3 µm pore size). The injection volume was 20 µL. Mobile phases A and B were prepared as H_2O/ ACN (95:5 v/v, 0.1 % FA) and ACN (0.1 % FA), respectively. The flow rate was set at 80 µL/min with a gradient elution as follows: 0 % B for 10 minutes, 0-30 %

B over 20 minutes, 30-90 % B over 15 minutes, hold at 90 % for 10 minutes, decrease to 0 % B over 0.1 min, hold at 0 % B for 4.9 minutes. The column temperature was maintained at 30 °C.

2.2.4: High Resolution Mass Spectrometric (HRMS) Conditions

Analysis was complete with a quadrupole time-of-flight mass spectrometer (Sciex QTOF x500R). The mass spectrometer was set to positive mode with an ion spray voltage of 5500 V. Other conditions are as follows: source gas 1(N_2, 35 arbitrary units), source gas 2 (N_2, 40 arbitrary units), curtain gas (N_2, 30 arbitrary units), temperature (500 °C), declustering potential (DP, 100 V), and collision energy (CE, 10 V). For the full scan, mass scan ranges from 50-1000 Da with an accumulation time of 0.25 s. For information dependent analysis (IDA), a threshold intensity of 1000 cps was used to trigger MS/MS collection, and a maximum of 10 candidate ions monitored per cycle. The MS/MS mass range scan was from 20-1000 Da. Typical instrument resolution is >20,000 (at full width half-height) and mass accuracy <5 ppm.

2.2.5: Data Analysis

Data was collected using Sciex OS software. Peak picking was performed using the Analytics tool in Sciex OS. The parameters used for peak are listed in **Table 2.1**. Manual hydrogen/deuterium (H/D) feature pair picking was done using excel. Automated H/D feature pair picking was performed using HDPairFinder that is run using R version 4.2.1 (https://www.r-project.org/).

Table 2.1- Analytics Parameters Chosen for Peak Picking.

Parameter	*Value*
Library Search	Smart Confirmation Search, All Libraries
Precursor Mass Tolerance	± 0.4 Da
Collision Energy	± 5 eV
Retention Time	NA
Use Polarity	0.05 Intensity Threshold
Use Collision Energy Spread	10 % minimal purity
Minimum Retention Time	3 min
Peak Detection Sensitivity	5/7 Exhaustive

2.3: Results and Discussion

2.3.1: Manual Data Analysis for HRMS data and Challenges

Non-target analysis generates large amounts of data. Studies have shown thousands of features, or a unique peak in the non-target data, from a single sample.[23,24] Without prioritization strategies, picking significant features for identification is difficult.[25] For example, **Figures 2.2** and **2.3** show all the labelled and unlabelled features for two different source water samples. For both samples, which were collected on two different dates, thousands of features are detected. To try and identify all these features would be challenging and time consuming. Thus, it is important to apply a strategy to prioritize the data. By applying the labelling reaction, I can focus on amine containing organic compounds. The mass difference between the hydrogen and deuterium ions will create specific patterns in the HRMS data. Two peaks that follow this isotopic pattern are identified as hydrogen/deuterium (H/D) feature pairs for a unique N-containing compound. H/D

feature pairs are a priority for identification as N-DBP precursors. Initially, I completed prioritization manually.

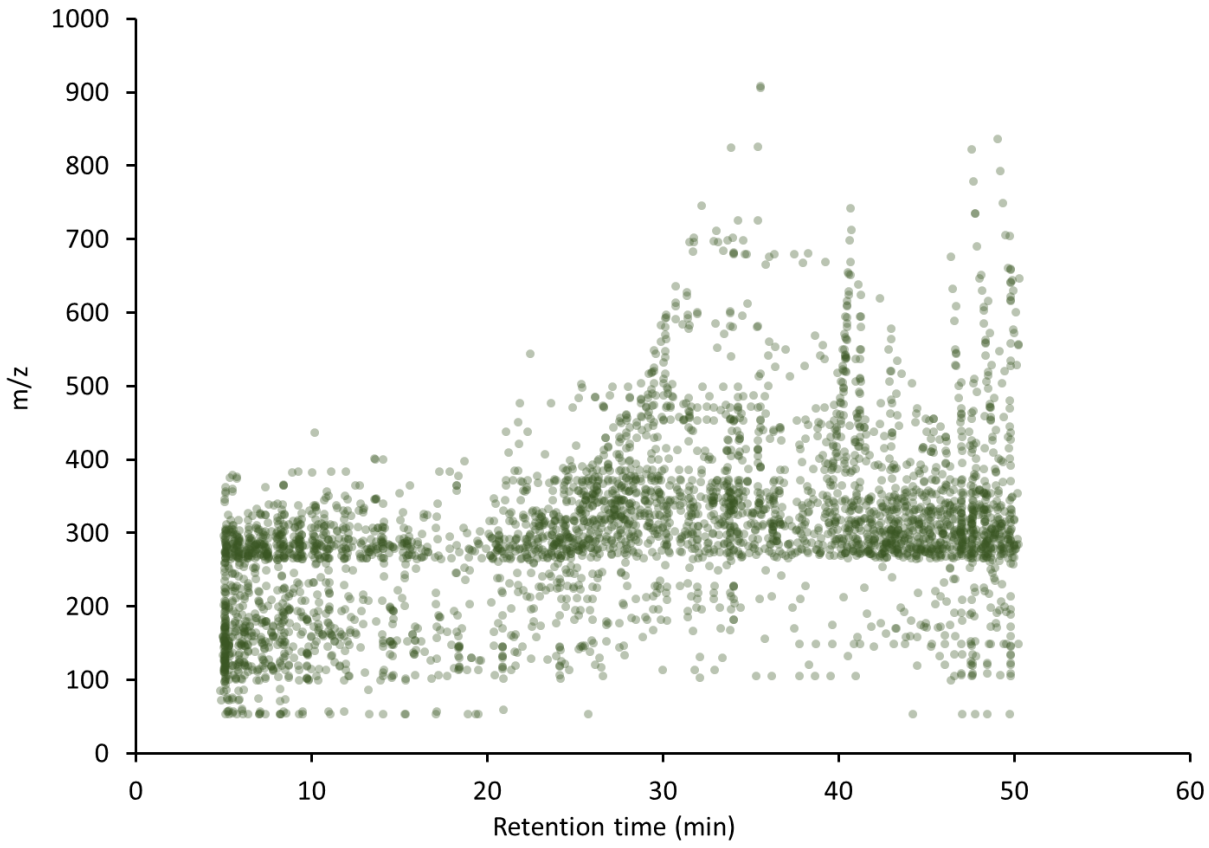


Figure 2.2 - Total features detected from ELS on March 21st, 2022. Retention time vs m/z of unprioritized data shows a total of 3895 features.

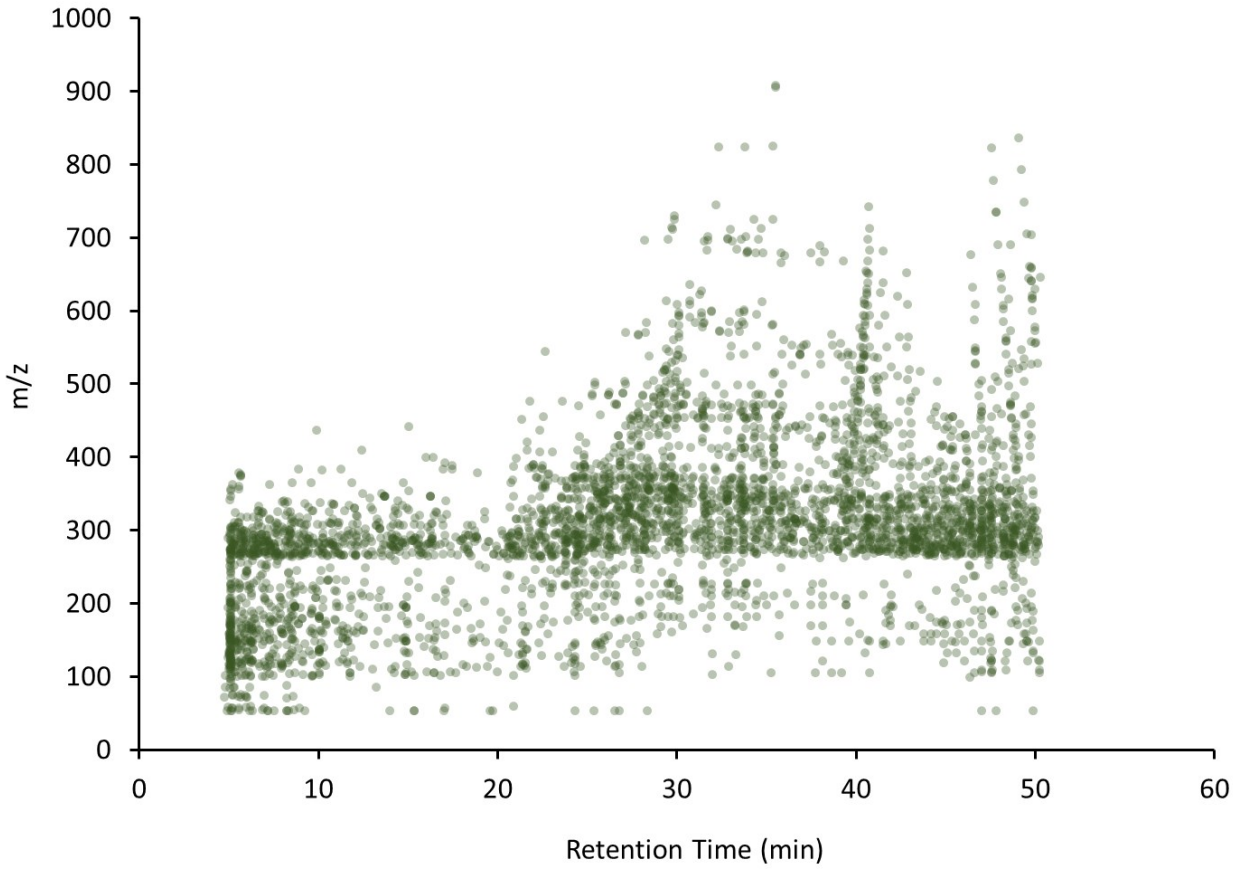


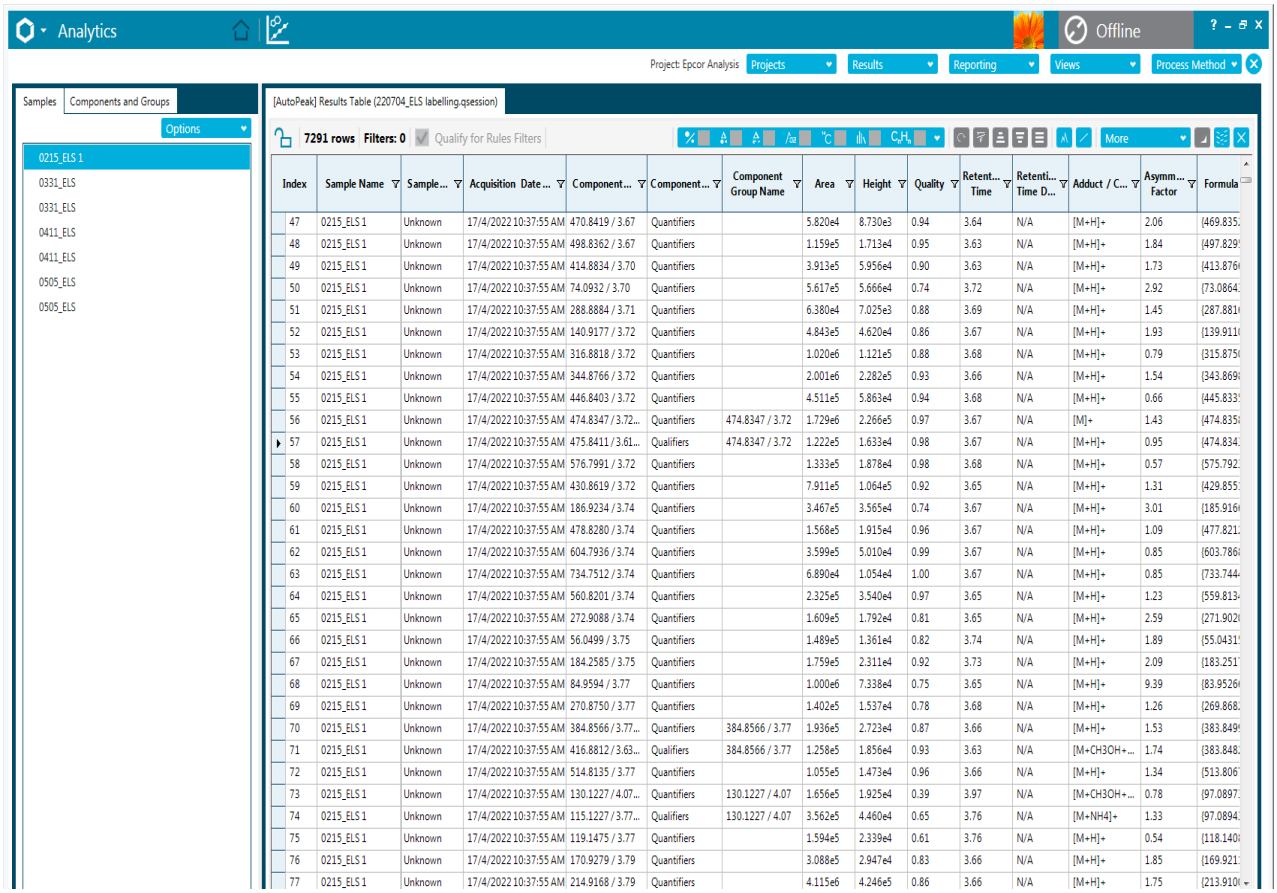
Figure 2.3 - Total features detected from ELS on March 24th, 2022. Retention time vs m/z of unprioritized data shows a total of 4181 features.

Manual searching for labelled H/D feature pairs uses Sciex OS software and excel.[18] An overview of the manual data analysis method is presented in **Figure 2.4**. First, a list of all the peaks (labelled and unlabelled) is generated using the built-in software, Analytics, in Sciex OS. The precursor mass and retention time columns are copied into an excel spreadsheet. The feature table is first sorted by increasing retention time. The precursor mass column is then copied five times to create six total columns (A through E). Each adjacent column is offset by one row. For example, the value in A2 would also be in B3, C4, D5, E6, and F7. To find the isotopic pattern between peaks, the precursor mass columns are subtracted from each other (i.e. A-B, A-C, A-D,

A-E, and A-F for each row). The absolute value of the differences is taken, and the sheet is filtered to look for the specific value ranges (Δm/z 2.012-2.013, 4.024-4.026, 6.036-6.039,8.048-8.052). This creates a list of precursor mass pairs that show the isotopic labelling pattern. Because the feature table is sorted by increasing retention time, this method assumes that the retention time of these two precursor masses is similar because they are within five rows in the table of peaks. When identified, these H/D feature pairs of precursor masses are copied into a new sheet. These H/D feature pairs could then be searched within Sciex OS software to ensure they have similar peak shapes and to confirm retention times. When applied to more than one sample, each sample needs separate analysis before manually comparing the two H/D feature pair lists.

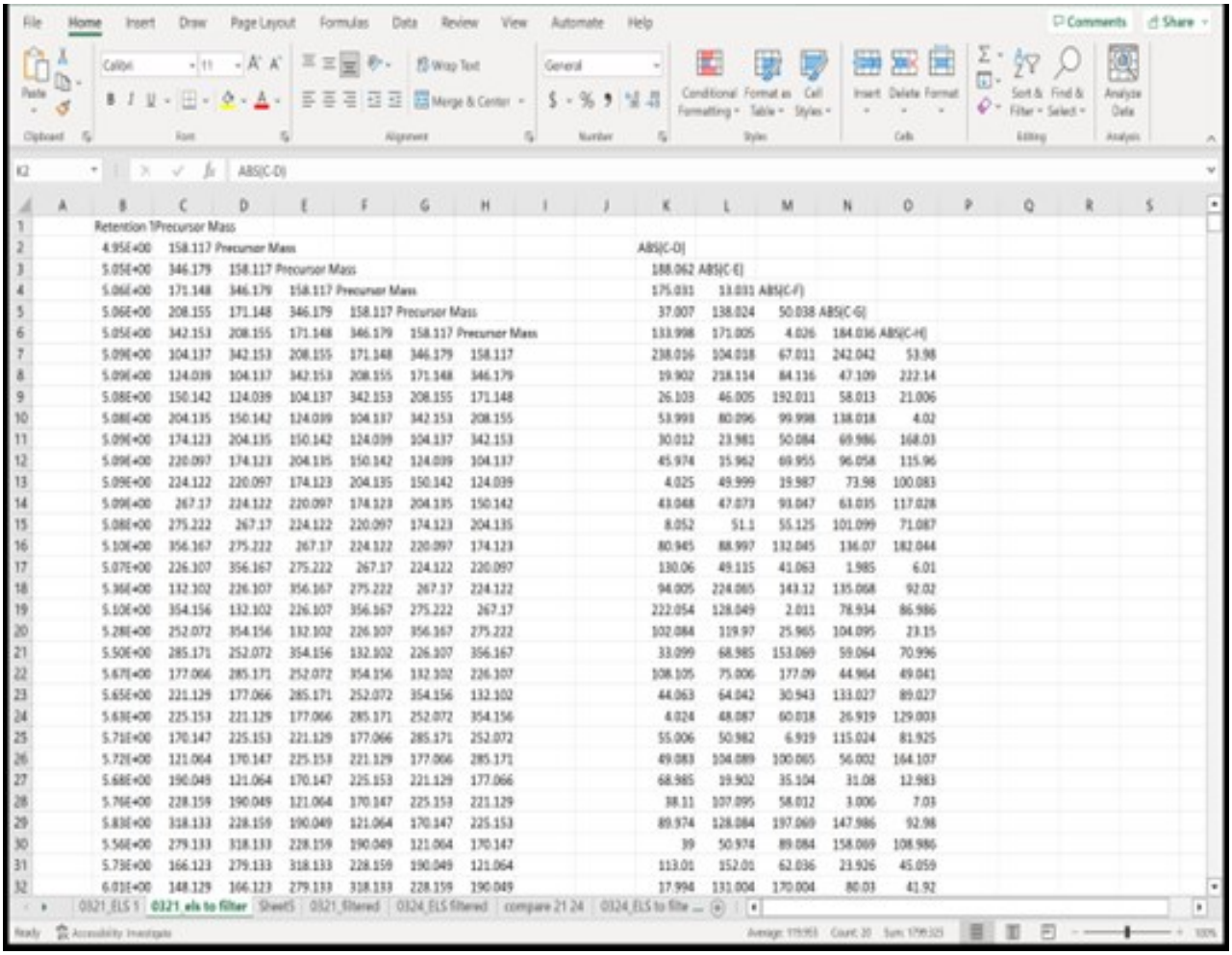
Figure 2.5 and **Figure 2.6** use March 21st and March 24th samples respectively, to show how prioritizing data will reduce the number of chemical features. In both cases, features are reduced from thousands to ~60. However, H/D feature pairs are only searched in rows in the dataset that are within five places of one another. Thus, any H/D feature pair outside of these limits is missed. Additionally, because the manual data analysis is time consuming, it is challenging to process large datasets. Therefore, I investigated automated data processing for H/D labelled non-targeted analysis data. Various programs, including patRoon, ISFrag, and OpenMS, can process non-target data.[26–28] These programs can search isotopic patterns, such as Chlorine$^{35/37}$ and Carbon$^{12/13}$. However, they can not be applied for processing the stable isotopic labelling data we obtained. Automating the search for the stable isotopic labelling pattern would streamline the analysis workflow. Therefore, I pursued the development of a program specifically for the stable isotopic labelling method.

Step 1: Generate a list of peaks using SciexOS built-in Analytics.



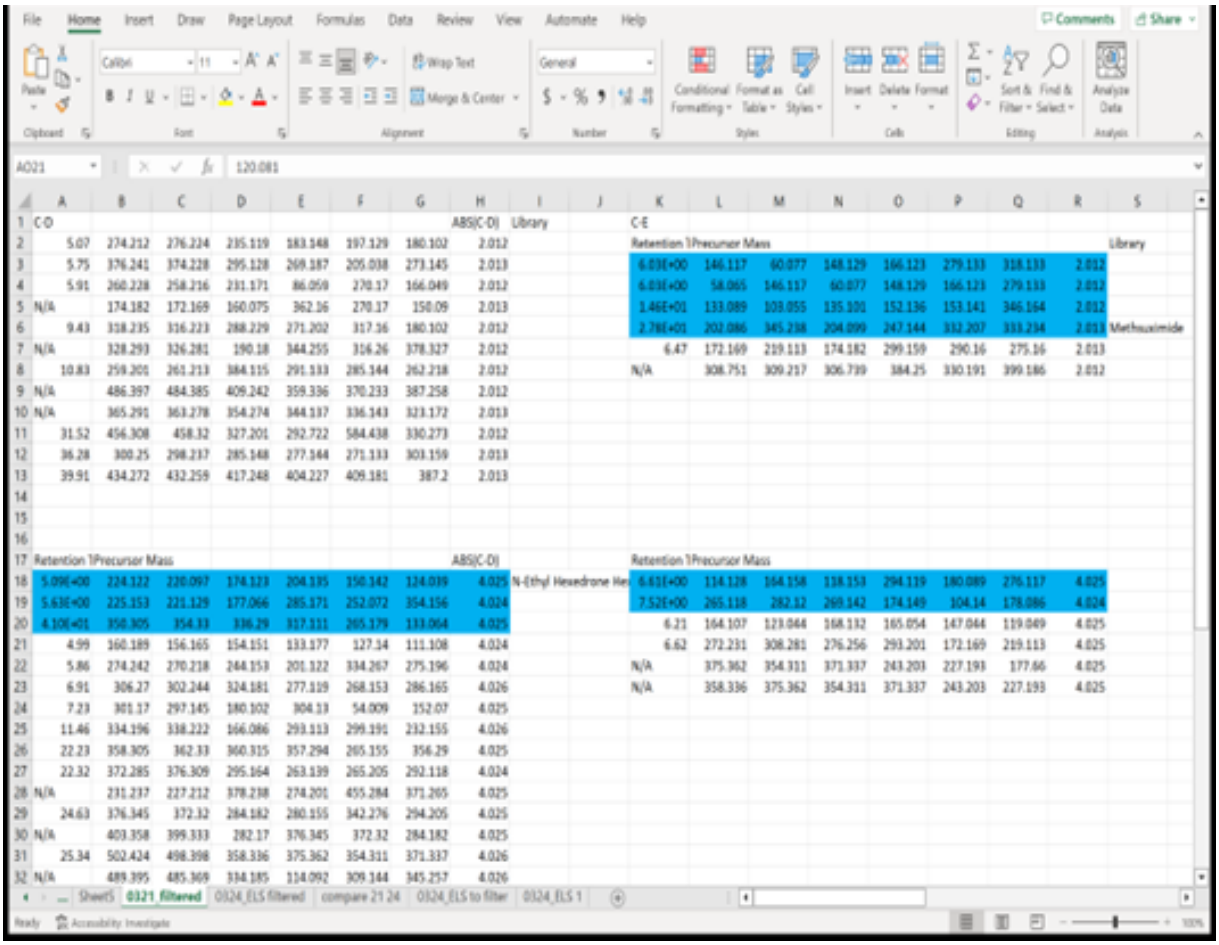
Index	Sample Name	Sample...	Acquisition Date...	Component...	Component...	Component Group Name	Area	Height	Quality	Retent... Time	Retenti... Time D...	Adduct / C...	Asymm... Factor	Formula
47	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	470.8419 / 3.67	Quantifiers		5.820e4	8.730e3	0.94	3.64	N/A	[M+H]+	2.06	{469.835
48	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	498.8362 / 3.67	Quantifiers		1.159e5	1.713e4	0.95	3.63	N/A	[M+H]+	1.84	{497.829
49	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	414.8834 / 3.70	Quantifiers		3.913e5	5.956e4	0.90	3.63	N/A	[M+H]+	1.73	{413.876
50	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	74.0932 / 3.70	Quantifiers		5.617e5	5.666e4	0.74	3.72	N/A	[M+H]+	2.92	{73.0864
51	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	288.8884 / 3.71	Quantifiers		6.380e4	7.025e3	0.88	3.69	N/A	[M+H]+	1.45	{287.881
52	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	140.9177 / 3.72	Quantifiers		4.843e5	4.620e4	0.86	3.67	N/A	[M+H]+	1.93	{139.911
53	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	316.8818 / 3.72	Quantifiers		1.020e6	1.121e5	0.88	3.68	N/A	[M+H]+	0.79	{315.875
54	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	344.8766 / 3.72	Quantifiers		2.001e6	2.282e5	0.93	3.66	N/A	[M+H]+	1.54	{343.869
55	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	446.8403 / 3.72	Quantifiers		4.511e5	5.863e4	0.94	3.68	N/A	[M+H]+	0.66	{445.833
56	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	474.8347 / 3.72...	Quantifiers	474.8347 / 3.72	1.729e6	2.266e5	0.97	3.67	N/A	[M]+	1.43	{474.835
57	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	475.8411 / 3.61...	Qualifiers	474.8347 / 3.72	1.222e5	1.633e4	0.98	3.67	N/A	[M+H]+	0.95	{474.834
58	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	576.7991 / 3.72	Quantifiers		1.333e5	1.878e4	0.98	3.68	N/A	[M+H]+	0.57	{575.792
59	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	430.8619 / 3.72	Quantifiers		7.911e5	1.064e5	0.92	3.65	N/A	[M+H]+	1.31	{429.855
60	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	186.9234 / 3.74	Quantifiers		3.467e5	3.565e4	0.74	3.67	N/A	[M+H]+	3.01	{185.916
61	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	478.8280 / 3.74	Quantifiers		1.568e5	1.915e4	0.96	3.67	N/A	[M+H]+	1.09	{477.821
62	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	604.7936 / 3.74	Quantifiers		3.599e5	5.010e4	0.99	3.67	N/A	[M+H]+	0.85	{603.786
63	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	734.7512 / 3.74	Quantifiers		6.890e4	1.054e4	1.00	3.67	N/A	[M+H]+	0.85	{733.744
64	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	560.8201 / 3.74	Quantifiers		2.325e5	3.540e4	0.97	3.65	N/A	[M+H]+	1.23	{559.813
65	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	272.9088 / 3.74	Quantifiers		1.609e5	1.792e4	0.81	3.65	N/A	[M+H]+	2.59	{271.902
66	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	56.0499 / 3.75	Quantifiers		1.489e5	1.361e4	0.82	3.74	N/A	[M+H]+	1.89	{55.0431
67	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	184.2585 / 3.75	Quantifiers		1.759e5	2.311e4	0.92	3.73	N/A	[M+H]+	2.09	{183.251
68	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	84.9594 / 3.77	Quantifiers		1.000e6	7.338e4	0.75	3.65	N/A	[M+H]+	9.39	{83.9526
69	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	270.8750 / 3.77	Quantifiers		1.402e5	1.537e4	0.78	3.68	N/A	[M+H]+	1.26	{269.868
70	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	384.8566 / 3.77...	Quantifiers	384.8566 / 3.77	1.936e5	2.723e4	0.87	3.66	N/A	[M+H]+	1.53	{383.849
71	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	416.8812 / 3.63...	Qualifiers	384.8566 / 3.77	1.258e5	1.856e4	0.93	3.63	N/A	[M+CH3OH+...	1.74	{383.848
72	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	514.8135 / 3.77	Quantifiers		1.055e5	1.473e4	0.96	3.66	N/A	[M+H]+	1.34	{513.806
73	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	130.1227 / 4.07...	Quantifiers	130.1227 / 4.07	1.656e5	1.925e4	0.39	3.97	N/A	[M+CH3OH+...	0.78	{97.0897
74	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	115.1227 / 3.77...	Qualifiers	130.1227 / 4.07	3.562e5	4.460e4	0.65	3.76	N/A	[M+NH4]+	1.33	{97.0894
75	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	119.1475 / 3.77	Quantifiers		1.594e5	2.339e4	0.61	3.76	N/A	[M+H]+	0.54	{118.140
76	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	170.9279 / 3.79	Quantifiers		3.088e5	2.947e4	0.83	3.66	N/A	[M+H]+	1.85	{169.921
77	0215_ELS 1	Unknown	17/4/2022 10:37:55 AM	214.9168 / 3.79	Quantifiers		4.115e6	4.246e5	0.86	3.66	N/A	[M+H]+	1.75	{213.910

Step 2: Copy retention time and precursor mass columns into excel spreadsheet. Copy precursor mass into five new columns, offsetting the row by one each time. Calculate the differences between the precursor mass columns for each row.

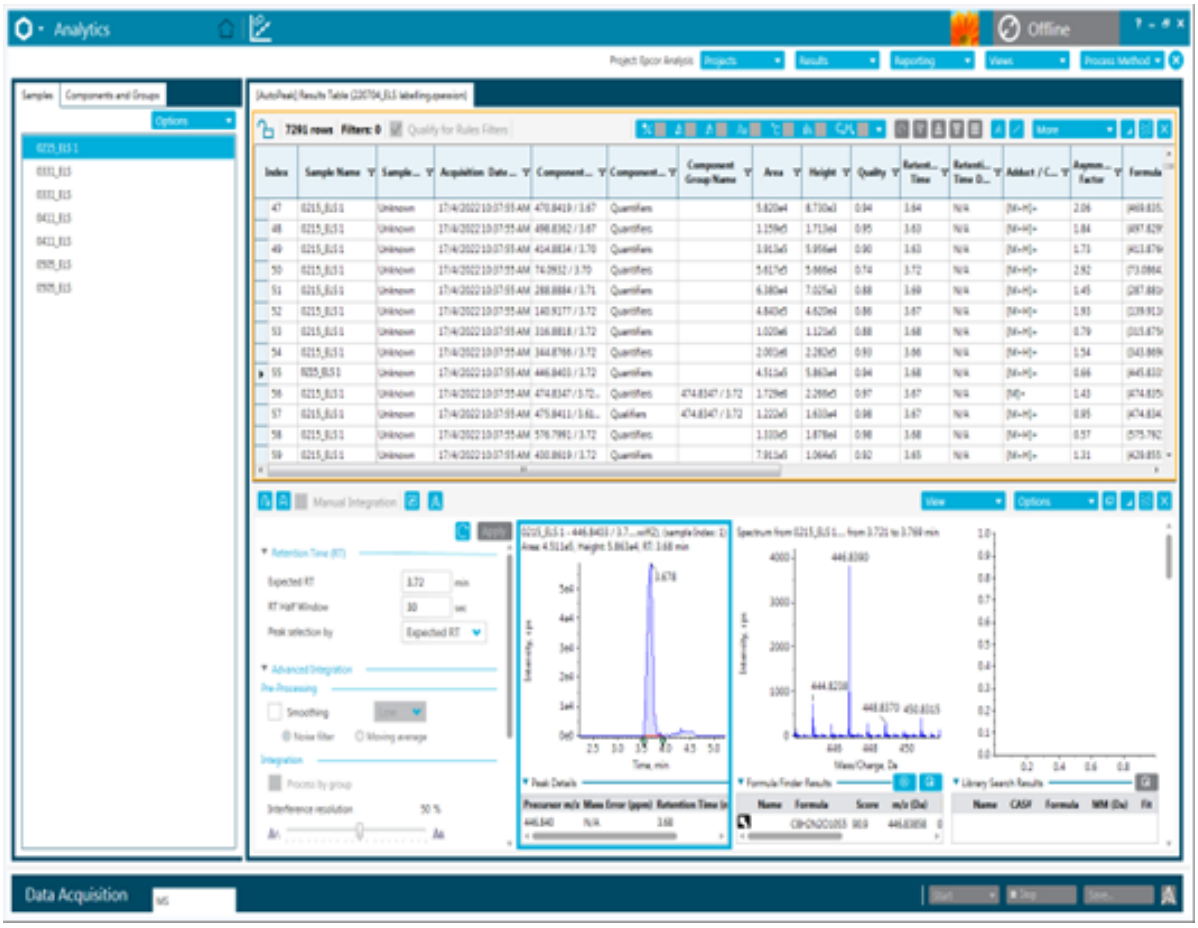


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		Retention T	Precursor Mass																
2		4.95E+00	158.117	Precursor Mass							ABS(C-D)								
3		5.05E+00	346.179	158.117	Precursor Mass						188.062	ABS(C-E)							
4		5.06E+00	171.148	346.179	158.117	Precursor Mass					175.031	13.031	ABS(C-F)						
5		5.06E+00	208.155	171.148	346.179	158.117	Precursor Mass				37.007	138.024	50.038	ABS(C-G)					
6		5.05E+00	342.153	208.155	171.148	346.179	158.117	Precursor Mass			133.998	171.005	4.026	184.036	ABS(C-H)				
7		5.09E+00	104.137	342.153	208.155	171.148	346.179	158.117			238.016	104.018	67.011	242.042	53.98				
8		5.09E+00	124.039	104.137	342.153	208.155	171.148	346.179			19.902	218.154	84.116	47.109	222.14				
9		5.08E+00	150.142	124.039	104.137	342.153	208.155	171.148			26.103	46.005	192.011	58.013	21.006				
10		5.08E+00	204.135	150.142	124.039	104.137	342.153	208.155			53.993	80.096	99.998	138.018	4.02				
11		5.09E+00	174.123	204.135	150.142	124.039	104.137	342.153			30.012	23.981	50.084	69.986	168.03				
12		5.09E+00	220.097	174.123	204.135	150.142	124.039	104.137			45.974	15.962	69.955	96.058	115.96				
13		5.09E+00	224.122	220.097	174.123	204.135	150.142	124.039			4.025	49.999	19.987	73.98	100.083				
14		5.09E+00	267.17	224.122	220.097	174.123	204.135	150.142			43.048	47.073	93.047	63.035	117.028				
15		5.08E+00	275.222	267.17	224.122	220.097	174.123	204.135			8.052	51.1	55.125	101.099	71.087				
16		5.10E+00	356.167	275.222	267.17	224.122	220.097	174.123			80.945	88.997	132.045	136.07	182.044				
17		5.07E+00	226.107	356.167	275.222	267.17	224.122	220.097			130.06	49.115	41.063	1.985	6.01				
18		5.36E+00	132.102	226.107	356.167	275.222	267.17	224.122			94.005	224.065	143.12	135.068	92.02				
19		5.10E+00	354.156	132.102	226.107	356.167	275.222	267.17			222.054	128.049	2.011	78.934	86.986				
20		5.28E+00	252.072	354.156	132.102	226.107	356.167	275.222			102.084	119.97	25.965	104.095	23.15				
21		5.50E+00	285.171	252.072	354.156	132.102	226.107	356.167			33.099	68.985	153.069	59.064	70.996				
22		5.67E+00	177.066	285.171	252.072	354.156	132.102	226.107			108.105	75.006	177.09	44.964	49.041				
23		5.65E+00	221.129	177.066	285.171	252.072	354.156	132.102			44.063	64.042	30.943	133.027	89.027				
24		5.63E+00	225.153	221.129	177.066	285.171	252.072	354.156			4.024	48.087	60.018	26.919	129.003				
25		5.71E+00	170.147	225.153	221.129	177.066	285.171	252.072			55.006	50.982	6.919	115.024	81.925				
26		5.72E+00	121.064	170.147	225.153	221.129	177.066	285.171			49.083	104.089	100.065	56.002	164.107				
27		5.68E+00	190.049	121.064	170.147	225.153	221.129	177.066			68.985	19.902	35.104	31.08	12.983				
28		5.76E+00	228.159	190.049	121.064	170.147	225.153	221.129			38.11	107.095	58.012	3.006	7.03				
29		5.83E+00	318.133	228.159	190.049	121.064	170.147	225.153			89.974	128.084	197.069	147.986	92.98				
30		5.56E+00	279.133	318.133	228.159	190.049	121.064	170.147			39	50.974	89.084	158.069	108.986				
31		5.73E+00	166.123	279.133	318.133	228.159	190.049	121.064			113.01	152.01	62.036	23.926	45.059				
32		6.01E+00	148.129	166.123	279.133	318.133	228.159	190.049			17.994	131.004	170.004	80.03	41.92				

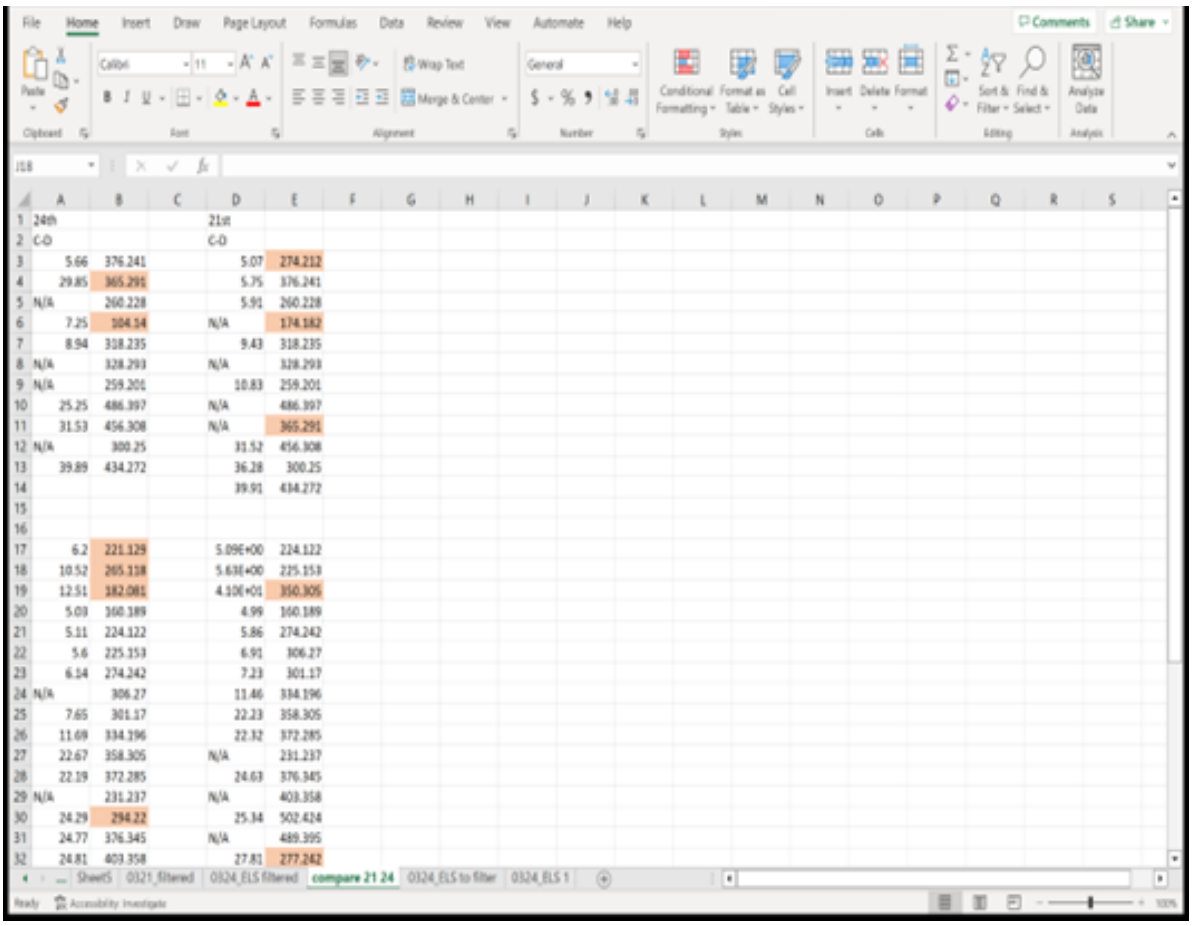
Step 3: Filter the mass differences between the precursor masses for the specific mass differences (Δm/z 2.012-2.013, 4.024-4.026, 6.036-6.039,8.048-8.052). Copy the retention time and precursor masses that show these specific differences to a new sheet to create a list of H/D feature pairs.



Step 4: Re-check the peak shapes and the similarity between the hydrogen and deuterium labelled in Sciex OS.



Step 5: Compare the H/D feature pairs across samples by manually comparing the values in Excel.



	A	B	C	D	E
1	24th			21st	
2	C-D			C-D	
3	5.66	376.241		5.07	274.212
4	29.85	365.291		5.75	376.241
5	N/A	260.228		5.91	260.228
6	7.25	184.14		N/A	174.182
7	8.94	318.235		9.43	318.235
8	N/A	328.293		N/A	328.293
9	N/A	259.201		10.83	259.201
10	25.25	486.397		N/A	486.397
11	31.53	456.308		N/A	365.291
12	N/A	300.25		31.52	456.308
13	39.89	434.272		36.28	300.25
14				39.91	434.272
15					
16					
17	6.2	221.129		5.09E+00	224.122
18	10.52	265.118		5.63E+00	225.153
19	12.51	182.081		4.10E+01	350.305
20	5.03	160.189		4.99	160.189
21	5.11	224.122		5.86	274.242
22	5.6	225.153		6.91	306.27
23	6.14	274.242		7.23	301.17
24	N/A	306.27		11.46	334.196
25	7.65	301.17		22.23	358.305
26	11.69	334.196		22.32	372.285
27	22.67	358.305		N/A	231.237
28	22.19	372.285		24.63	376.345
29	N/A	231.237		N/A	403.358
30	24.29	294.22		25.34	502.424
31	24.77	376.345		N/A	489.395
32	24.81	403.358		27.81	277.242

Figure 2.4- Overview of Manual Data Processing Procedure for Stable Isotopic Methyl Labelling Analysis.

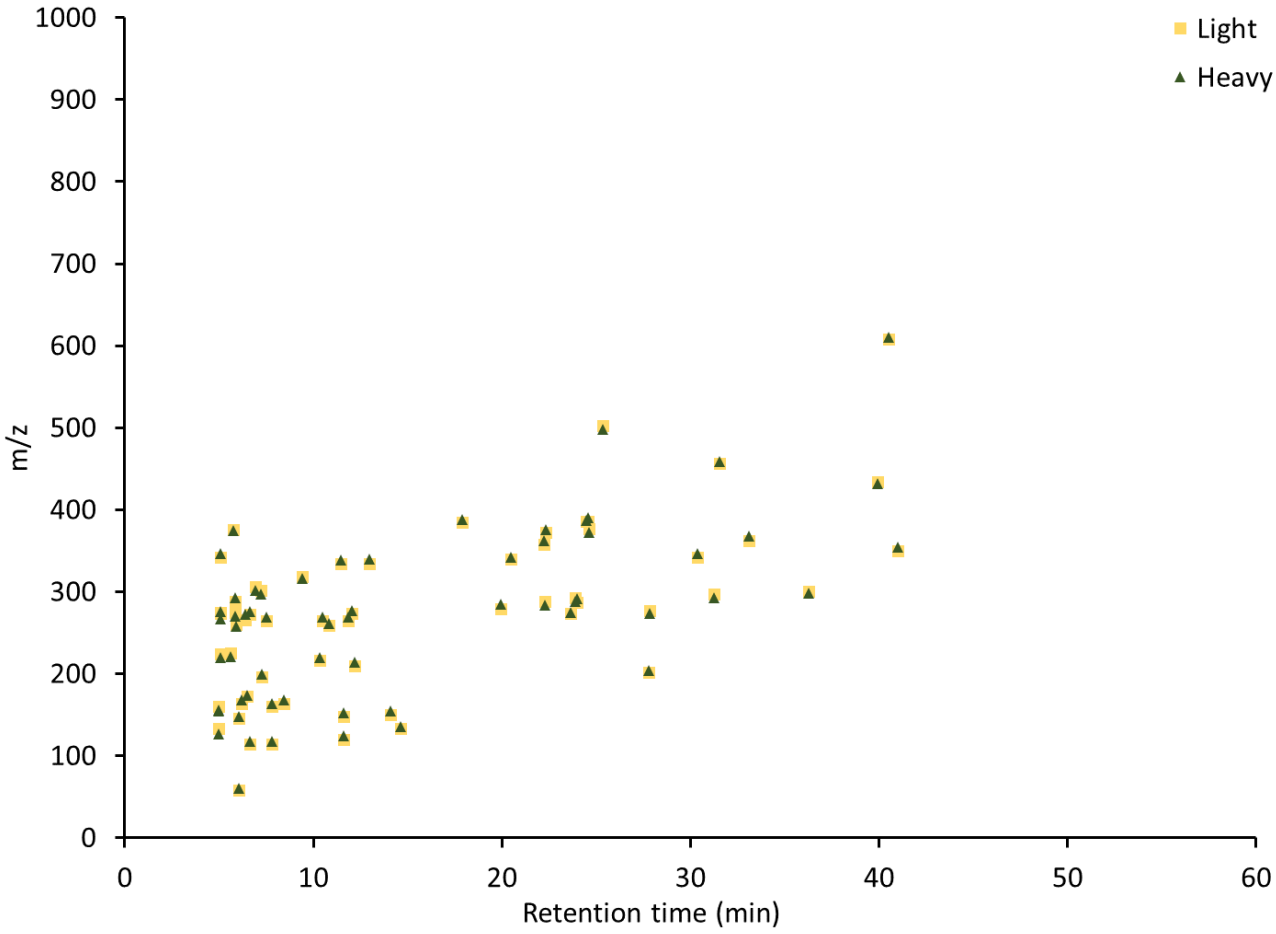


Figure 2.5 - H/D feature pairs (light/heavy labelled) of ELS on March 21st, 2022. Using manual analysis, 62 H/D feature pairs are found.

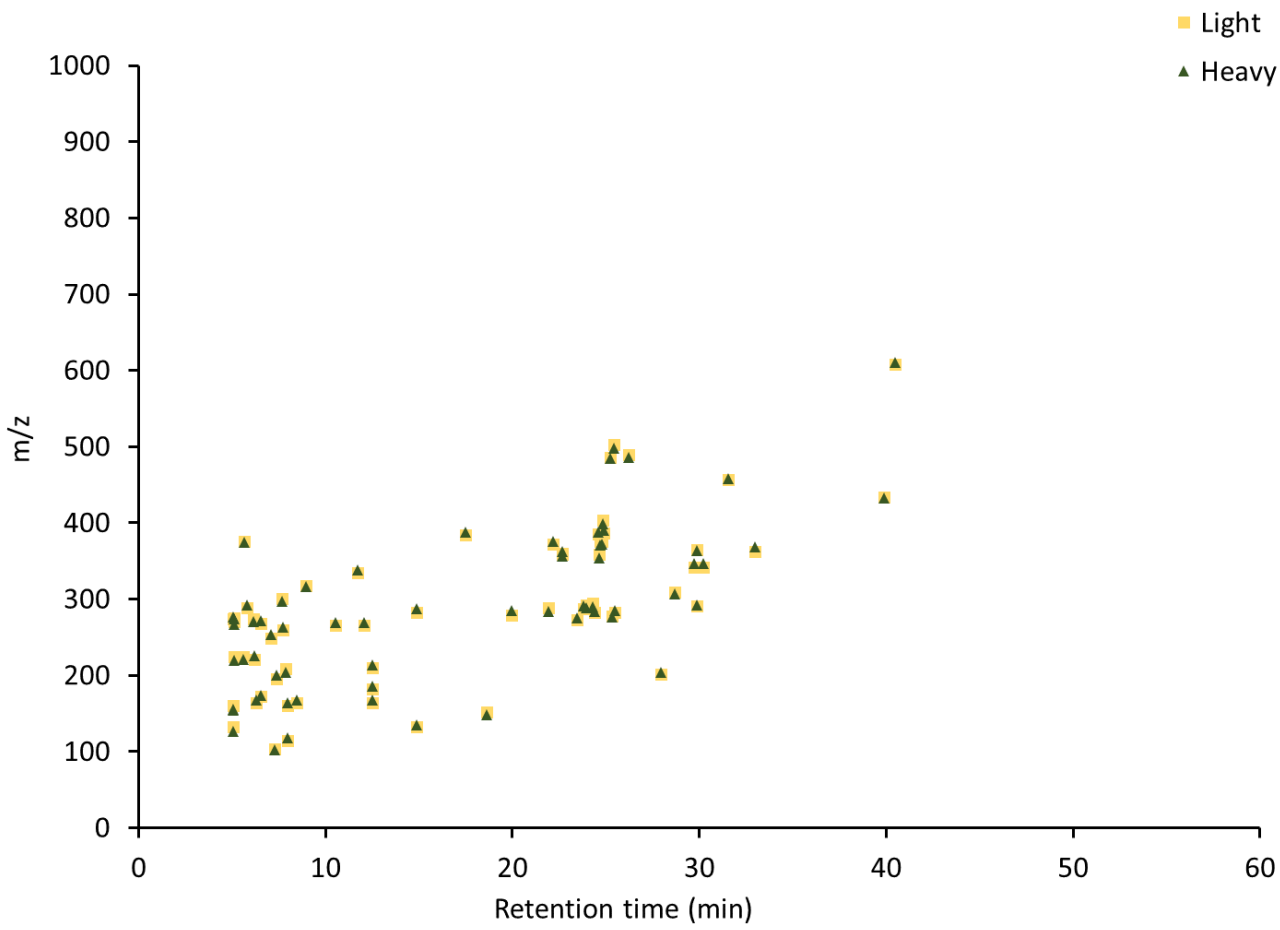


Figure 2.6 - H/D feature pairs (light/heavy labelled) of ELS on March 24th, 2022. Using manual analysis, 66 H/D feature pairs are found.

2.3.2: Pilot HDPairFinder Development and Application

To overcome the difficulties associated with data processing of the methyl labelled features, I collaborated with Dr. Tao Huan and his team from UBC. This led to the development of a data analysis program, HDPairFinder. HDPairFinder automatically searches for H/D feature pairs in the stable isotopic labelling data. In this collaborative effort, I collected samples, carried out sample analysis using H/D labelling, SPE extraction, and HPLC-HRMS non-targeted analysis. Then I performed initial testing of the pilot program and provided feedback for

Tingting Zhao, the co-first author on this project, to improve the program. The initial development and testing are presented in this chapter, while the final development and application will be presented in **Chapter 3**.[22] To test the initial performance of the program I analyzed the March 21st, 2022 sample. I used HDPairFinder to analyze the sample and compared the results to the manual data analysis procedure.

The first step of developing HDPairFinder is designing how to automatically select H/D feature pairs. To select pairs, we consider various parameters and evaluate the tolerance values. The first parameter is the m/z ranges for the identification of the hydrogen-deuterium isotopic pattern. In the pilot program, the m/z ratio ranges are Δm/z 2.012-2.013, 4.024-4.026, 6.036-6.039, 8.048-8.052, and are associated with 1, 2, 3, and 4 methyl tags reacted respectively. These ranges are selected to cover potential instrument errors while still being selective to the hydrogen and deuterium. The next parameter to consider is the retention time differences between the H/D feature pair peaks. Previous studies have shown that deuterium ions can impact reverse phase separation. Therefore the retention time differences between the hydrogen and deuterium pair require careful examination.[33] The initial tolerance for the retention time is set at ± 0.1 min, and the examination of peaks missed by this tolerance is considered later. Finally, the ratio of the two peak intensities needs consideration. Assuming the hydrogen and deuterium extracts are mixed at a 1:1 ratio, the intensity ratio should be close to 1.[18] The pilot program uses the peak area ratio of 1 with a tolerance of ± 0.2. While these parameters are considered in manual data analysis, the values are clearly defined for the program.

In the pilot form, HDPairFinder looks for peaks in a .csv file created using the Sciex OS software. It searches for all H/D feature pairs within a certain retention time and intensity (peak area) ratio tolerances (± 0.1 min/ 1 ± 0.2 respectively). **Figure 2.7** shows the initial results for the

March 21st, 2022 sample. A total of 34 H/D feature pairs are detected compared to the 62 H/D feature pairs from the manual method (**Figure 2.5/Figure 2.7)**. The retention time and intensity ratio tolerances are then increased to see the influence they may have on the number of H/D feature pairs detected. Tolerance values are adjusted from the initial ± 0.1 min/ 1 ± 0.2 (**Figure 2.7)**, to ± 0.1 min/ 1 ± 0.5 (**Figure 2.8)**, ± 0.2 min/ 1 ± 0.2 (**Figure 2.9)**, and ± 0.2 min/ 1 ± 0.5 (**Figure 2.10)**. Increasing the tolerance values shows the number of H/D feature pairs increase to close to the manual method. However, there are some differences in detected H/D feature pairs that requires further investigation.

I further investigated similarities between the peaks from the manual method and pilot program for the March 21st sample. Of the 62 manual peaks, 23 are detected in the pilot program with tolerance values of ± 0.2 min and ratio tolerance of ± 0.5. Peaks only found by the manual method are often explained by exceeding the retention time tolerance. On the other hand, those detected only by HDPairFinder are often because of limited distance between rows in the feature table used for the manual method. This is one of the known limitations of the manual method. The automated program overcomes this limit by comparing all peaks that meet the threshold values. **Table 2.2** shows the H/D feature pairs found through the manual analysis method and comparison to HDPairFinder. Further addressing these differences is considered in later development of the program.

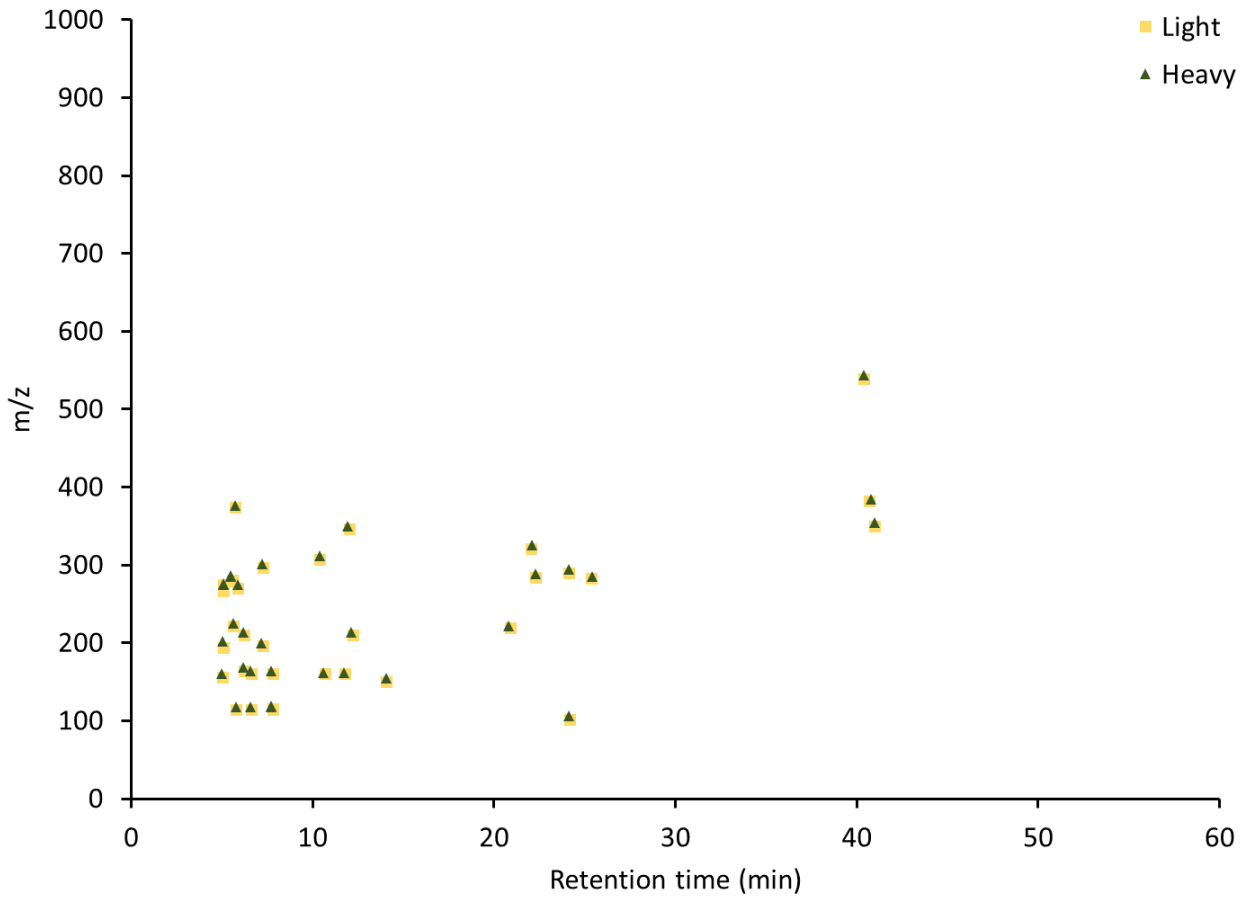


Figure 2.7 - H/D feature pairs (light/heavy labelled) of ELS on March 21st, 2022, with pilot program HDPairFinder. Tolerance values initially set to ± 0.2 intensity ratio, ± 0.1 min RT showed 34 labelled H/D feature pairs.

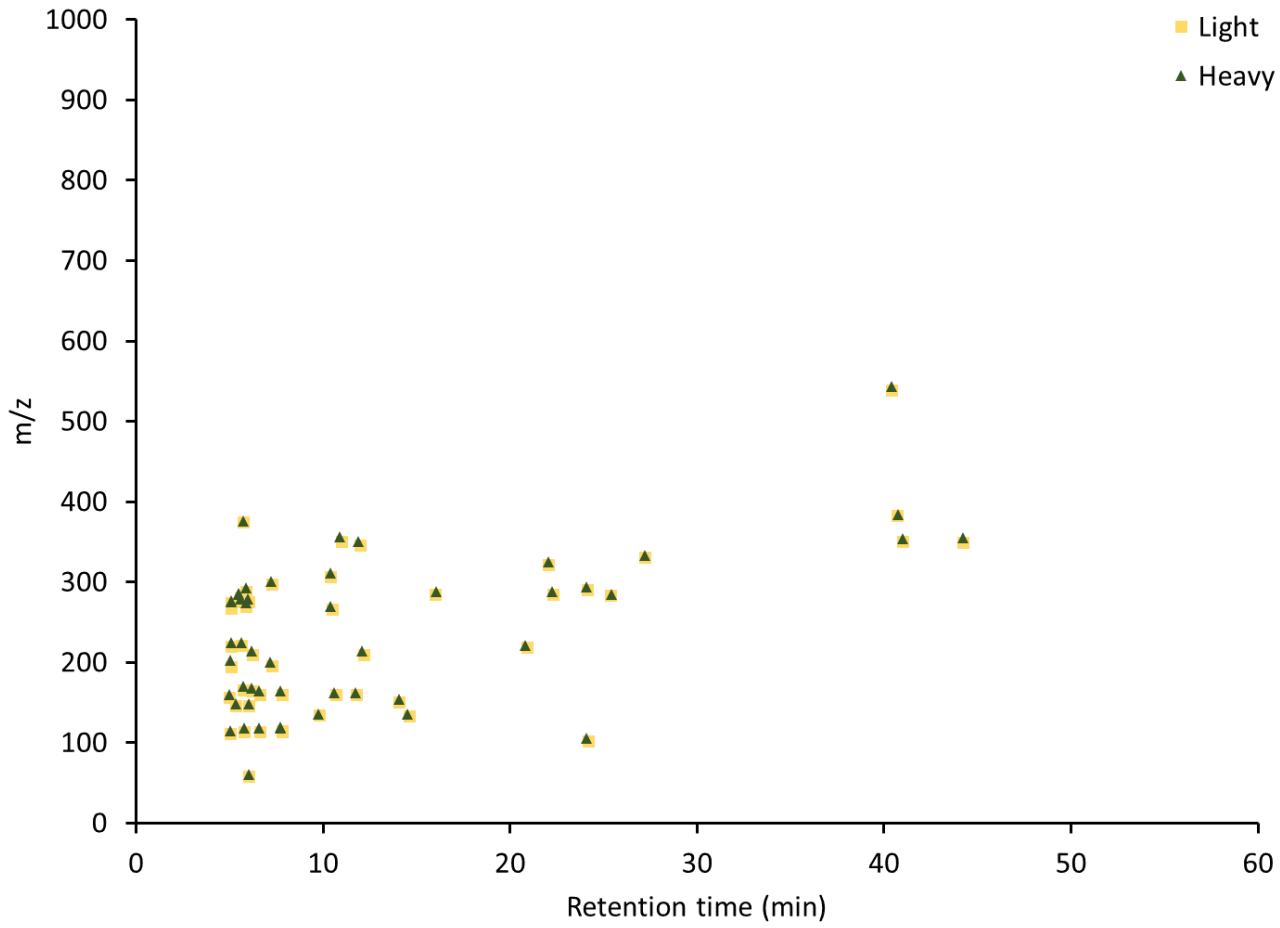


Figure 2.8 – H/D feature pairs (light/heavy) of ELS on March 21st, 2022, with pilot program HDPairFinder. Tolerance values initially set to ± 0.5 intensity ratio, ± 0.1 min RT showed 50 labelled H/D feature pairs.

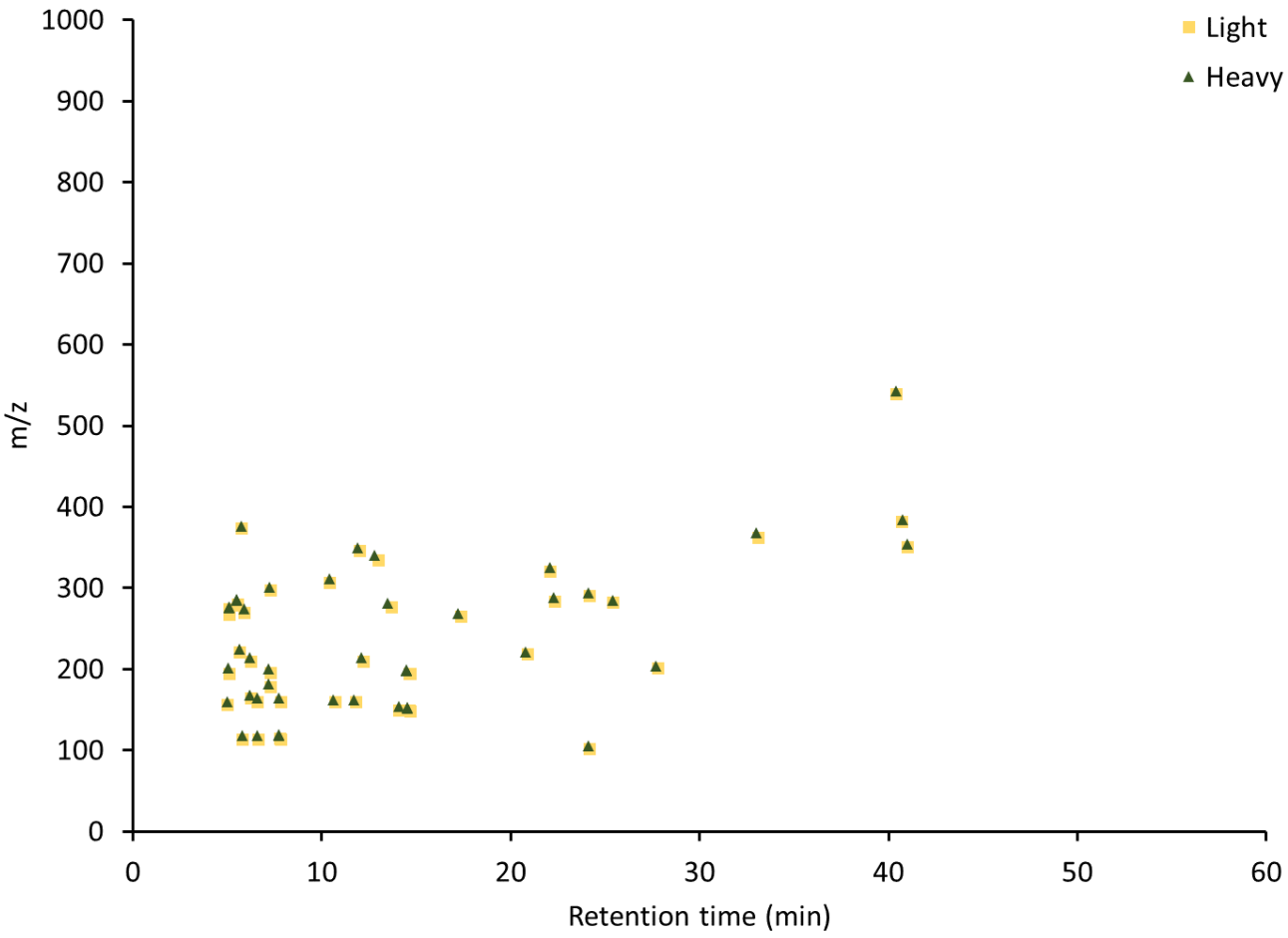


Figure 2.9 – H/D feature pairs (light/heavy labelled) of ELS on March 21st, 2022, with pilot program HDPairFinder. Tolerance values initially set to ± 0.2 intensity ratio, ± 0.2 min RT showed 45 labelled H/D feature pairs.

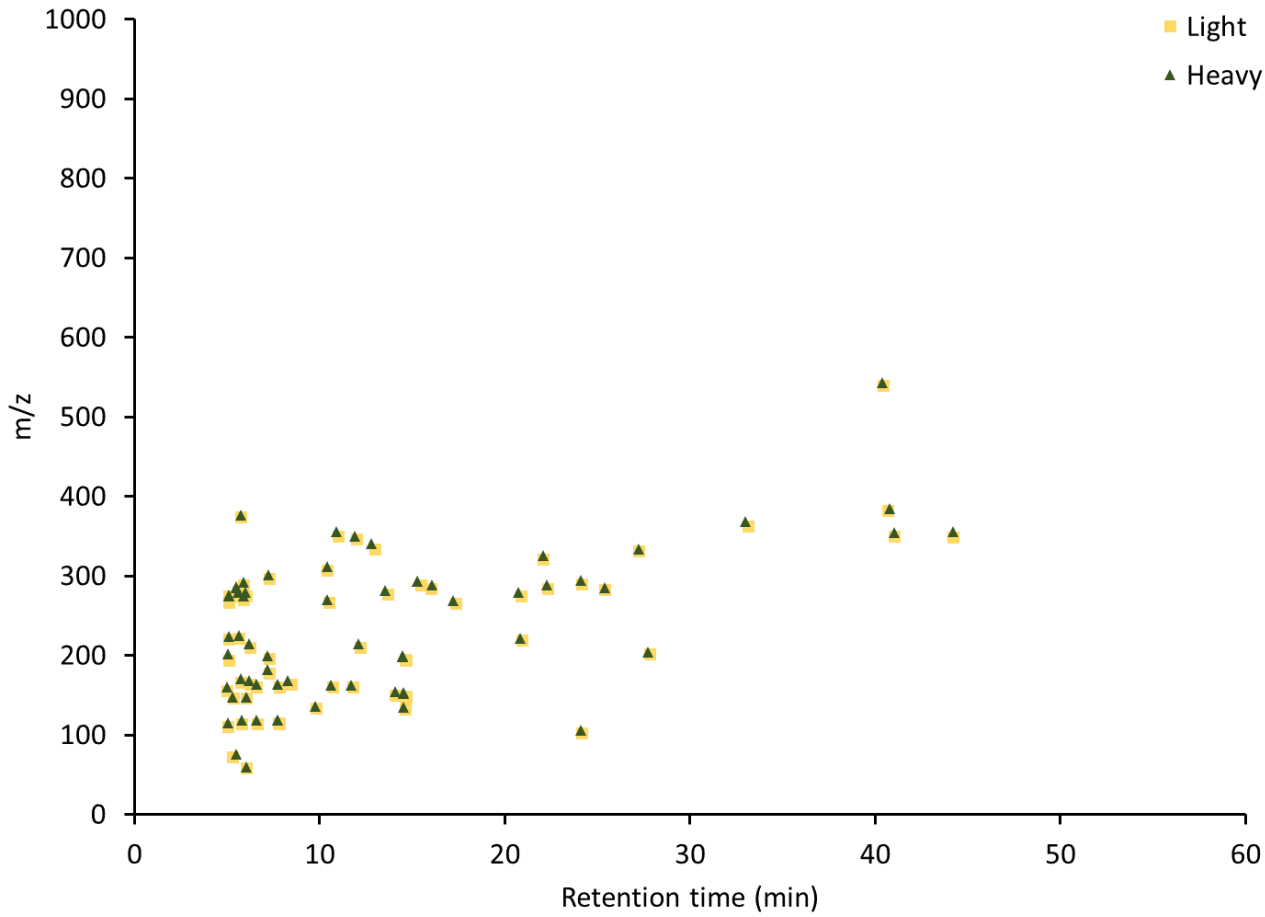


Figure 2.10 – H/D feature pairs (light/heavy labelled) of ELS on March 21st, 2022, with pilot program HDPairFinder. Tolerance values initially set to ± 0.5 intensity ratio, ± 0.2 min RT showed 66 H/D feature pairs.

Table 2.2 - List of features found by manual method and if they are found by the pilot HDPairFinder (Yes/No). This table includes the threshold values where the H/D feature pairs are found in both the pilot program. The threshold values include retention time (RT ± threshold in minutes) and intensity ratio (1 ± threshold). Reasons that pairs are not detected are listed. NA retention time indicates that one of the H/D feature pairs did not have a value in this specific sample when processed by Sciex OS. Therefore, it would have later been removed as a H/D feature pair. Those detected in a later version are attributed to changes in H/D feature pair searching algorithm.

Retention Time	**Light m/z**	**Heavy m/z**	**In pilot program (Yes/No)**	**What threshold (**±Int/RT)	**Why different**
6.03	58.065	60.077	Y	0.5/0.1	Was found by program
7.80	114.127	118.152	Y	0.2/0.1	Was found by program
6.61	114.128	118.153	Y	0.2/0.1	Was found by program
11.61	120.081	124.106	N	N/A	± 0.3 RT
4.98	127.14	133.177	N	N/A	0.3 Intensity Ratio
14.63	133.089	135.101	Y	0.5/0.1	Was found by program

6.03	146.117	148.129	Y	0.5/0.1	Was found by program
11.61	148.075	152.1	N	N/A	± 0.3 RT
14.08	150.128	154.152	Y	0.2/0.1	Was found by program
4.99	154.151	160.189	N	N/A	0.38 Intensity Ratio
4.99	156.165	160.189	Y	0.2/0.1	Was found by program
7.80	160.133	164.158	Y	0.2/0.1	Was found by program
8.44	164.071	168.096	Y	0.5/0.2	Was found by program
6.21	164.107	168.132	Y	0.2/0.1	Was found by program
6.47	172.169	174.182	N	N/A	4 Intensity Ratio
7.26	196.096	200.121	Y	0.2/0.1	Was found by program
27.79	202.086	204.099	Y	0.2/0.2	Was found by program
12.19	210.113	214.138	Y	0.2/0.1	Was found by program
10.32	216.159	220.184	N	N/A	± 1 RT

5.09	220.097	224.122	Y	0.5/0.1	Was found by program
5.63	221.129	225.153	Y	0.2/0.1	Was found by program
5.91	258.216	260.228	N	N/A	2 Intensity Ratio
10.83	259.201	261.213	N	N/A	NA Retention Time
7.52	265.118	269.142	N	N/A	In later version
10.49	265.118	269.143	N	N/A	In later version
11.84	265.118	269.143	N	N/A	In later version
6.39	266.193	272.231	N	N/A	± 0.23 RT, 0.06 Intensity Ratio
5.08	267.17	275.222	N	N/A	In later version
5.86	270.218	274.242	N	N/A	In later version
6.62	272.231	276.256	N	N/A	41 Intensity Ratio
12.03	273.216	277.241	N	N/A	50 Intensity Ratio
27.81	273.217	277.242	N	N/A	± 0.4 RT
23.64	273.218	275.23	N	N/A	20 Intensity Ratio
5.07	274.212	276.224	Y	0.2/0.1	Was found by program
19.93	279.133	285.171	N	N/A	In later version
22.27	284.196	288.221	Y	0.2/0.1	Was found by program

23.96	287.232	291.257	N	N/A	NA Retention Time
5.87	288.191	292.216	Y	0.5/0.1	Was found by program
23.87	288.236	292.261	N	N/A	0.44 Intensity Ratio
31.25	293.186	297.211	N	N/A	In later version
7.23	297.145	301.17	Y	0.2/0.1	Was found by program
36.28	298.237	300.25	N	N/A	NA Retention Time
6.91	302.244	306.27	N	N/A	0.09 Intensity Ratio
9.43	316.223	318.235	N	N/A	± 0.46 RT
12.98	334.176	340.214	Y	0.2/0.2	Was found by program
11.46	334.196	338.222	N	N/A	2.2 Intensity Ratio
20.48	340.295	342.308	N	N/A	NA Retention Time
5.05	342.153	346.179	N	N/A	NA Retention Time
30.39	342.311	346.336	N	N/A	NA Retention Time

41.01	350.305	354.33	Y	0.2/0.1	Was found by program
22.23	358.305	362.33	N	N/A	20 Intensity Ratio
33.12	362.207	368.243	Y	0.2/0.2	Was found by program
22.32	372.285	376.309	N	N/A	NA Retention Time
24.63	372.32	376.345	N	N/A	0.03 Intensity Ratio
5.75	374.228	376.241	Y	0.2/0.1	Was found by program
17.87	384.213	388.239	N	N/A	NA Retention Time
24.46	385.317	387.33	N	N/A	NA Retention Time
24.59	386.301	390.326	N	N/A	NA Retention Time
39.91	432.259	434.272	N	N/A	56 Intensity Ratio
31.52	456.308	458.32	N	N/A	0.23 Intensity Ratio
25.34	498.398	502.424	N	N/A	0.04 Intensity Ratio

40.50	608.363	610.376	N	N/A	96 Intensity Ratio

2.3.3: Benefits and Limitation of Manual vs. HDPairFinder Data Analysis

While manual data analysis allows for prioritization of the H/D feature pairs, it has many remaining limitations. For example, manual analysis of non-target data is difficult to apply to large datasets as analysis is time consuming. In the stable isotopic labelling method, manual analysis can also miss important H/D feature pairs because the search for a pair occurs only within five rows on the feature table. If samples have several peaks eluted at the same retention time and therefore many rows, H/D feature pairs could be missed by not being in proximity. Another concern is the deuterium isotopic effect where retention times can shift because of deuterium ions.[29] In the manual analysis method, this is not accounted for as H/D feature pair search occurs in a very tight range. Peak intensity is also not an initial consideration, so the list of pairs could show two peaks with significant differences. Differences in the peak intensities could be indicative of a false feature. Because of the limited ability to look through peaks, manual analysis is limiting the potential of non-target analysis. Thus, the potential benefits of automated programs for methyl-based labelling should be further investigated.

The pilot HDPairFinder program improves the detection of H/D feature pairs relative to the manual method. The processing of a single sample takes only minutes with HDPairFinder compared to days by the manual method. In the future, this would be beneficial to running multiple samples. However, further optimization and development is still required to ensure proper peak selection. Through optimizing threshold values for parameters or using a “scoring”

system based on multiple parameters, feature detection could be improved. Improving feature detection is necessary to ensure the identification of "true" H/D feature pairs.

A limitation in the pilot program and manual data analysis is the choice of peak picking algorithm. Various algorithms have been reviewed, and differences in raw data processing can be seen depending on the peak picking algorithm.[30] Manual data analysis still requires the use of Sciex OS algorithms for the initial peak picking. Specifically for Sciex OS, the algorithm is not named and the limitations are not completely understood. Pursuing selection of a new peak picking algorithm will improve the overall workflow. Consistent use of the algorithm will also provide a direction for standardization across research groups. By further developing HDPairFinder, it has a high potential for its application compared to manual analysis.

2.4: Conclusion

Detection of nitrogen containing compounds is important in improving our understanding of the composition of organic compounds in our source water. Thus, we can improve our knowledge of potential DBP precursors. However, the large amount of data generated by non-target analysis poses challenges to manual data analysis. The analysis workflow is time consuming, often requiring days of work for a single sample, and does not produce an exhaustive list of peaks. To improve this workflow, I collaborated with Dr. Tao Huan and Tinging Zhao to build HDPairFinder. Compared to other available programs, HDPairFinder can reliably process data of non-targeted analysis of methyl labelled samples. In the initial testing, the pilot HDPairFinder shows the potential to reduce time required for analysis. The pilot program is successful in detecting nitrogen containing compounds in source water. Further development and application of HDPairFinder will be presented in **Chapter 3**.

2.5: References

(1) Crittenden, J. C.; Trussell, R. R.; Hand, D. W.; Howe, K. J.; Tchbanoglous, G. *MWH's Water Treatment: Principles and Design*, Third.; John Wiley and Sons Inc., *2012*.

(2) Richardson, S. D.; Plewa, M. J.; Wagner, E. D.; Schoeny, R.; Demarini, D. M. Occurrence , Genotoxicity , and Carcinogenicity of Regulated and Emerging Disinfection by-Products in Drinking Water : A Review and Roadmap for Research. *Mutat. Res. - Rev. Mutat.* **2007**, 636, 178–242. https://doi.org/10.1016/j.mrrev.2007.09.001.

(3) Han, J.; Zhang, X.; Jiang, J.; Li, W. How Much of the Total Organic Halogen and Developmental Toxicity of Chlorinated Drinking Water Might Be Attributed to Aromatic Halogenated DBPs? *Environ. Sci. Technol.* **2021**, 55 (9), 5906–5916. https://doi.org/10.1021/acs.est.0c08565.

(4) Liu, C.; Ersan, M. S.; Wagner, E.; Plewa, M. J.; Amy, G.; Karanfil, T. Toxicity of Chlorinated Algal-Impacted Waters: Formation of Disinfection Byproducts vs. Reduction of Cyanotoxins. *Water Res.* **2020**, 184, 116145. https://doi.org/10.1016/j.watres.2020.116145.

(5) Zhang, D.; Chu, W.; Yu, Y.; Krasner, S. W.; Pan, Y.; Shi, J.; Yin, D.; Gao, N. Occurrence and Stability of Chlorophenylacetonitriles: A New Class of Nitrogenous Aromatic DBPs in Chlorinated and Chloraminated Drinking Waters. *Environ. Sci. Technol. Lett.* **2018**, 5 (6), 394–399. https://doi.org/10.1021/acs.estlett.8b00220.

(6) Cai, L.; Li, L.; Yu, S. Formation of Odorous Aldehydes , Nitriles and N -Chloroaldimines from Combined Leucine in Short Oligopeptides during Chlorination. *Water Res.* **2020**, 177, 115803. https://doi.org/10.1016/j.watres.2020.115803.

(7) Essaïed, K. A.; Brown, L. V.; von Gunten, U. Reactions of Amines with Ozone and Chlorine: Two Novel Oxidative Methods to Evaluate the N-DBP Formation Potential from Dissolved Organic Nitrogen. *Water Res.* **2022,** 209, 117864. https://doi.org/10.1016/j.watres.2021.117864.

(8) Zhang, H.; Gao, P.; Liu, Y.; Du, Z.; Feng, L.; Zhang, L. Effects of Different Types of Nitrogen Sources in Water on the Formation Potentials of Nitrogenous Disinfection By-Products in Chloramine Disinfection Process Based on Isotope Labeling. *Sci. Total Environ.* **2022**, 842, 156692. https://doi.org/10.1016/J.SCITOTENV.2022.156692.

(9) Huang, G.; Jiang, P.; Jmaiff Blackstock, L. K.; Tian, D.; Li, X. F. Formation and Occurrence of Iodinated Tyrosyl Dipeptides in Disinfected Drinking Water. *Environ. Sci. Technol.* **2018**, 52 (7), 4218–4226. https://doi.org/10.1021/acs.est.7b06276.

(10) Huang, G.; Jiang, P.; Li, X. F. Mass Spectrometry Identification of N-Chlorinated Dipeptides in Drinking Water. *Anal. Chem.* **2017**, 89 (7), 4204–4209. https://doi.org/10.1021/acs.analchem.7b00228.

(11) Qiu, J.; Craven, C.; Wawryk, N.; Carroll, K.; Li, X.-F. Integration of Solid Phase Extraction with HILIC-MS/MS for Analysis of Free Amino Acids in Source Water. *J. Environ. Sci. (China)* **2022**, 117, 190-196. https://doi.org/10.1016/j.jes.2022.04.025.

(12) How, Z. T.; Busetti, F.; Linge, K. L.; Kristiana, I.; Joll, C. A.; Charrois, J. W. A. Analysis of Free Amino Acids in Natural Waters by Liquid Chromatography – Tandem Mass Spectrometry. *J. Chromatogr. A* **2014**, 1370, 135–146. https://doi.org/10.1016/j.chroma.2014.10.040.

(13) Wallis, P. M.; Hynes, H. B. N.; Telang, S. A. The Importance of Groundwater in the Transportation of Allochthonous Dissolved Organic Matter to the Streams Draining a Small Mountain Basin. *Hydrobiologia* **1981**, 79 (1), 77–90. https://doi.org/10.1007/BF00005821.

(14) Bulloch, D. N.; Nelson, E. D.; Carr, S. A.; Wissman, C. R.; Armstrong, J. L.; Schlenk, D.; Larive, C. K. Occurrence of Halogenated Transformation Products of Selected Pharmaceuticals and Personal Care Products in Secondary and Tertiary Treated Wastewaters from Southern California. *Environ. Sci. Technol*. **2015**, 49 (4), 2044–2051. https://doi.org/10.1021/es504565n.

(15) Jasemizad, T.; Bromberg, L.; Hatton, T. A.; Padhye, L. P. Oxidation of Betrixaban to Yield N-Nitrosodimethylamine by Water Disinfectants. *Water Res.* **2020**, 186, 116309. https://doi.org/10.1016/j.watres.2020.116309.

(16) Bedner, M.; MacCrehan, W. A. Transformation of Acetaminophen by Chlorination Produces the Toxicants 1,4-Benzoquinone and N-Acetyl-p-Benzoquinone Imine. *Environ. Sci. Technol.* **2006**, 40 (2), 516–522. https://doi.org/10.1021/es0509073.

(17) Cao, F.; Zhang, M.; Yuan, S.; Feng, J.; Wang, Q.; Wang, W.; Hu, Z. Transformation of Acetaminophen during Water Chlorination Treatment: Kinetics and Transformation Products Identification. *Environ. Sci. Pollut. Res*. **2016**, 23 (12), 12303–12311. https://doi.org/10.1007/s11356-016-6341-x.

(18) Liu, Z.; Craven, C. B.; Huang, G.; Jiang, P.; Wu, D.; Li, X. F. Stable Isotopic Labeling and Nontarget Identification of Nanogram/Liter Amino Contaminants in Water. *Anal. Chem.* **2019**, 91 (20), 13213–13221. https://doi.org/10.1021/acs.analchem.9b03642.

(19) Hollender, J.; van Bavel, B.; Dulio, V.; Farmen, E.; Furtmann, K.; Koschorreck, J.; Kunkel, U.; Krauss, M.; Munthe, J.; Schlabach, M.; Slobodnik, J.; Stroomberg, G.; Ternes, T.; Thomaidis, N. S.; Togola, A.; Tornero, V. High Resolution Mass Spectrometry-Based Non-Target Screening Can Support Regulatory Environmental Monitoring and Chemicals Management. *Environ. Sci. Eur.* **2019**, 31, 42. https://doi.org/10.1186/s12302-019-0225-x.

(20) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N. S.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; De Boer, R.; Reid, M. J.; Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.; Bogialli, S.; Stipaničev, D.; Rostkowski, P.; Hollender, J. Non-Target Screening with High-Resolution Mass Spectrometry: Critical Review Using a Collaborative Trial on Water Analysis. *Anal. Bioanal. Chem*. **2015**, 407 (21), 6237–6255. https://doi.org/10.1007/s00216-015-8681-7.

(21) Xing, S.; Shen, S.; Xu, B.; Huan, T. Molecular Formula Discovery via Bottom-up MS/MS Interrogation. *Nat. Methods* **2023**, 20 (6), 881-890. https://doi.org/10.1038/s41592-023-01850-x.

(22) Zhao, T.; Carroll, K.; Craven, C. B.; Wawryk, N. J. P.; Xing, S.; Guo, J.; Li, X.; Huan, T. HDPairFinder : A Data Processing Platform for Hydrogen / Deuterium Isotopic Labeling-Based Nontargeted Analysis of Trace-Level Amino-Containing Chemicals in Environmental Water. *J. Environ. Sci.* **2024**, 136, 583–593. https://doi.org/10.1016/j.jes.2023.02.033.

(23) Krauss, M.; Singer, H.; Hollender, J. LC-High Resolution MS in Environmental Analysis: From Target Screening to the Identification of Unknowns. *Anal. Bioanal. Chem*. **2010**, 397 (3), 943–951. https://doi.org/10.1007/s00216-010-3608-9.

(24) Bonnefille, B.; Karlsson, O.; Rian, M. B.; Raqib, R.; Parvez, F.; Papazian, S.; Islam, M. S.; Martin, J. W. Nontarget Analysis of Polluted Surface Waters in Bangladesh Using Open Science Workflows. *Environ. Sci. Technol.* **2023**. 57(17), 6808-6824. https://doi.org/10.1021/acs.est.2c08200.

(25) Wawryk, N. J. P.; Craven, C. B.; Blackstock, L. K. J.; Li, X. F. New Methods for Identification of Disinfection Byproducts of Toxicological Relevance: Progress and Future Directions. *J. Environ. Sci. (China)* **2021**, 99, 151–159. https://doi.org/10.1016/j.jes.2020.06.020.

(26) Helmus, R.; van de Velde, B.; Brunner, A. M.; ter Laak, T. L.; van Wezel, A. P.; Schymanski, E. L. patRoon 2.0: Improved Non-Target Analysis Workflows Including Automated Transformation Product Screening. *J. Open Source Softw.* **2022**, 7 (71), 4029. https://doi.org/10.21105/joss.04029.

(27) Guo, J.; Shen, S.; Xing, S.; Yu, H.; Huan, T. ISFrag: De Novo Recognition of In-Source Fragments for Liquid Chromatography-Mass Spectrometry Data. *Anal. Chem.* **2021**, 93 (29), 10243–10250. https://doi.org/10.1021/acs.analchem.1c01644.

(28) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H. C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat. Methods* **2016,** 13 (9), 741–748. https://doi.org/10.1038/nmeth.3959.

(29) Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E. R. Deuterium Isotope Effects on Hydrophobic Interactions: The Importance of

Dispersion Interactions in the Hydrophobic Phase. *J. Am. Chem. Soc.* **2003**, 125 (45), 13836–13849. https://doi.org/10.1021/ja036006g.

(30) Guo, J.; Huan, T. Mechanistic Understanding of the Discrepancies between Common Peak Picking Algorithms in Liquid Chromatography-Mass Spectrometry-Based Metabolomics. *Anal. Chem*. **2022**. 95(14), 5894–5902. https://doi.org/10.1021/acs.analchem.2c04887.

Chapter 3

Application of HDPairFinder to Study Source Water Changes over Spring Runoff

3.1: Introduction

Following the initial development of HDPairFinder in **Chapter 2**, I further developed the program with our collaborators and published the final version. In the publication, the optimized default values for pair picking parameters were investigated. The final program is freely available for download on github (https://github.com/HuanLab/HDPairFinder).[1] An overview of the workflow of this program can be seen in **Figure 3.1**.[1] The final version of HDPairFinder includes four modules: **Module 1**) extraction of H/D feature pairs **Module 2**) alignment across samples **Module 3**) gap-filling of missing values and **Module 4**) putative annotation.

Module 1 completes the extraction of Hydrogen/Deuterium (H/D) feature pairs. It can be further divided into three components: peak picking, H/D feature pair searching, and data cleaning (**Figure 3.1a)**. First, the peak picking from raw data is completed using the well developed *CentWave* algorithm from XCMS R package.[2,3] Using the list of peaks, the H/D feature pairs are searched using multiple criteria including peak-peak cross correlation, intensity ratio, and retention time tolerances between the hydrogen and deuterium labelled peaks. Following H/D feature pair picking, data cleaning step is implemented using a previously developed program, ISFrag.R, is incorporated. Data cleaning using ISFrag allows for the

*Parts of Chapter 3 (i.e Section 3.1 and Figure 3.1) were published in Zhao, T.; Carroll, K.; Craven, C. B.; Wawryk, N. J. P.; Xing, S.; Guo, J.; Li, X.; Huan, T. HDPairFinder : A Data Processing Platform for Hydrogen / Deuterium Isotopic Labeling-Based Nontargeted Analysis of Trace-Level Amino-Containing Chemicals in Environmental Water. J. Environ. Sci. 2024, 136, 583–593. Reprinted with permission. Copyright 2023 Elsevier.

removal of any H/D feature pairs caused by in-source fragmentation, salt adduct formation, and naturally occurring ^{13}C isotopes.[4] While MS parameters could be adjusted to help minimize in-source fragmentation and adduct formation, with electrospray ionization it will be inevitable. Therefore, the data cleaning is a necessary step to ensure higher quality data. Following the data cleaning, the list of H/D feature pairs are output in a csv file.

Module 2 performs alignment of all data from multiple samples. Feature alignment is performed using an in-house built program to merge the same feature across multiple samples (**Figure 3.1b**). This prepares an integrated feature intensity table for downstream quantitative comparison and statistical analysis. The alignment process is based on the m/z of H/D feature pairs, and the retention time of hydrogen labelled compound. The outcome of alignment is a feature table of all the labelled compounds and their labelled m/z, retention time, and the intensity of H- features in all the samples.

Module 3 performing gap-filling of missing values (**Figure 3.1c)**. Missing values are largely presented in aligned feature intensity tables.[5] Missing values can be classified into two kinds. The first kind is contributed from chemical compounds of low signal abundance and thus are not detected in the HPLC-HRMS analysis. The second kind is contributed from the poor performance of peak-picking algorithm that does not recognize the chromatographic peak even if it has a high abundance.[6] A common practice of treating missing value is to replace it by a small value or using machine learning to make predictions. These approaches can address missing values of the first kind but not the second kind. In the algorithm design, HDPairFinder implements an evidence-based missing value imputation to retrieve missing values of the second kind. This algorithm uses the values of m/z and retention time to look into the raw data. If the chromatographic peaks of the H/D feature pair exist, they will be extracted and evaluated to

retrieve the missing value. The detailed quality evaluation includes three criteria: (1) the maximum intensity of H/D feature pair is larger than certain intensity threshold; (2) the intensity ratio of H/D feature pair is in a specified range; (3) The peak-peak cross correlation between the H/D feature pair is larger than or equal to a threshold. Only the H/D feature pairs that pass all the three criteria are considered missing values of the second kind and the missing value will be filled by the peak intensity of recovered pairs. Otherwise, the missing value will be considered as the first kind and be replaced by a small value in the downstream data processing. The gap filling results are output into a csv file.

Module 4 of the program assists in compound identification through putative annotations based on an accurate mass match **(Figure 3.1d)**. An inhouse library (AMINES) was created through the combination of amine-containing compounds from available databases including the Human Metabolome Database (HMDB), Toxin and Toxin-Target Database (T3DB), MassBank of North America (MoNA) and National Institute of Standards and Technology (NIST). The final library contains 38,000 structure-disjoint primary and secondary amino compounds. The AMINES library is used for matching an accurate mass to the unlabelled mass from the alignment or gap-filled data. To calculate the unlabelled mass, a $[M+H]^+$ adduct is assumed and the accurate mass of the methyl tags subtracted (14.0096 for hydrogen methyl tag and 16.0267 for deuterated methyl tag). Together, the four modules of HDPairFinder allows for automation of data analysis from raw data files to a final list of H/D feature pairs across multiple samples.

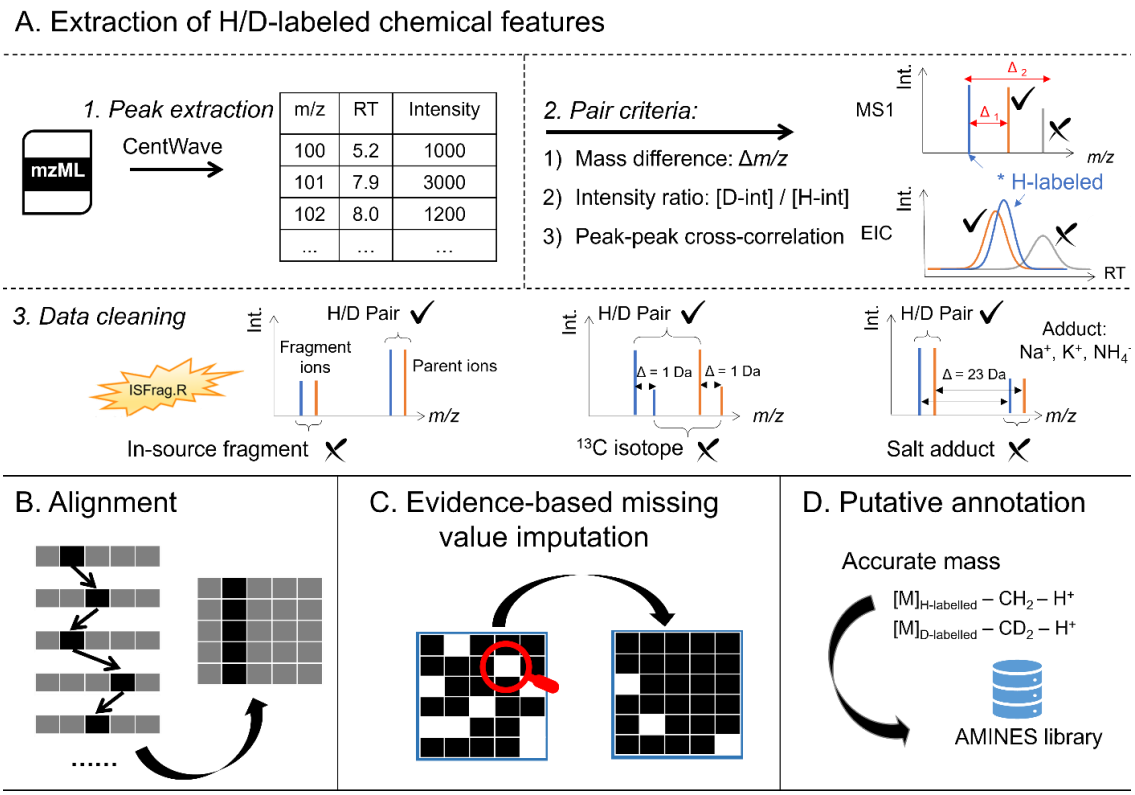


Figure 3.1- Workflow of HDPairFinder. A) the peak picking from the raw data, H/D feature pair picking, and data cleaning B) alignment across multiple samples C) gap-filling using the raw data D) putative annotation using accurate mass matches to unlabelled m/z.

The capabilities of HDPairFinder provide a strong starting point for in-depth investigation into non-targeted data. Therefore, I decided to apply the program for studying seasonal changes in organic compounds in source water samples. A particular interest in studying seasonal changes is spring runoff, where a major melting event leads to a significant influx of natural organic matter (NOM) entering source water.[7,8] NOM can react with disinfectants during the treatment process to form disinfection by-products (DBPs). Therefore, as NOM increases during the spring, DBPs formation will increase.[9,10] Epidemiological studies have observed potential association of DBP exposure with chronic adverse effects on human health. Thus, it is

important to try and minimize their formation.[11] To minimize DBP formation and control an influx of NOM, a common water treatment practice is NOM removal before disinfectant addition.

Treatment processes to remove NOM often include coagulation and filtration, which can remove large molecular mass compounds in the water.[12] Additional treatment can include the addition of powdered activated carbon (PAC), which removes a wide range of organic compounds including some small water-soluble molecules present in water.[13,14] However, a lack of characterization of the NOM in incoming source limits the understanding of the effectiveness of the NOM removal. Therefore, comprehensive analysis of organic compounds in incoming source water during the spring is required.

While common water quality parameters including total organic nitrogen (TON), ammonia as nitrogen (NH3-N), and Total Kjeldahl Nitrogen (TKN) are used by water treatment facilities to track nitrogen in source water, do not provide information to identify compounds. Therefore, a method outside these common parameters is required to gain a comprehensive understanding of the composition of nitrogen containing NOM. Here, I use a stable isotopic labelling method in conjunction with data analysis using HDPairFinder to investigate how amine-containing compounds are changing over the course of the spring.

3.2: **Experimental**

3.2.1: Chemicals and Reagents

Formaldehyde (CH_2O, 37 wt. % in H_2O, contains 10−15% methanol as stabilizer), deuterated formaldehyde (CD_2O, ~20 wt. % in D_2O, 98 atom % D), sodium cyanoborohydride ($NaBH_3CN$, 95%), formic acid (FA, 99%), and nylon disk filters (0.45 μm) were obtained from Sigma-Aldrich (St. Louis, MO). Optima water, methanol, acetonitrile (ACN), ammonium

hydroxide (30% wt.), and glass microfiber filters (1.5 µm) were purchased from Fisher Scientific (Fair Lawn, NJ). Oasis MCX cartridges (6 mL, 150 mg of sorbent) were obtained from Waters (Milford, MA). Syringe filters (0.45 µm, PVDF) were obtained from Dikma (Markham, ON). Amino acids, Leucine, Isoleucine, Tyrosine, Tryptophan, Threonine, and Phenylalanine were purchased from Sigma-Aldrich (St. Louis, MO)

3.2.2: Sample Collection

Authentic water samples were collected from two water treatment plants, E.L. Smith (ELS) and Rossdale (ROS), from February to May 2022. Raw water samples were collected on February 15, February 23, February 28, March 7, March 14, March 16, March 21, March 24, March 28, March 31, April 11, and May 5, 2022. In 2023, water samples were collected only from ELS but at two different treatment points (raw source water and clarifier effluent (ClarE)). Raw water is the water that is entering the treatment plant from the North Saskatchewan River, before any treatment steps. Clarifier Effluent (ClarE) samples are taken following primary treatment at the water treatment plant (WTP) which includes coagulation, flocculation, and PAC (when it is being used). Importantly, this sample is collected before any disinfection steps. Sampling dates in 2023 are as follows: February 23rd, March 2nd, March 9th, March 16th, arch 20th, March 23rd, March 30th, April 2nd (raw only), April 5th, April 8th, April 11th, April 13th. Samples collected on February 3rd (raw only), February 9th, and February 23rd were used in pooled samples. In both years, collection bottles were rinsed three times with the water being collected before the sample was collected and capped with no headspace. Samples were filtered using 1.5 µm glass microfiber filters, followed by 0.45 µm nylon membrane filters and stored at 4°C before analysis. The labelling reaction was complete on the water samples within three days of collection.

3.2.3: Reaction Conditions

A previously developed method for stable isotopic labelling was followed.[15] Solutions of CH_2O (1.8 M), CD_2O (1.8 M), and $NaBH_3CN$ (0.6 M) were prepared separately before each reaction. For the reaction, water was split into two 1 L parts. To one part, CH_2O was added to a final concentration of 3.6 mM and $NaBH_3CN$ to 1.2 mM. To the other part, CD_2O and $NaBH_3CN$ were added to concentrations of 3.6 mM and 1.2 mM, respectively. The reaction was left stirring for four hours, after which 2 mL of formic acid was added.

SPE was performed using Oasis MCX cartridges on a Supelco vacuum manifold. Cartridges were rinsed with methanol (2 mL) followed by water (4 mL, 0.2 % FA v/v) before sample loading. Samples (1 L) were passed through at a rate ~2-3 mL/min. Following the sample, the cartridge was rinsed with water (2 mL, 0.2 % FA v/v), and eluted with ammonium hydroxide solution (10 mL, 5 % wt. in methanol). Eluent was concentrated to 0.1 mL under nitrogen stream. Extracts of hydrogen and deuterium labelled, each ~0.1 mL, were mixed and filtered using a syringe filter. Finally, they were analyzed using high performance liquid chromatography (HPLC) - high resolution mass spectrometry (HRMS).

3.2.4: HPLC-HRMS Conditions

Separations were performed on an Agilent 1260 Series HPLC system, with a Luna C18 column (100 mm × 2 mm × 3 μm pore size). The injection volume was 20 μL. Mobile phases A and B were prepared as H_2O/ACN (95:5 v/v, 0.1 % FA) and ACN (0.1 % FA) respectively. The flow rate was set to 80 μL/min with a gradient elution as follows: 0 % B for 10 minutes, 0-30 % B over 20 minutes, 30-90 % B over 15 minutes, hold at 90 % for 10 minutes, decrease to 0 % B over 0.1 min, hold at 0 % B for 4.9 minutes. The column temperature was maintained at 30 °C.

Analysis was complete with a quadrupole time-of-flight mass spectrometer (Sciex QTOF x500R). The mass spectrometer was set to positive mode with an ion spray voltage of 5500 V. Other conditions are as follows: source gas 1(N_2, 35 arbitrary units), source gas 2 (N_2, 40 arbitrary units), curtain gas (N_2, 30 arbitrary units), temperature (500 °C), declustering potential (DP, 100 V), and collision energy (CE, 10 V). For the full scan, mass scan range from 50-1000 Da with an accumulation time of 0.25 s. For IDA analysis, a threshold intensity of 1000 cps was used to trigger MS/MS collection, and a maximum of 10 candidate ions monitored per cycle. The MS/MS mass range scan was from 20-1000 Da.

3.2.5: Pooled Source Water Sample Tests

Pooled source water samples were tested in 2023 to investigate sample degradation. Two types of pooled samples (pooled extracts and pooled source) from both 2022 and 2023 were analyzed with each set of spring runoff source water samples. The first type, pooled extracts, combined the final labelled extract, after labelling and SPE, from multiple sample dates. Specifically, the 2022 pooled extracts were made by taking 80 µL from extracts stored from 2022 sample dates and combining them into one vial. The extracts taken for the pooled sample had been run through the labelling procedure in 2022 and analyzed with the HPLC-HRMS method. The dates selected for the 2022 extracts include February 15th, March 14th, March 31st, March 28th, April 11th, and May 5th. The 2023 pooled extracts sample took 80 µL each of the February 3 raw, February 9th raw, and February 9th ClarE extracts.

The second type of pooled sample was pooled source. These were multiple water samples that were first combined and then run through the entire labelling procedure. The labelling reaction procedure was run in batches (2 L each) and the final extracts combined into one to generate enough volume to continue analysis throughout the spring. For the 2022 pooled source

sample, 500 mL of eight stored water samples from 2022 (February 28th, March 7th, March 14th, March 16th, March 24th, March 28th, March 31st, and April 11th) were combined. The 2022 pooled sample (4 L) was spiked with six amino acids (Leucine, Isoleucine, Tyrosine, Tryptophan, Threonine, and Phenylalanine) to a final concentration of 100 µg/L. Following, this sample underwent reactive nitrogen labelling using the normal labelling procedure described in **Section 3.2.3**. Because of the volume of the pooled sample, it was split in half and analyzed in two batches simultaneously before mixing and filtering the final extracts into one vial. Similarly, the 2023 pooled source sample was made by collecting a large volume of sample (20 L) on one date (February 23rd, 2023) which underwent reactive nitrogen labelling. Again, because the volume of the pooled sample the sample was split into four batches, ran through the normal labelling procedure, and the final extracts combined and filtered into one vial.

3.2.6: Data Analysis Parameters

Raw data (Sciex wiff2 files) were first converted to mzML using MSConvert (version 3.0). The MSConvert parameters can be seen in **Figure 3.2**. A desktop computer (i9-10900X 3.7GHz, 10 cores, 2x32 GB memory; Windows 10; 64-bit operating system) was used to process the data with HDPairFinder.R (https://github.com/HuanLab/HDPairFinder). The various tolerance values for HDPairFinder were left at default values and are listed in **Table 3.1**. The default values were chosen due to their effectiveness being proved during HDPairFinder development. The gap-filled data was used for further data analysis. For median fold change calculations, features with >50% of missing values were removed, replacement of the remaining missing values was done with 1/5th the minimum value of each H/D feature pair. Fold change was calculated for each H/D feature pair relative to the first sampling date (February 15th for

2022 and February 23rd for 2023). The median of all the features was taken for each sampling date.

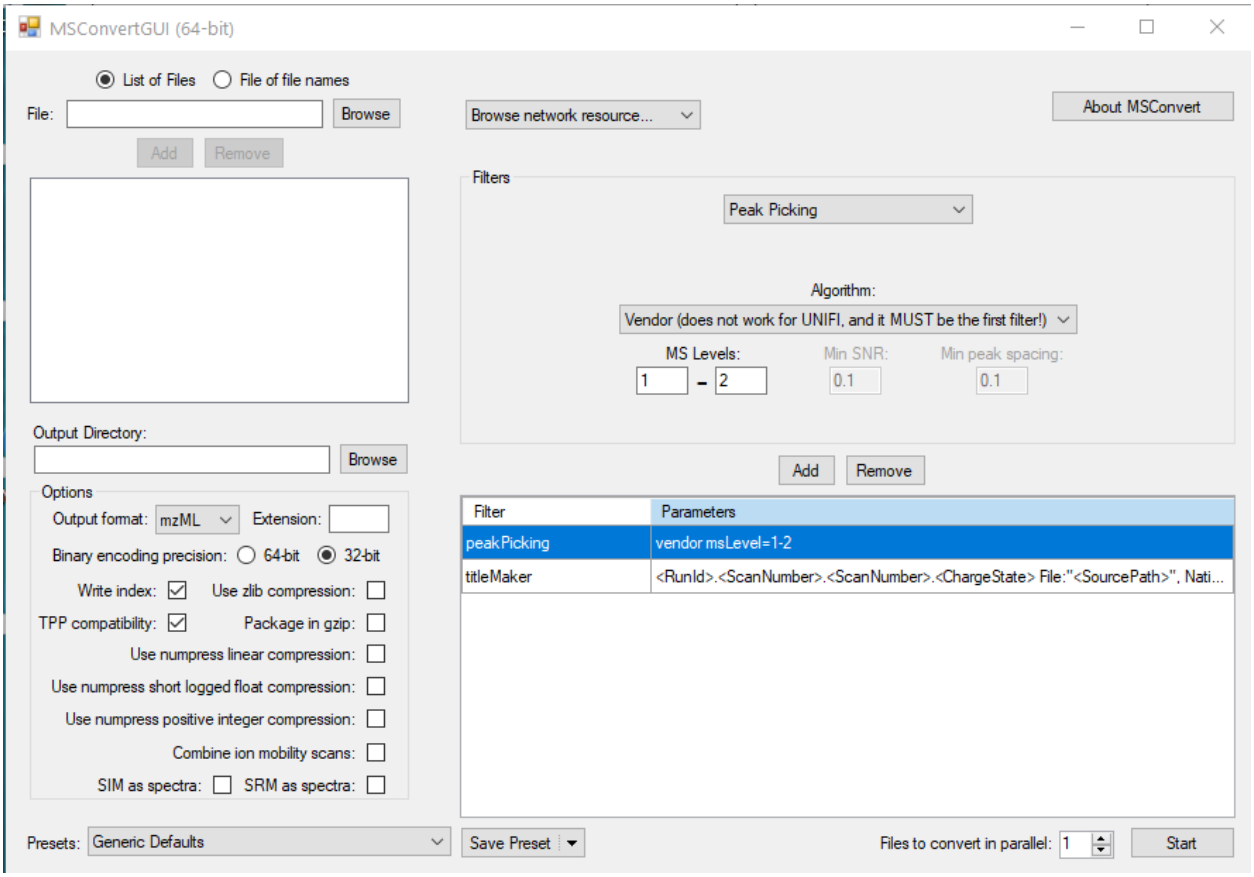


Figure 3.2- MSConvert Parameters.

Table 3.1- HDPairFinder Parameters.

Parameters	Value
heavy_mz_tol	20 ppm
rt_diff	-0.2 ~ 0.1 min
int_ratio	0.4 ~ 1.4
cc_threshold	0.7
run_inSourceFrag	True
align_mz_tol	50 ppm
align_rt_tol	0.5 min
gap_mz_tol	20 ppm
gap_rt_tol	0.5 min
int_threshold	1000
anno_mz_tol	30 ppm

3.2.7: Common Water Quality Parameter Analysis and Daily Mean Temperatures

Common water quality parameter data was obtained from EPCOR, the drinking water provider in Edmonton. Total Kjeldahl Nitrogen (TKN) is the sum of ammonia nitrogen and organic nitrogenous compounds. It is measured by the "Standard Methods for Examination of Water and Wastewater, APHA, AWWA and WEF, Washington DC, Method 4500Norg D." (approved in 1997, editorial revisions in 2011). It is measured in mg/L. Ammonia as Nitrogen (NH3-N), measures the amount of ammonia in a water sample. It is measured according to the "Standard Methods for Examination of Water and Wastewater, 4500 NH3 A, D" (Current version). It is measured in mg/L. Daily mean temperature data was obtained online from the Government of Canada, past weather and climate data. The data was obtained for the Edmonton Blachford station.[16]

3.3: Results and Discussion

3.3.1: Application of HDPairFinder to Authentic Water Samples

I first applied HDPairFinder to raw water from two treatment plants, ELS and ROS. Both treatment plants are in Edmonton and use the North Saskatchewan River as their source. Samples were collected from both treatment plants on the same dates in 2022. To look at changes over time in the H/D feature pairs I used median fold change relative to the February 15th sample. Both treatment plants showed a similar trend, showing a peak on March 28th, with ROS being slightly higher (**Figure 3.3**). Differences between the two trends could be attributed to city runoff, as previous studies have shown that runoff from within cities contributes some organic compounds.[17,18]

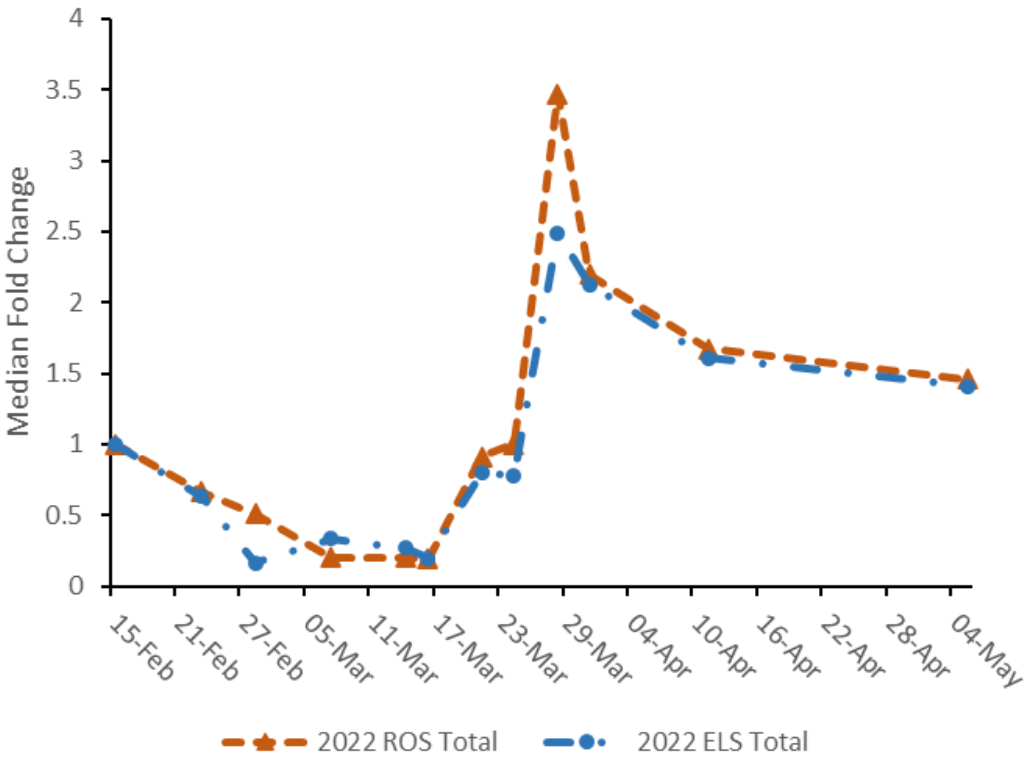


Figure 3.3- Median fold change for H/D feature pairs for the two water treatment plants, ELS and ROS, in 2022.

Next, I compared samples collected over two years. Specifically, raw water samples were collected from ELS in both 2022 and 2023 (**Figure 3.4)**. In 2022, there was an increase in the H/D feature pair intensity observed in mid to late March samples. This increase would be indicative of the influx of NOM occurring during the snowmelt. However, in 2023 there was no clear trend seen over the course of the spring. The lack of significant increase could be attributed to a more gradual reduction to the snowpack. In turn, this leaves less snow to contribute to the "big" melt in the spring and creates a milder spring runoff event. To verify the trends observed by the median fold change, common water quality parameters from the water treatment plant and daily mean temperature data from Environment Canada were obtained.

Two water quality parameters, TKN and NH3-N, are commonly used to measure nitrogen content in incoming raw water. Therefore, they would be a good comparison for the reactive nitrogen labelling method. In 2022, there was a significant increase in the TKN and NH3-N, whereas baseline levels were observed throughout the 2023 sampling period (**Figure 3.5**). This agrees with the trends observed for the reactive nitrogen compounds from the stable isotopic labelling method. When comparing the daily mean temperature data, a steeper increase in temperature in 2022 was observed compared to 2023 (**Figure 3.6**). In 2022, this indicates a fast snowpack melt, causing the sudden and large increase in the NOM. In 2023, the steep increase is not observed which would indicate a more gradual melt. Like the water quality parameter data, the daily mean temperature supports the trend observed in the stable isotopic labelling method.

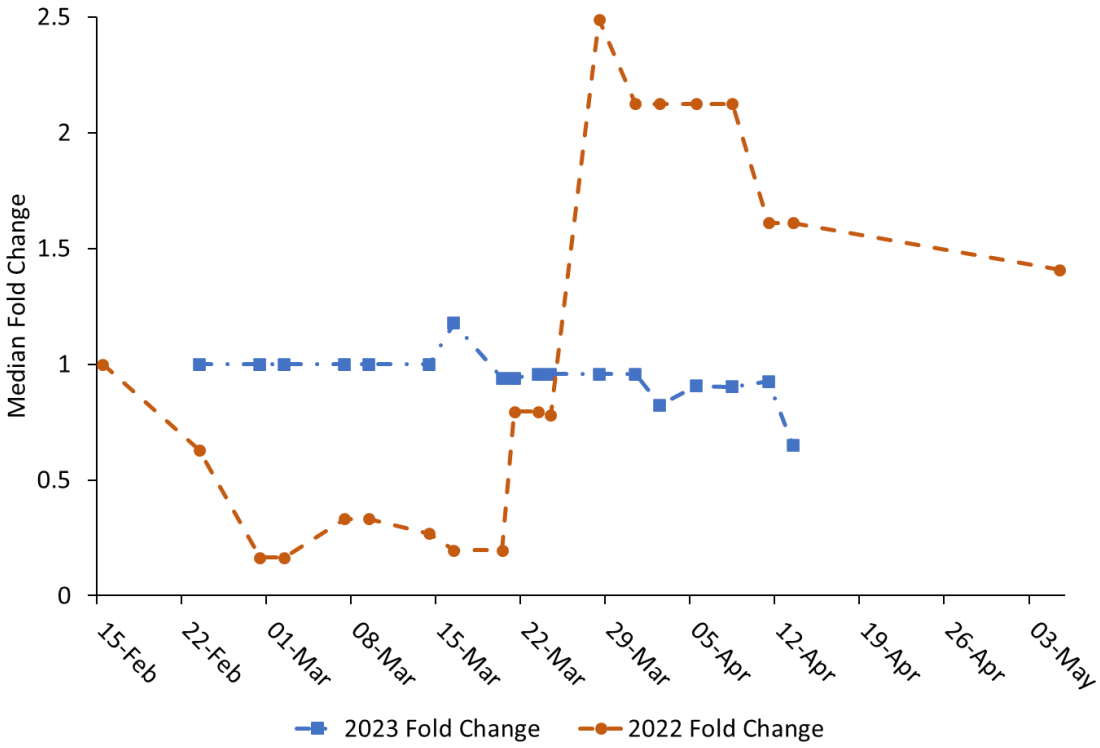


Figure 3.4- Median fold change of H/D feature pairs for 2023 and 2022 at the ELS treatment plant. Fold change is relative to February 15th for 2022 and February 23rd for 2023.

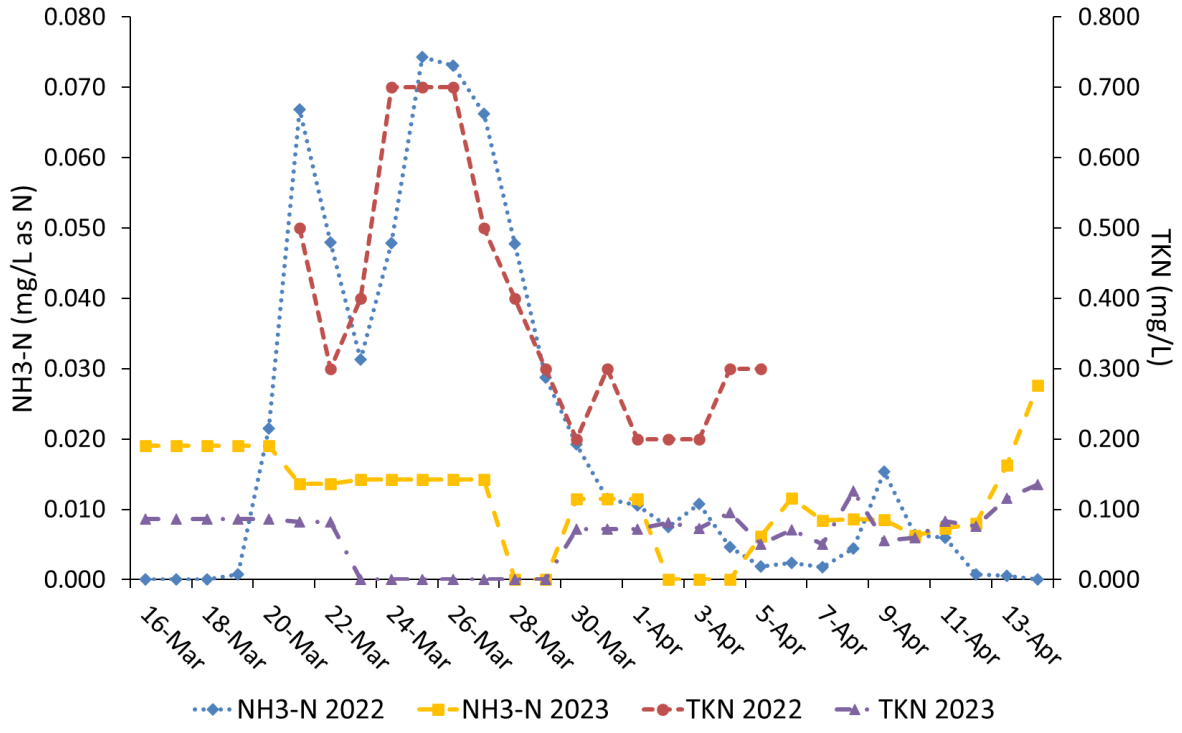


Figure 3.5- Common water quality parameters related to nitrogen, TKN and NH3-N, over 2022 and 2023.

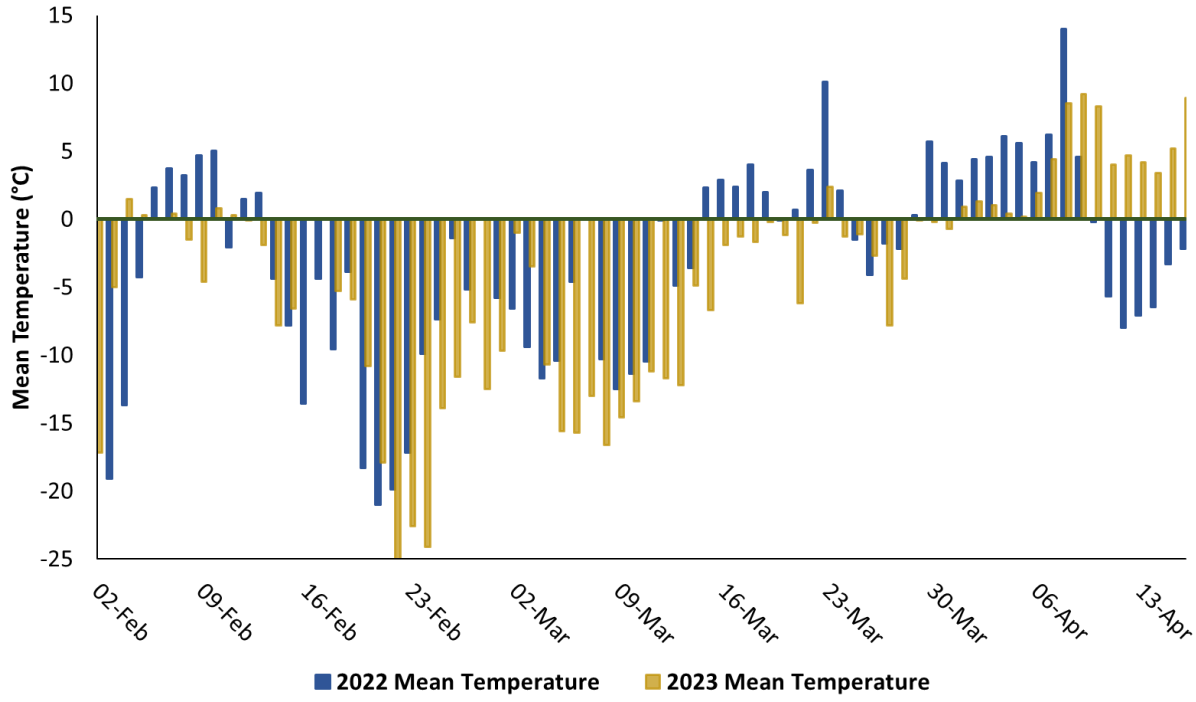


Figure 3.6- 2022 and 2023 daily mean temperature data over the course of sampling periods. Data obtained from Government of Canada, Historical Climate Data.[16]

We initially applied the HDPairFinder program to analyze non-target data of raw water samples. Here, I further applied this program to study the changes of N-containing compounds during the treatment process, specifically after the clarification step. In 2023 two types of water samples were collected, the first being the raw water entering the treatment plant and the second being the ClarE. ClarE samples were taken following primary treatment (coagulation and PAC when applicable) at the WTP. By tracking the changes over time, the removal of the reactive nitrogen compounds by primary treatment were examined. The median fold change for the 2023 raw and ClarE samples are presented in **Figure 3.7**. As the water quality parameters showed for 2023 (**Figure 3.5**), the composition of these samples would not be expected to be changing over time. The ClarE sample showed a median fold change stable around 1 for most of the sampling period. As this sample is taken following coagulation, it could be indicative of the consistent

removal of the organic compounds by this step. However, the ClarE did show a sharp drop in intensity on April 8th, which corresponds well to the beginning of PAC dosing at the water treatment plant on April 5th. As the PAC is added, a drop in the intensity is expected as the removal of the small organic compounds is improved. The ability to clearly recognize this trend shows the future applicability of this method to measure organic compounds removed during water treatment.

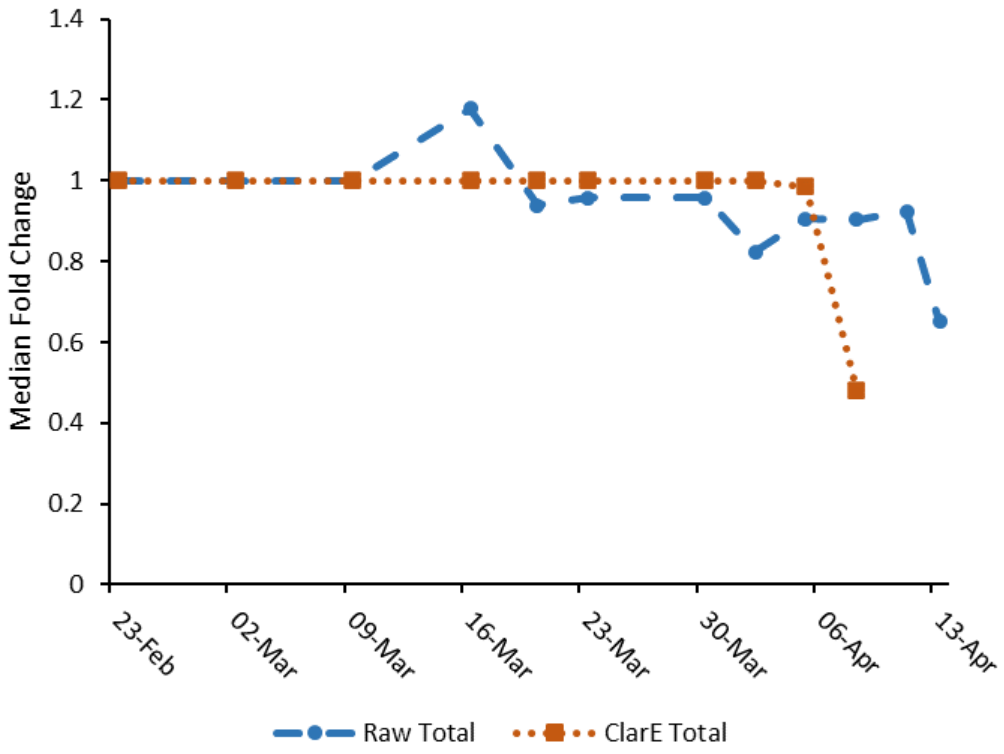


Figure 3.7- Median fold change for H/D feature pairs in raw and ClarE samples in 2023. Fold change is relative to February 23rd sample.

3.3.2: Characterization of Source Water over the Course of Spring Runoff

Following the application of HDPairFinder to look at overall trends, I attempted to broadly characterize the amine-containing organic matter present in the source water. A recent

study showed that disinfection by-products (DBPs) <1000 Da account for the highest toxicity risk posed by DBPs.[19] To understand the formation of these DBPs during water treatment, I looked at the organics in this range that could be acting as DBP precursors. I divided the <1000 Da mass range further to identify what may be the most significant contributor to changes over spring runoff. To do this, I chose to split the H/D feature pairs into three range; unlabelled m/z ratios of <250 which captures mainly amino acids, 250-500 which includes many dipeptides, and 500-1000 which would include polypeptides. Other small nitrogen-containing organic compounds would be expected to be found in all the ranges.

First, I looked at the median fold change for the unlabelled m/z of <250 across the two sampling years. In this range a large spike was observed in 2022 and a small spike in 2023 (**Figure 3.8**). Previous studies have shown small water-soluble compounds, likely in this mass range, increase before the large influx of the remaining NOM.[7,8] Specifically, amino acids are found in this range and have been shown in studies done by our group to increase during the spring in the North Saskatchewan River.[20] Interestingly, the spike in this specific mass range can be seen during spring runoff in both 2022 and 2023. As this mass range was shown to have small increases before the large increase of organic matter, it could indicate that compounds in this range could be used as indicators for the start of a spring runoff event. This supports previous studies done by our group predicting amino acids to be a marker of spring runoff. [20] However, it did not account for the largest changes seen during the spring.

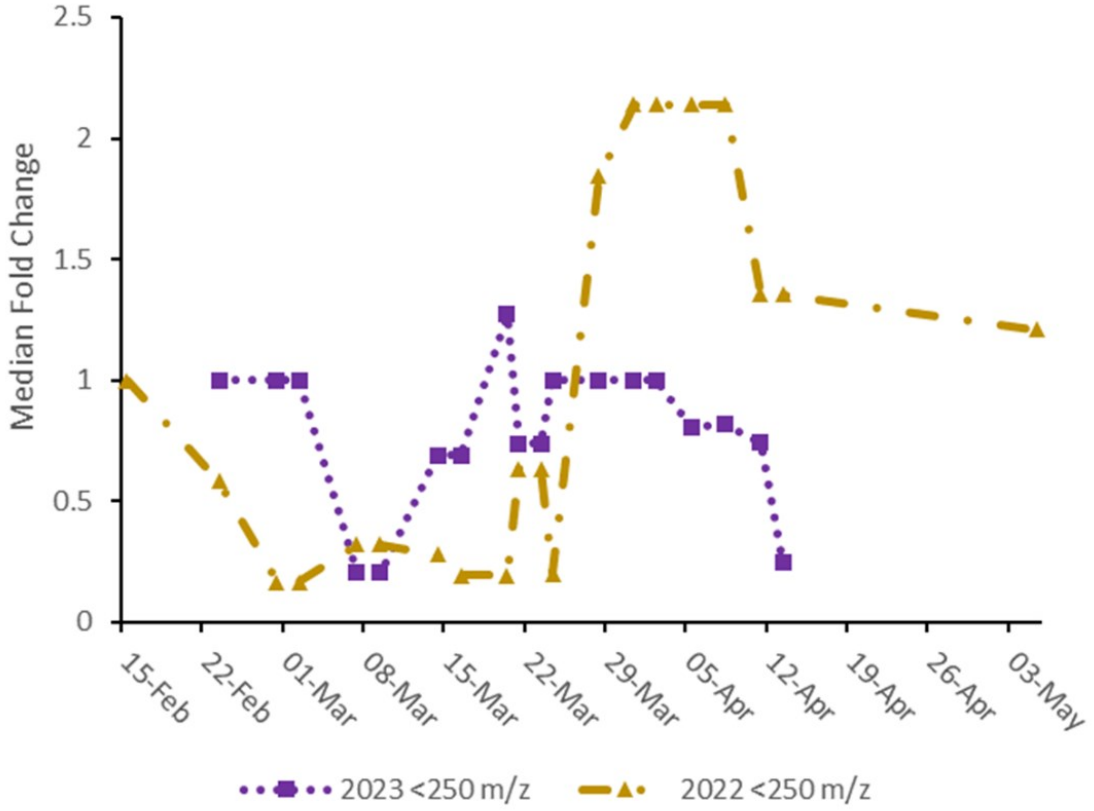


Figure 3.8- 2022 and 2023 ELS median fold change for H/D feature pairs with an unlabelled m/z ratio of <250. Fold change is relative to February 15th for 2022 and February 23rd for 2023.

Next, I looked at the median fold change for the unlabelled m/z of 250-500 range. In 2022 clear spikes could be seen in the middle of spring whereas 2023 showed little change (**Figure 3.9**). In 2022, the median fold change for this range and water quality parameters showed a similar trend. Both can be seen to start increasing around March 20th and start returning to baseline around March 30th (**Figure 3.9**/**Figure 3.5**). Therefore, this could suggest that this range is making up significant portion of the amine-containing NOM in spring runoff events where increases in the water quality parameters are observed. Additionally, this trend agreed with the increase of temperatures in 2022, indicating that the melting event would be the contributor

of this mass range (**Figure 3.9/Figure 3.6**). However, there was still the larger mass range that required consideration.

Next, I looked at the unlabelled m/z range of 500-1000. In both 2022 and 2023 this region only contained only a few H/D feature pairs but had the highest median fold change value (**Figure 3.10**). In both years, the features with an unlabelled m/z greater than did not follow the same trend as the other mass ranges. Additionally, it did not follow a similar trend to the water quality parameters or the daily mean temperature data. By identification of these compounds, it could allow for better explanations of why the difference in patterns and higher median fold change are occurring when compared to the other ranges. For example, the possibility of their degradation before entering the water treatment plant. Therefore, the identification of the compounds isotopically labelled in this mass range should be a priority to identify in the future.

When comparing all the different unlabelled mass ranges, the overall trends followed the closest to m/z of 250-500 range (**Figure 3.4/3.9**). Additionally, the similarity in trend to the water quality parameters highlights the impact that this range could be having on the water treatment plant. One possible explanation for this trend is dipeptides and organic compounds in this region could be higher in abundance. As this region seems to be having a significant impact and showed clear differences between the two years, future investigation into the identity of these compounds would be warranted. This could provide insight into if there are unique compounds changing in 2022, or if they are the same and just changing to a higher degree. As the changes in the incoming source water during spring runoff had been identified, the next step was to see the fate of these different mass ranges during water treatment.

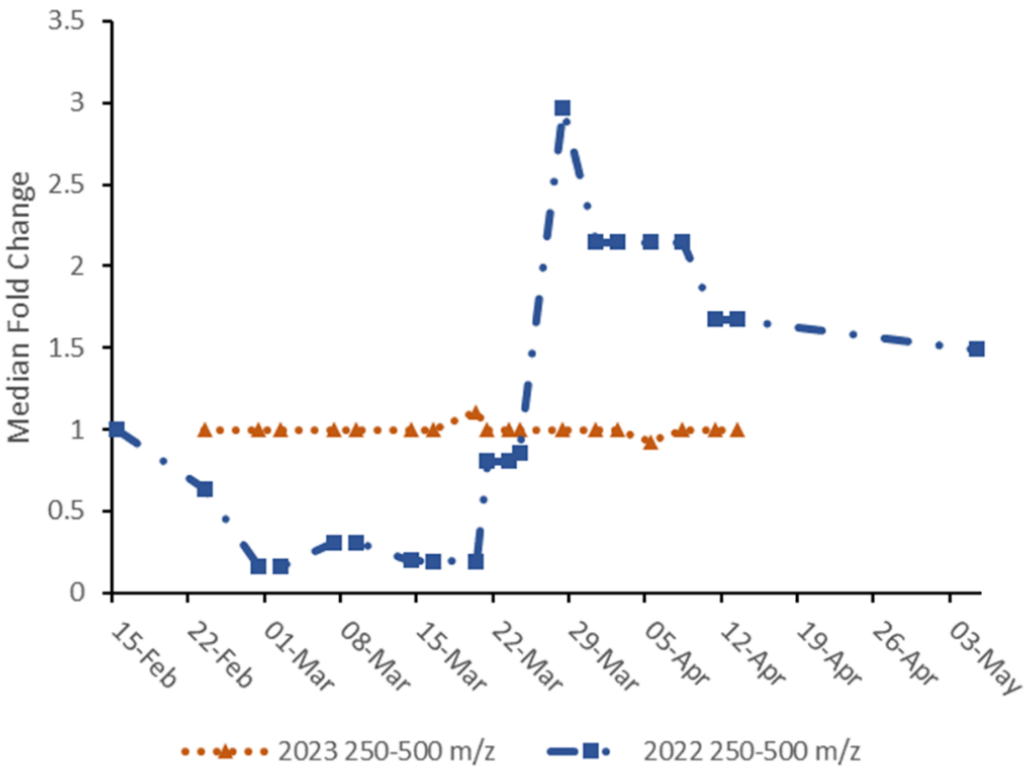


Figure 3.9- 2022 and 2023 ELS median fold change for H/D feature pairs with an unlabelled m/z ratio of 250-500. Fold change is relative to February 15th for 2022 and February 23rd for 2023.

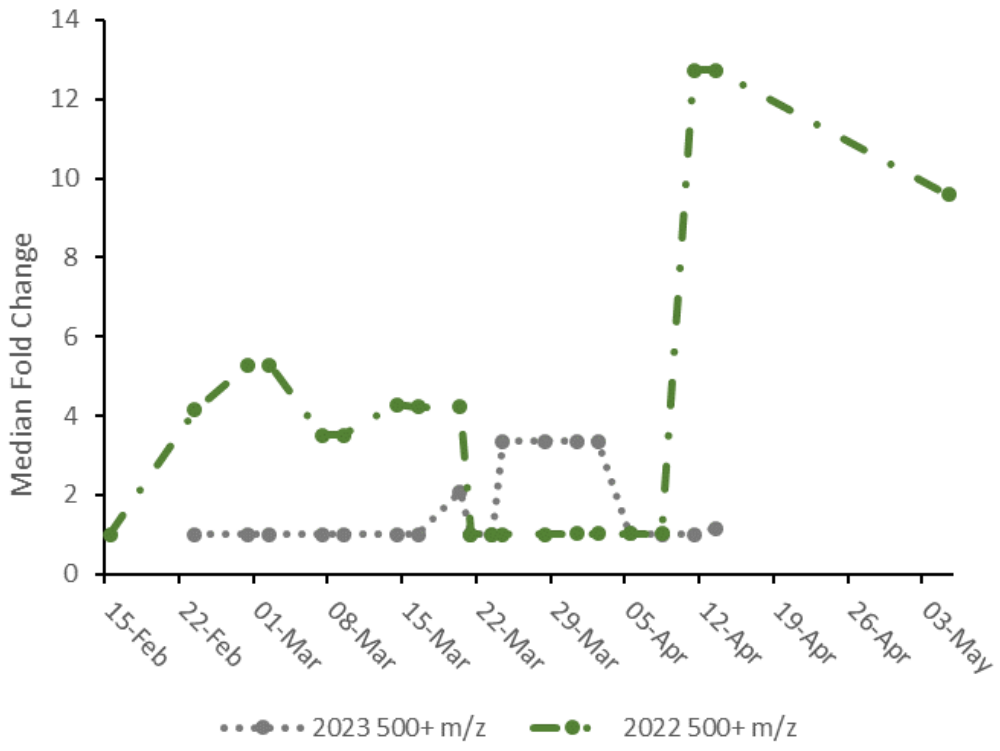


Figure 3.10- 2022 and 2023 ELS median fold change relative for H/D feature pairs with an unlabelled m/z ratio of >500. Fold change is relative to February 15th for 2022 and February 23rd for 2023.

To look at how primary water treatment steps are removing the nitrogen containing organics, I compared the different mass ranges for the raw and ClarE samples in 2023. The unlabelled m/z range of <250 was seen to change more in the ClarE compared to the raw sample. In the ClarE sample, a spike can be seen followed by a significant drop **(Figure 3.11c)**. This small portion will increase in the spring, as they are expected to enter source water with the snowmelt. The increase is then observed until the PAC addition, beginning on April 5th. Therefore, it could be predicted that these compounds are not removed well during the coagulation steps and therefore PAC addition during the spring is required for their removal.

When looking at the 500-1000 range various spikes in both the raw and ClarE were observed (**Figure 3.1a**). With a spike observed in the ClarE, we can predict that these compounds are not entirely removed during coagulation. Investigation into the identify of these compounds could provide insight into why the large fold change observed. The overall trend in median fold change followed the closest to the 250-500 range (**Figure 3.11b/d**). For the 250-500 range the median fold change for the ClarE sample remained constant until the addition of PAC. This indicates that the coagulation would consistently remove this range, however, PAC allows for additional removal. Further application of this method could lead to identification compounds that are making it through these initial water treatment steps and could be forming DBPs. Although these mass ranges can provide insight into the potential groups of compounds of interest, further information on the structure of these compounds was investigated.

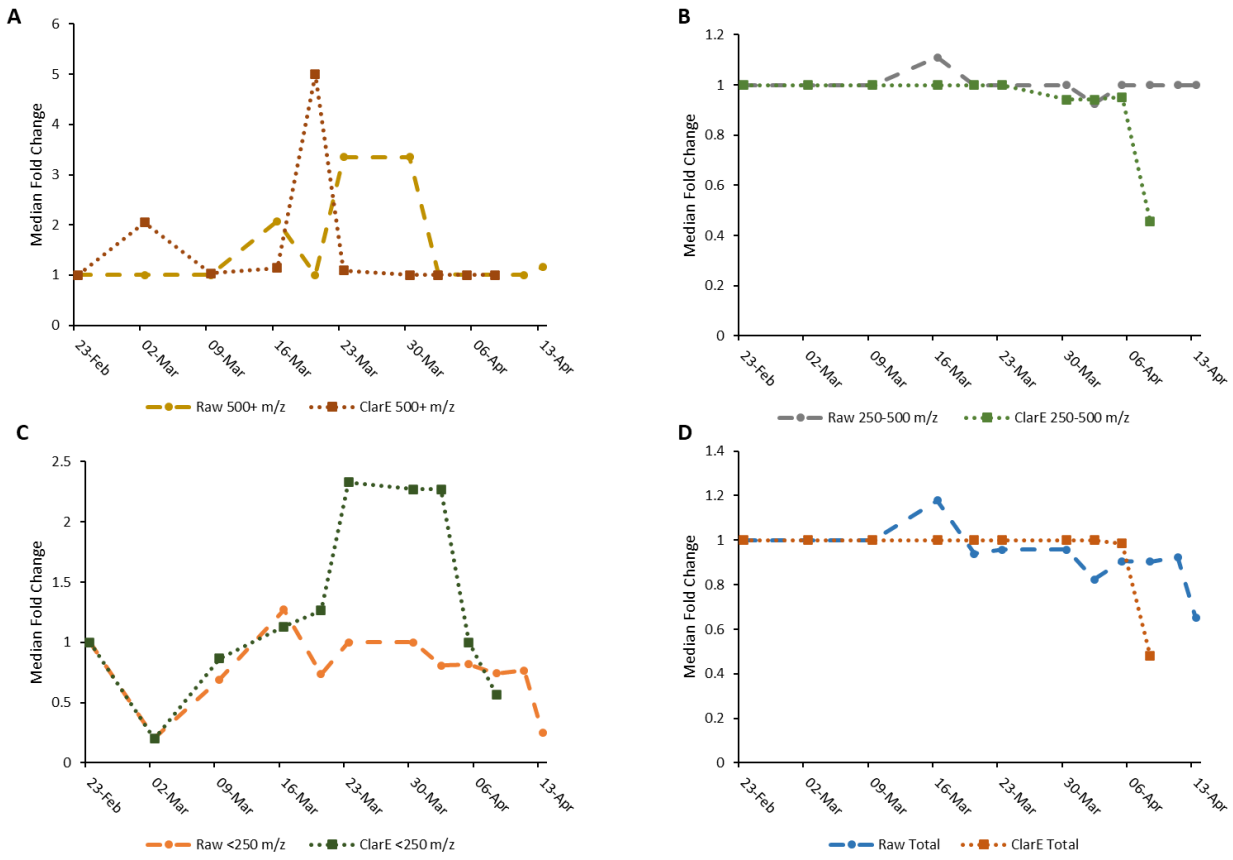


Figure 3.11- Median fold change relative to February 23rd sample for raw and ClarE samples for different mass ranges A) H/D feature pairs with unlabelled m/z ratio of >500 B) H/D feature pairs with unlabelled m/z ratio of 250-500 C) H/D feature pairs with unlabelled m/z ratio of <250 D) total median fold change for the combination of all mass ranges.

Another way I attempted to characterize the water it by looking at the number of methyl groups, or tags, that reacted during the stable isotopic labelling reaction. The number of methyl groups can be determined from the mass difference used to detect the H/D feature pair in the data analysis. This can characterize what type of amine is reacting, as shown in **Figure 3.12**. For all sample types and locations, the 1 and 2 tags made up the most significant portion (**Figure 3.13)**. These groups are generally associated with the smaller amino acids and peptides which allows

shows that these small peptides are the most significant portion being captured by this method.[15] Overall, the combination of different mass ranges and number of tags highlights the importance of the small amino containing compounds for their contributions during spring runoff.

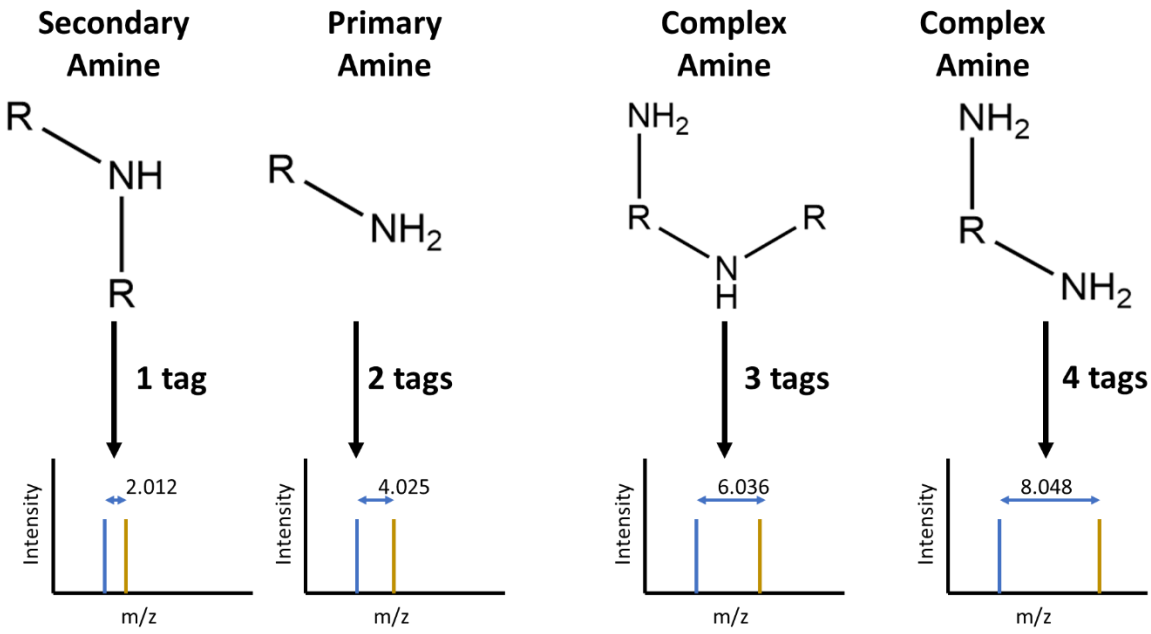


Figure 3.12- Number of methyl groups (tags), added to different types of amines.

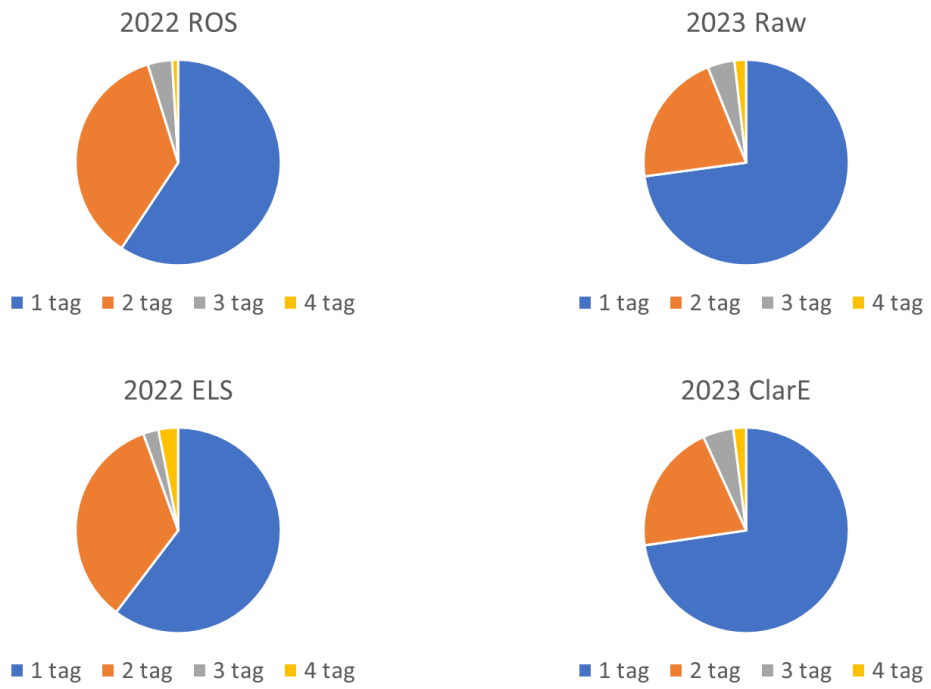


Figure 3.13- Number of methyl-labelled tags attached to each H/D feature pair for different groups of samples.

3.3.3: Pooled Source Water Samples and Future Considerations

Various pooled samples were tested in 2023 for preliminary investigation of sample decomposition. Four types of pooled samples were used, as described in **Section 3.2.5**. In brief, the four types of pooled samples were 2023 and 2022 pooled source and 2023 and 2022 pooled extracts. The pooled source ran multiple labelling reactions on the same sample and combined them into one vial. The pooled extracts took multiple samples where the labelling method had been previously complete and combined small volumes of each. In each case, I compared the median fold change over time.

Overall, no clear pattern among the four pooled samples could be observed (**Figure 3.14**). With no consistent trend, there is no indication that instrument response had significantly

changed on a particular day. When looking at each pooled sample, the 2023 pooled extracts showed the highest consistency across days. This sample was made by using a newly collected source water sample and running through the labelling method immediately. Therefore, it could be expected to have less degradation compared to the 2022 samples which had been stored for a year before the pooled analysis. However, a steep drop in intensity was observed in the 2023 pooled source. This could be because of potential decomposition of the sample.

However, further analysis with better instrument response controls will be required for decomposition testing. For example, using a targeted method to track the methyl-labelled compounds over time. Another method that could be applied in the future for testing sample decomposition is collecting data using multiple injection volumes of the same sample. By collecting multiple injection volumes other available programs, such as MAFFIN, could be applied compare data using among analysis dates.[21] Multiple injection volume data also allows for the determination of "true" features. However, because of the limited volume of the pooled samples made in 2023, this data was not collected. The pooled samples should be further investigated in the future with the additional data to gain a more comprehensive understanding of sample degradation.

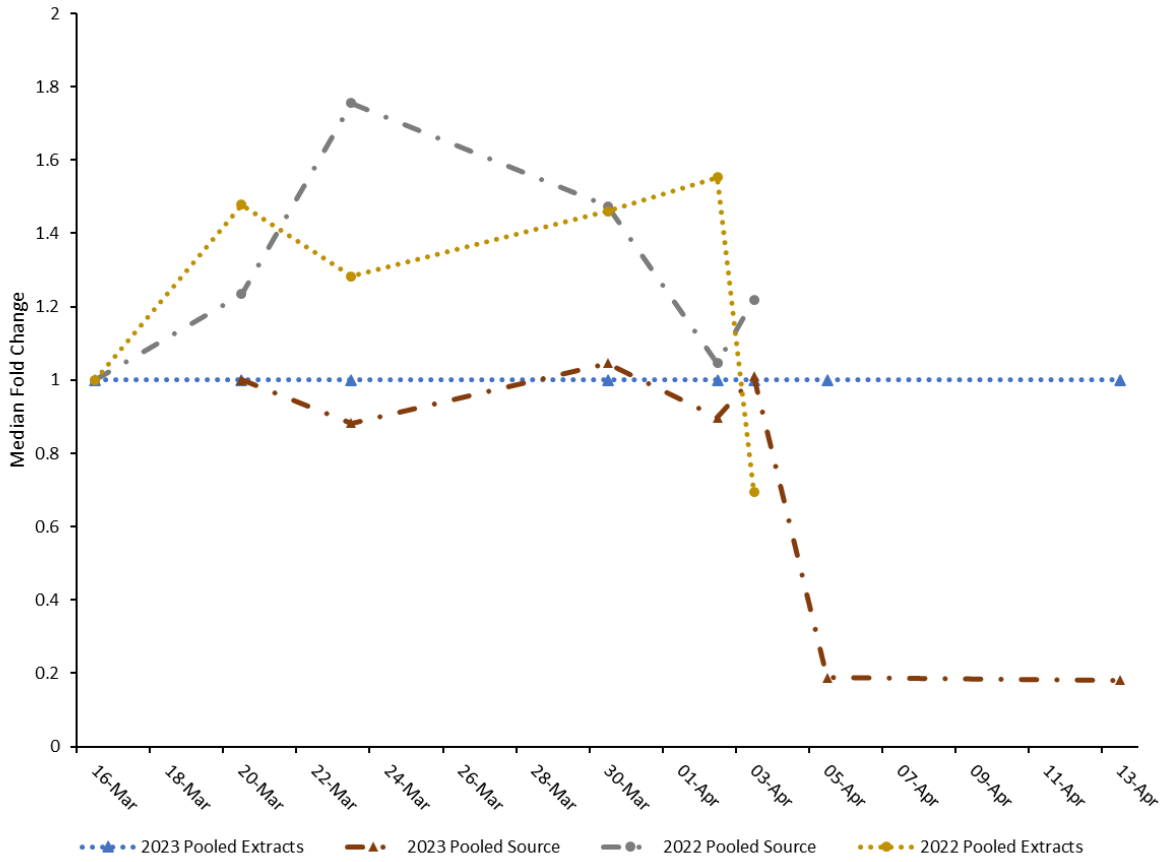


Figure 3.14- Median fold change for H/D feature pairs in pooled source water samples ran alongside 2023 samples.

3.4: Conclusion

In this chapter, HDPairFinder was applied to look for trends across different sampling sites, different years, and different types of water samples. The median fold change of the two sampling locations in Edmonton showed similar trends, with differences potentially accounted for by city runoff. The two different sampling years showed different trends, with 2022 showing a large spike in the median fold change and 2023 remaining constant. These differences between sampling years were confirmed using common water quality parameters and temperature data. Additionally, HDPairFinder was applied to study the removal of N-containing compounds in

ClarE samples, where primary water treatment had taken place. This data showed a drop in the N-containing organics following the addition of PAC. These results highlight the importance of PAC addition for efficient NOM removal and control of DBP formation.

Further application of HDPairFinder for identification of N-containing organics could be done in the future. In particular, the tentative identification feature provided by HDPairFinder could provide direction into what standards to order and test, ultimately leading to identification of some of compounds of interest. Following identification targeted methods could be developed to quantify potential compounds of interest. Further improvements could be made in the understanding of the decomposition of labelled extracts, as limitations remained with the pooled samples. Overall by improving the characterization of these source water samples we increase our understanding of NOM present in source water.

3.5: References

(1) Zhao, T.; Carroll, K.; Craven, C. B.; Wawryk, N. J. P.; Xing, S.; Guo, J.; Li, X.; Huan, T. HDPairFinder : A Data Processing Platform for Hydrogen / Deuterium Isotopic Labeling-Based Nontargeted Analysis of Trace-Level Amino-Containing Chemicals in Environmental Water. *J. Environ. Sci*. **2024**, 136, 583–593. https://doi.org/10.1016/j.jes.2023.02.033.

(2) Tautenhahn, R.; Bottcher, C.; Neumann, S. Highly Sensitive Feature Detection for High Resolution LC/MS. *BMC Bioinformatics* **2008**, 9, 1–16. https://doi.org/10.1186/1471-2105-9-504.

(3) Guo, J.; Huan, T. Mechanistic Understanding of the Discrepancies between Common Peak Picking Algorithms in Liquid Chromatography-Mass Spectrometry-Based Metabolomics. *Anal. Chem*. **2022**. 95(14), 5894–5902. https://doi.org/10.1021/acs.analchem.2c04887.

(4) Guo, J.; Shen, S.; Xing, S.; Yu, H.; Huan, T. ISFrag: De Novo Recognition of In-Source Fragments for Liquid Chromatography-Mass Spectrometry Data. *Anal. Chem*. **2021**, 93 (29), 10243–10250. https://doi.org/10.1021/acs.analchem.1c01644.

(5) Huan, T.; Li, L. Counting Missing Values in a Metabolite-Intensity Data Set for Measuring the Analytical Performance of a Metabolomics Platform. *Anal. Chem*. **2015**, 87 (2), 1306–1313. https://doi.org/10.1021/ac5039994.

(6) Hu, Y.; Cai, B.; Huan, T. Enhancing Metabolome Coverage in Data-Dependent LC-MS/MS Analysis through an Integrated Feature Extraction Strategy. *Anal. Chem*. **2019**, 91, 14433-14441. https://doi.org/10.1021/acs.analchem.9b02980.

(7) Sebestyen, S. D.; Boyer, E. W.; Shanley, J. B.; Kendall, C.; Doctor, D. H.; Aiken, G. R.; Ohte, N. Sources, Transformations, and Hydrological Processes That Control Stream Nitrate and Dissolved Organic Matter Concentrations during Snowmelt in an Upland Forest. *Water Resources Research* **2008**, 44, 1–14. https://doi.org/10.1029/2008WR006983.

(8) Du, Z.; Ding, S.; Xiao, R.; Fang, C.; Song, W.; Jia, R.; Chu, W. Does Snowfall Introduce Disinfection By-Product Precursors to Surface Water ? *Environ. Sci. Technol*. **2022**, 56, 14487-14497. https://doi.org/10.1021/acs.est.2c04408.

(9) Krasner, S. W.; Roback, S.; Qian, Y.; Li, X. F.; Marfil-Vega, R.; Bukhari, Z. Occurrence of Nitrosamines and Their Precursors in North American Drinking Waters. *AWWA Water Sci*. **2020**, 2 (6), 1–15. https://doi.org/10.1002/aws2.1208.

(10) Hrudey, S. E.; Rector, D.; Motkosky, N. Characterization of Drinking Water Odour Arising from Spring Thaw for an Ice-Covered Upland River Source. *Wat. Sci. Tech*. **1992**, 25 (2), 65–72.

(11) Hrudey, S. E. Chlorination Disinfection By-Products, Public Health Risk Tradeoffs and Me. *Water Res.* **2009**, 43 (8), 2057–2092. https://doi.org/10.1016/j.watres.2009.02.011.

(12) Lee, W.; Westerhoff, P.; Esparza-Soto, M. Occurrence and Removal of Dissolved Organic Nitrogen in US Water Treatment Plants. *AAWA*. **2006**, 98(10), 102–110. https://doi.org/10.1002/j.1551-8833.2006.tb07782.x.

(13) Chen, C.; Zhang, X.; Zhu, L.; He, W.; Han, H. Changes in Different Organic Matter Fractions during Conventional Treatment and Advanced Treatment. *J. Environ. Sci*. **2011**, 23 (4), 582–586. https://doi.org/10.1016/S1001-0742(10)60423-8.

(14) Andersson, A.; Lavonen, E.; Harir, M.; Gonsior, M.; Hertkorn, N.; Schmitt-Kopplin, P.; Kylin, H.; Bastviken, D. Selective Removal of Natural Organic Matter during Drinking Water Production Changes the Composition of Disinfection By-Products. *Environ. Sci. Water Res. Technol*. **2020**, 6(3), 779–794. https://doi.org/10.1039/c9ew00931k.

(15) Liu, Z.; Craven, C. B.; Huang, G.; Jiang, P.; Wu, D.; Li, X. F. Stable Isotopic Labeling and Nontarget Identification of Nanogram/Liter Amino Contaminants in Water. *Anal. Chem*. **2019**, 91 (20), 13213–13221. https://doi.org/10.1021/acs.analchem.9b03642.

(16) *Government of Canada: Past Weather and Climate Data*. https://climat.metro.gc.ca/historical_data/search_historic_data_e.html (Accessed Jan 3, 2023).

(17) Telang, S. A. Effects of Reservoir-Dam, Urban, Industrial, and Sewage Treatment Run-off on the Presence of Oxygen and Organic Compounds in the Bow River. *Water. Air. Soil Pollut*. **1990**, 50 (1–2), 77–90. https://doi.org/10.1007/BF00284785.

(18) Seiwert, B.; Nihemaiti, M.; Troussier, M.; Weyrauch, S.; Reemtsma, T. Abiotic Oxidative Transformation of 6-PPD and 6-PPD Quinone from Tires and Occurrence of Their Products in Snow from Urban Roads and in Municipal Wastewater. *Water Research*. **2022**, 212, 118122. https://doi.org/10.1016/j.watres.2022.118122.

(19) Mitch, W. A.; Richardson, S. D.; Zhang, X.; Gonsior, M. High-Molecular-Weight by-Products of Chlorine Disinfection. *Nat. Water* **2023**, 1, 336-347. https://doi.org/10.1038/s44221-023-00064-x.

(20) Qiu, J.; Craven, C.; Wawryk, N.; Carroll, K.; Li, X.-F. Integration of Solid Phase Extraction with HILIC-MS/MS for Analysis of Free Amino Acids in Source Water. *J. Environ. Sci. (China)* **2022**, 117, 190-196. https://doi.org/10.1016/j.jes.2022.04.025.

(21) Yu, H.; Huan, T. Systems Biology MAFFIN : Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities. *Bioinformatics* **2022**, 38 (13), 3429–3437. https://doi.org/10.1093/bioinformatics/btac355.

Chapter 4

Conclusions and Future Work

4.1: Conclusions

4.1.1: Manual Data Analysis and Development of HDPairFinder

Application of a stable isotopic labelling method is effective in studying reactive amine-containing compounds in source water is through the application of a stable isotopic labelling method.[1] This method uses methyl-based labelling to look for specific isotopic patterns in high resolution mass spectrometry (HRMS) data. While this provides a sensitive detection method for amine-containing compounds, the data analysis is slow and challenging. As the specific isotopic patterns are searched for manually, it can be time consuming. It is also unable to fully create lists of exhaustive peaks and can miss compounds if retention times are slightly offset. Therefore, an automated program to assist in data analysis is needed. Various programs had been previously developed to aid in the data analysis of HPLC-HRMS data. While programs such as OpenMS and MZmine have been applied in the past, they lack capabilities to identify specific isotopic patterns in data.[2,3] Therefore, a newly developed program was required. This led to the development of HDPairFinder, a fully automated analysis program to identify the isotopic pattern associated with the methyl-based labelling procedure.[4]

HDPairFinder provides many improvements to the data analysis procedure. Processing time for samples is reduced and data cleaning helps to provide higher quality data. With the integration of alignment features, a single output table is given for peak intensity across multiple samples. An additional module incorporated into HDPairFinder allows for tentative identification of amine containing compounds based on an accurate mass match. Overall, HDPairFinder helps

in providing a streamlined data workflow that can aid in the identification of new nitrogen containing species in source water.

4.1.2: Application of HDPairFinder to Real Samples

Following the development of HDPairFinder, I applied it to source water samples across two years. Spring runoff samples were analyzed with the goal of detecting changes in reactive nitrogen containing compounds over the course of spring. This analysis could provide valuable information regarding the influx of natural organic material occurring with the snow and ice melt.[5] Through the study of different treatment plants, years, and types of samples we can improve our understanding of the characterization of incoming source water. This knowledge can help to direct the areas of interest for further investigation for their impact over the course of spring.

The analysis of two different sampling years showed how different spring runoff events may be occurring. In 2022, a spike in the reactive nitrogen containing compounds was observed, whereas 2023 remained constant. Common water quality parameters and daily mean temperatures were found to be in agreeance with the data collected using the stable isotopic labelling method. By having a complete peak list I was also able to analyze different mass ranges to help characterize the source water. Although each range showed different trends, a m/z ratio of 250-500 seemed to have the most significant contribution to the overall trend. Through the application of a stable isotopic labelling method in conjunction with data analysis using HDPairFinder, I was able to study the spring runoff changes to amine-containing compounds in source water. This work could help to provide direction in targeted areas of study for water treatment in the future.

4.2: Future Work

4.2.1: Future Application of HDPairFinder

In this work, I was able to apply HDPairFinder to a selection of different sample locations, types, and years. While this helped to provide some preliminary data, further application could still be done. As each spring runoff can vary, the application in future years can help to provide more insight into these melting events. In particular, the continued collection of clarifier effluent samples in years where water quality parameter changes are observed, could help to provide insight into how well organics are removed during water treatment. To confirm the changes in results, samples collected throughout the year could be analyzed. This could help to provide a larger picture of the seasonal variation of each of these compounds compared to just over the course of the spring. Additionally, the wider application of HDPairFinder could be done for source water samples from other geographical locations. This could help to provide insight into how amine containing compounds in source water change with the environment.

As HDPairFinder is further applied, some additional considerations towards analysis should be made. For example, triplicates for each sample should be run to help in later statistical analysis. Additionally, different quality controls to help with sample normalization should be carefully considered and applied. Although fold change can help to account for some of this and help with looking at overall trends, it is still far from ideal. Internal standards could be further investigated, although they are still limited as commonly used ones are deuterated. This may have the potential to impact the programs' ability to detect real features so needs careful application before implementation. Pooled samples were run in 2023, however, sample degradation could be better systematically studies. Additionally, to use many normalizations software, it is required to either have a serial dilution through multiple injection volumes, or to

have multiple replicates. This was limited in 2023 due to the amounts of samples provided but should be considered in any future work.

4.2.2: Confirmation of Tentative Identifications

The tentative identification feature of the software allows for a quick idea of what compounds may be present in the source water. When used in parallel with information such as mass to charge ratio and number of tags it can help to provide information on the potential identity of the compound. However, without running a standard, it still is not a definitive identification.[6] To help with this, more standards could be run through the entire labelling procedure to check for identifications. Where possible, purchasing already synthesized standards would be helpful, however it is limited as the addition of the methyl label, especially with deuterium is not common. Therefore the ability to confirm the identify of a peak with high certainty is limited. However, by running available standards through the labelling procedure and matching to the non-target data, a direction for targeted methods could be identified. With the identification of new compounds, a more comprehensive understanding of the composition of source water could be gained.

4.3: Implications of Work

The development and application of HDPairFinder has been able to help improve data analysis for the analysis of nitrogen containing compounds in source water. By improving our knowledge of these compounds in source water, we can help to improve water treatment in the future. With the recognition of compounds of significance in specific mass ranges and work towards specific identification, we can help direct future targeted research. By gaining this knowledge we can show how effective water treatment is in removing these organic compounds.

Additionally, in their identification we can look at the potential of these organics to form disinfection by-products and how to minimize their formation.

Specifically focussing on nitrogen containing compounds will become more important in the future. As global warming progresses and water reuse increases, nitrogen in drinking water treatment is expected to increase. This has been shown from previous studies as the use of potable reuse water increases nitrogen content in source water.[7–9] By having a method that can help in analyzing these compounds, we will be able to track their potential impacts and understand how they are changing in the future. Overall, this work will help in ensuring the ongoing safety of our drinking water.

4.4: References

(1) Liu, Z.; Craven, C. B.; Huang, G.; Jiang, P.; Wu, D.; Li, X. F. Stable Isotopic Labeling and Nontarget Identification of Nanogram/Liter Amino Contaminants in Water. *Anal. Chem.* **2019**, 91 (20), 13213–13221. https://doi.org/10.1021/acs.analchem.9b03642.

(2) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H. C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat. Methods* **2016**, 13 (9), 741–748. https://doi.org/10.1038/nmeth.3959.

(3) 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.

(4) Zhao, T.; Carroll, K.; Craven, C. B.; Wawryk, N. J. P.; Xing, S.; Guo, J.; Li, X.; Huan, T. HDPairFinder : A Data Processing Platform for Hydrogen / Deuterium Isotopic Labeling-Based Nontargeted Analysis of Trace-Level Amino-Containing Chemicals in Environmental Water. *J. Environ. Sci*. **2024**, 136, 583–593. https://doi.org/10.1016/j.jes.2023.02.033.

(5) Sebestyen, S. D.; Boyer, E. W.; Shanley, J. B.; Kendall, C.; Doctor, D. H.; Aiken, G. R.; Ohte, N. Sources, Transformations, and Hydrological Processes That Control Stream Nitrate and Dissolved Organic Matter Concentrations during Snowmelt in an Upland Forest. *Water Resources Research* **2008**, 44, 1–14. https://doi.org/10.1029/2008WR006983.

(6) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol*. **2014**, 48 (4), 2097–2098. https://doi.org/10.1021/es5002105.

(7) Lau, S. S.; Bokenkamp, K.; Tecza, A.; Wagner, E. D.; Plewa, M. J.; Mitch, W. A. Toxicological Assessment of Potable Reuse and Conventional Drinking Waters. *Nat. Sustain*. **2023**, 6 (1), 39–46. https://doi.org/10.1038/s41893-022-00985-7.

(8) Scheurer, M.; Brauch, H. J.; Lange, F. T. Analysis and Occurrence of Seven Artificial Sweeteners in German Waste Water and Surface Water and in Soil Aquifer Treatment (SAT). *Anal. Bioanal. Chem*. **2009**, 394 (6), 1585–1594. https://doi.org/10.1007/s00216-009-2881-y.

(9) Peng, J.; Huang, H.; Zhong, Y.; Yin, R.; Wu, Q.; Shang, C.; Yang, X. Transformation of Dissolved Organic Matter during Biological Wastewater Treatment and Relationships with the Formation of Nitrogenous Disinfection Byproducts. *Water Res*. **2022**, 222 (7), 118870. https://doi.org/10.1016/j.watres.2022.118870.

Bibliography

Andrade-Eiroa, A.; Canle, M.; Leroy-Cancellieri, V.; Cerdà, V. Solid-Phase Extraction of Organic Compounds: A Critical Review. Part II. *Trends Anal. Chem*. **2016**, 80, 655–667. https://doi.org/10.1016/j.trac.2015.08.014.

Andersson, A.; Lavonen, E.; Harir, M.; Gonsior, M.; Hertkorn, N.; Schmitt-Kopplin, P.; Kylin, H.; Bastviken, D. Selective Removal of Natural Organic Matter during Drinking Water Production Changes the Composition of Disinfection By-Products. *Environ. Sci. Water Res. Technol*. **2020**, 6(3), 779–794. https://doi.org/10.1039/c9ew00931k.

Avtonomov, D. M.; Raskind, A.; Nesvizhskii, A. I. BatMass: A Java Software Platform for LC-MS Data Visualization in Proteomics and Metabolomics. *J. Proteome Res*. **2016**, 15 (8), 2500–2509. https://doi.org/10.1021/acs.jproteome.6b00021.

Bedner, M.; MacCrehan, W. A. Transformation of Acetaminophen by Chlorination Produces the Toxicants 1,4-Benzoquinone and N-Acetyl-p-Benzoquinone Imine. *Environ. Sci. Technol*. **2006**, 40 (2), 516–522. https://doi.org/10.1021/es0509073.

Bonnefille, B.; Karlsson, O.; Rian, M. B.; Raqib, R.; Parvez, F.; Papazian, S.; Islam, M. S.; Martin, J. W. Nontarget Analysis of Polluted Surface Waters in Bangladesh Using Open Science Workflows. *Environ. Sci. Technol*. **2023**. 57(17), 6808-6824. https://doi.org/10.1021/acs.est.2c08200.

Bulloch, D. N.; Nelson, E. D.; Carr, S. A.; Wissman, C. R.; Armstrong, J. L.; Schlenk, D.; Larive, C. K. Occurrence of Halogenated Transformation Products of Selected Pharmaceuticals and Personal Care Products in Secondary and Tertiary Treated Wastewaters from Southern California. *Environ. Sci. Technol*. **2015**, 49 (4), 2044–2051. https://doi.org/10.1021/es504565n.

Buszewski, B.; Noga, S. Hydrophilic Interaction Liquid Chromatography (HILIC)-a Powerful Separation Technique. *Anal. Bioanal. Chem*. **2012**, 402 (1), 231–247. https://doi.org/10.1007/s00216-011-5308-5.

Cai, L.; Li, L.; Yu, S. Formation of Odorous Aldehydes , Nitriles and N -Chloroaldimines from Combined Leucine in Short Oligopeptides during Chlorination. *Water Res*. **2020**, 177, 115803. https://doi.org/10.1016/j.watres.2020.115803.

Cao, F.; Zhang, M.; Yuan, S.; Feng, J.; Wang, Q.; Wang, W.; Hu, Z. Transformation of Acetaminophen during Water Chlorination Treatment: Kinetics and Transformation Products Identification. *Environ. Sci. Pollut. Res*. **2016**, 23 (12), 12303–12311. https://doi.org/10.1007/s11356-016-6341-x.

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.

Charbonnet, J. A.; Mcdonough, C. A.; Xiao, F.; Schwichtenberg, T.; Cao, D.; Kaserzon, S.; Thomas, K. V; Dewapriya, P.; Place, B. J.; Schymanski, E. L.; Field, J. A.; Helbling, D. E.; Higgins, C. P. Communicating Confidence of Per- and Polyfluoroalkyl Substance Identification

via High-Resolution Mass Spectrometry. *Environ. Sci. Technol. Lett.* **2022**, 9, 473-481. https://doi.org/10.1021/acs.estlett.2c00206.

Chen, C.; Zhang, X.; Zhu, L.; He, W.; Han, H. Changes in Different Organic Matter Fractions during Conventional Treatment and Advanced Treatment. *J. Environ. Sci*. 2011, 23 (4), 582–586. https://doi.org/10.1016/S1001-0742(10)60423-8.

Crittenden, J. C.; Trussell, R. R.; Hand, D. W.; Howe, K. J.; Tchbanoglous, G. *MWH's Water Treatment: Principles and Design,* Third.; John Wiley and Sons Inc., **2012**.

Cuthbertson, A. A.; Kimura, S. Y.; Liberatore, H. K.; Knappe, D. R. U.; Stanford, B.; Summers, R. S.; Dickenson, E. R.; Maness, J. C.; Glover, C.; Selbes, M.; Richardson, S. D. GAC to BAC: Does It Make Chloraminated Drinking Water Safer? *Water Res*. **2020**, 172, 115432. https://doi.org/10.1016/j.watres.2019.115432.

Cuthbertson, A. A.; Bach, C.; Richardson, S. D.; Dauchy, X. A Novel Automated Method for the Quantification of Ten Halobenzoquinones in Drinking Water Using Online Solid-Phase Extraction Coupled with Liquid Chromatography Tandem Mass Spectrometry. *J. Chromatogr. A* **2020**, 1612, 460642. https://doi.org/10.1016/j.chroma.2019.460642.

Dator, R.; von Weymarn, L. B.; Villalta, P. W.; Hooyman, C. J.; Maertens, L. A.; Upadhyaya, P.; Murphy, S. E.; Balbo, S. In Vivo Stable-Isotope Labeling and Mass-Spectrometry-Based Metabolic Profiling of a Potent Tobacco-Specific Carcinogen in Rats. *Anal. Chem*. **2018**, 90 (20), 11863–11872. https://doi.org/10.1021/acs.analchem.8b01881.

Diemert, S.; Wang, W.; Andrews, R. C.; Li, X. F. Removal of Halo-Benzoquinone (Emerging Disinfection By-Product) Precursor Material from Three Surface Waters Using Coagulation. *Water Res*. **2013**, 47 (5), 1773–1782. https://doi.org/10.1016/j.watres.2012.12.035.

Domingues, J. T.; Orlando, R. M.; Almeida, M. R.; de Lemos, L. R.; Mageste, A. B.; Rodrigues, G. D. Extraction of Estrogen Hormones from Water Samples Using an Aqueous Two-Phase System: A New Approach for Sample Preparation in the Analysis of Emerging Contaminants. *Microchem. J.* **2021**, 166 (March), 106231. https://doi.org/10.1016/j.microc.2021.106231.

Du, Z.; Ding, S.; Xiao, R.; Fang, C.; Song, W.; Jia, R.; Chu, W. Does Snowfall Introduce Disinfection By-Product Precursors to Surface Water ? *Environ. Sci. Technol*. **2022**, 56, 14487-14497. https://doi.org/10.1021/acs.est.2c04408.

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.

Essaïed, K. A.; Brown, L. V.; von Gunten, U. Reactions of Amines with Ozone and Chlorine: Two Novel Oxidative Methods to Evaluate the N-DBP Formation Potential from Dissolved Organic Nitrogen. *Water Res*. **2022**, 209, 117864. https://doi.org/10.1016/j.watres.2021.117864.

Government of Canada: Past Weather and Climate Data. https://climat.metro.gc.ca/historical_data/search_historic_data_e.html (Accessed Jan 3, 2023).

Gros, M.; Rodríguez-Mozaz, S.; Barceló, D. Fast and Comprehensive Multi-Residue Analysis of a Broad Range of Human and Veterinary Pharmaceuticals and Some of Their Metabolites in Surface and Treated Waters by Ultra-High-Performance Liquid Chromatography Coupled to Quadrupole-Linear Ion Trap Tandem. *J. Chromatogr. A* **2012**, 1248, 104–121. https://doi.org/10.1016/j.chroma.2012.05.084.

Guo, J.; Shen, S.; Xing, S.; Yu, H.; Huan, T. ISFrag: De Novo Recognition of In-Source Fragments for Liquid Chromatography-Mass Spectrometry Data. *Anal. Chem*. **2021**, 93 (29), 10243–10250. https://doi.org/10.1021/acs.analchem.1c01644.

Guo, J.; Huan, T. Mechanistic Understanding of the Discrepancies between Common Peak Picking Algorithms in Liquid Chromatography-Mass Spectrometry-Based Metabolomics. *Anal. Chem*. **2022**. 95(14), 5894–5902. https://doi.org/10.1021/acs.analchem.2c04887.

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.

Han, J.; Zhang, X.; Jiang, J.; Li, W. How Much of the Total Organic Halogen and Developmental Toxicity of Chlorinated Drinking Water Might Be Attributed to Aromatic Halogenated DBPs? *Environ. Sci. Technol*. **2021**, 55 (9), 5906–5916. https://doi.org/10.1021/acs.est.0c08565.

Health Canada. *Guidelines for Canadian Drinking Water Quality—Summary Tables*. Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch: Ottawa, ON, 2022; Vol. 24. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/water-quality/guidelines-canadian-drinking-water-quality-summary-table.html. (Accessed July 20th, 2023).

Health Canada. *Guidelines for Canadian Drinking Water Quality: Guideline Technical Document—Chlorine*. Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, 2009. https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-chlorine-guideline-technical-document/page-2-guidelines-

canadian-drinking-water-quality-chlorine-guideline-technical-document.html. (Accessed July 20th, 2023).

Helmus, R.; ter Laak, T. L.; van Wezel, A. P.; de Voogt, P.; Schymanski, E. L. patRoon: Open Source Software Platform for Environmental Mass Spectrometry Based Non-Target Screening. *J. Cheminform*. **2021**, 13 (1), 1–25. https://doi.org/10.1186/s13321-020-00477-w.

Helmus, R.; van de Velde, B.; Brunner, A. M.; ter Laak, T. L.; van Wezel, A. P.; Schymanski, E. L. patRoon 2.0: Improved Non-Target Analysis Workflows Including Automated Transformation Product Screening. *J. Open Source Softw*. **2022**, 7 (71), 4029. https://doi.org/10.21105/joss.04029.

Hollender, J.; van Bavel, B.; Dulio, V.; Farmen, E.; Furtmann, K.; Koschorreck, J.; Kunkel, U.; Krauss, M.; Munthe, J.; Schlabach, M.; Slobodnik, J.; Stroomberg, G.; Ternes, T.; Thomaidis, N. S.; Togola, A.; Tornero, V. High Resolution Mass Spectrometry-Based Non-Target Screening Can Support Regulatory Environmental Monitoring and Chemicals Management. *Environ. Sci. Eur.* **2019**, 31, 42. https://doi.org/10.1186/s12302-019-0225-x.

How, Z. T.; Busetti, F.; Linge, K. L.; Kristiana, I.; Joll, C. A.; Charrois, J. W. A. Analysis of Free Amino Acids in Natural Waters by Liquid Chromatography – Tandem Mass Spectrometry. *J. Chromatogr. A* **2014**, 1370, 135–146. https://doi.org/10.1016/j.chroma.2014.10.040.

Hrudey, S. E.; Rector, D.; Motkosky, N. Characterization of Drinking Water Odour Arising from Spring Thaw for an Ice-Covered Upland River Source. *Wat. Sci. Tech*. **1992**, 25 (2), 65–72.

Hrudey, S. E. Chlorination Disinfection By-Products, Public Health Risk Tradeoffs and Me. *Water Res*. **2009**, 43 (8), 2057–2092. https://doi.org/10.1016/j.watres.2009.02.011.

Hrudey, S. E.; Hrudey, E. J.; Pollard, S. J. T. Risk Management for Assuring Safe Drinking *Water. Environ. Int*. **2006**, 32 (8), 948–957. https://doi.org/10.1016/j.envint.2006.06.004.

Hu, Y.; Cai, B.; Huan, T. Enhancing Metabolome Coverage in Data-Dependent LC-MS/MS Analysis through an Integrated Feature Extraction Strategy. *Anal. Chem*. **2019**, 91, 14433-14441. https://doi.org/10.1021/acs.analchem.9b02980.

Hu, H. Y.; Du, Y.; Wu, Q. Y.; Zhao, X.; Tang, X.; Chen, Z. Differences in Dissolved Organic Matter between Reclaimed Water Source and Drinking Water Source. *Sci. Total Environ*. **2016**, 551–552, 133–142. https://doi.org/10.1016/j.scitotenv.2015.12.111.

Hu, S.; Gong, T.; Ma, J.; Tao, Y.; Xian, Q. Simultaneous Determination of Iodinated Haloacetic Acids and Aromatic Iodinated Disinfection Byproducts in Waters with a New SPE-HPLC-MS/MS Method. *Chemosphere* **2018**, 198, 147–153. https://doi.org/10.1016/j.chemosphere.2018.01.124.

Huan, T.; Li, L. Counting Missing Values in a Metabolite-Intensity Data Set for Measuring the Analytical Performance of a Metabolomics Platform. *Anal. Chem*. **2015**, 87 (2), 1306–1313. https://doi.org/10.1021/ac5039994.

Huang, G.; Jiang, P.; Jmaiff Blackstock, L. K.; Tian, D.; Li, X. F. Formation and Occurrence of Iodinated Tyrosyl Dipeptides in Disinfected Drinking Water. *Environ. Sci. Technol*. **2018**, 52 (7), 4218–4226. https://doi.org/10.1021/acs.est.7b06276.

Huang, G.; Jiang, P.; Li, X. F. Mass Spectrometry Identification of N-Chlorinated Dipeptides in Drinking Water. *Anal. Chem*. **2017**, 89 (7), 4204–4209. https://doi.org/10.1021/acs.analchem.7b00228.

Jacangelo, J. G.; DeMarco, J.; Owen, D. M.; Stephen, J.; Jacangelo, J. G.; Demarco, J.; Owen, D. M.; Randtke, S. J. Selected Processes for Removing NOM : An Overview. *AWWA*. **1995**, 87 (1), 64–77.

Jasemizad, T.; Bromberg, L.; Hatton, T. A.; Padhye, L. P. Oxidation of Betrixaban to Yield N-Nitrosodimethylamine by Water Disinfectants. *Water Res*. **2020**, 186, 116309. https://doi.org/10.1016/j.watres.2020.116309.

Jennings, E. K.; Sierra Olea, M.; Kaesler, J. M.; Hübner, U.; Reemtsma, T.; Lechtenfeld, O. J. Stable Isotope Labeling for Detection of Ozonation Byproducts in Effluent Organic Matter with FT-ICR-MS. *Water Res*. **2023**, 229, 119477. https://doi.org/10.1016/j.watres.2022.119477.

Kimura, S.Y.; Zheng, W.; Hipp, T.N.; Allen, J.M.; Richardson, S.D. Total Organic Halogen (TOX) in Human Urine: A Halogen-Specific Method for Human Exposure Studies. *J. Environ. Sci. (China)* **2017**, 58, 285–295. https://doi.org/10.1016/j.jes.2017.04.008.

Kosaka, K.; Nakai, T.; Hishida, Y.; Asami, M.; Ohkubo, K.; Akiba, M. Formation of 2,6-Dichloro-1,4-Benzoquinone from Aromatic Compounds after Chlorination. *Water Res*. **2017**, 110, 48–55. https://doi.org/10.1016/j.watres.2016.12.005.

Krasner, S. W.; Roback, S.; Qian, Y.; Li, X. F.; Marfil-Vega, R.; Bukhari, Z. Occurrence of Nitrosamines and Their Precursors in North American Drinking Waters. *AWWA Water Sci*. **2020**, 2 (6), 1–15. https://doi.org/10.1002/aws2.1208.

Krasner, S. W. The Formation and Control of Emerging Disinfection By-Products of Health Concern. *Philos. Trans. Math. Phys. Eng. Sci*. **2009**, 367 (1904), 4077–4095. https://doi.org/10.1098/rsta.2009.0108.

Krauss, M.; Singer, H.; Hollender, J. LC-High Resolution MS in Environmental Analysis: From Target Screening to the Identification of Unknowns. *Anal. Bioanal. Chem*. **2010**, 397 (3), 943–951. https://doi.org/10.1007/s00216-010-3608-9.

Lau, S. S.; Bokenkamp, K.; Tecza, A.; Wagner, E. D.; Plewa, M. J.; Mitch, W. A. Toxicological Assessment of Potable Reuse and Conventional Drinking Waters. *Nat. Sustain*. **2023**, 6 (1), 39–46. https://doi.org/10.1038/s41893-022-00985-7.

Lee, W.; Westerhoff, P.; Esparza-Soto, M. Occurrence and Removal of Dissolved Organic Nitrogen in US Water Treatment Plants. *AAWA*. **2006**, 98(10), 102–110. https://doi.org/10.1002/j.1551-8833.2006.tb07782.x.

Lepistö, A.; Räike, A.; Sallantaus, T.; Finér, L. Increases in Organic Carbon and Nitrogen Concentrations in Boreal Forested Catchments — Changes Driven by Climate and Deposition. *Sci. Total Environ*. **2021**, 780, 146627. https://doi.org/10.1016/j.scitotenv.2021.146627.

Li, W.; Zhang, X.; Han, J. Formation of Larger Molecular Weight Disinfection Byproducts from Acetaminophen in Chlorine Disinfection. *Environ. Sci. Technol*. **2022**, 56 (23), 16929–16939. https://doi.org/10.1021/acs.est.2c06394.

Linge, K. L.; Kristiana, I.; Liew, D.; Holman, A.; Joll, C. A. Halogenated Semivolatile Acetonitriles as Chloramination Disinfection By-Products in Water Treatment: A New Formation Pathway from Activated Aromatic Compounds. *Environ. Sci. Process. Impacts* **2020**, 22 (3), 653–662. https://doi.org/10.1039/c9em00603f.

Liu, C.; Ersan, M. S.; Wagner, E.; Plewa, M. J.; Amy, G.; Karanfil, T. Toxicity of Chlorinated Algal-Impacted Waters: Formation of Disinfection Byproducts vs. Reduction of Cyanotoxins. *Water Res*. **2020**, 184, 116145. https://doi.org/10.1016/j.watres.2020.116145.

Liu, Z.; Craven, C. B.; Huang, G.; Jiang, P.; Wu, D.; Li, X. F. Stable Isotopic Labeling and Nontarget Identification of Nanogram/Liter Amino Contaminants in Water. Anal. Chem. 2019, 91 (20), 13213–13221. https://doi.org/10.1021/acs.analchem.9b03642.

Lloyd R. Snyder, Joseph J. Kirkland, J. W. D. *Introduction to Modern Liquid Chromatography*, Third.; John Wiley & Sons, Inc: Hoboken, NJ, USA, 2011.

Marczak, M.; Wolska, L.; Chrzanowski, W.; Namieśnik, J. Microanalysis of Volatile Organic Compounds (VOCs) in Water Samples - Methods and Instruments. *Microchim. Acta* **2006**, 155, 331–348. https://doi.org/10.1007/s00604-006-0630-x.

Mitch, W. A.; Richardson, S. D.; Zhang, X.; Gonsior, M. High-Molecular-Weight by-Products of Chlorine Disinfection. *Nat. Water* **2023**, 1, 336-347. https://doi.org/10.1038/s44221-023-00064-x.

Neuwald, I. J.; Hübner, D.; Wiegand, H. L.; Valkov, V.; Borchers, U.; Nödler, K.; Scheurer, M.; Hale, S. E.; Arp, H. P. H.; Zahn, D. Occurrence, Distribution, and Environmental Behavior of Persistent, Mobile, and Toxic (PMT) and Very Persistent and Very Mobile (vPvM) Substances in the Sources of German Drinking Water. *Environ. Sci. Technol*. **2022**, 56 (15), 10857–10867. https://doi.org/10.1021/acs.est.2c03659.

Pang, Z.; Zhou, G.; Ewald, J.; Chang, L.; Hacariz, O.; Basu, N.; Xia, J. Using MetaboAnalyst 5.0 for LC–HRMS Spectra Processing, Multi-Omics Integration and Covariate Adjustment of Global Metabolomics Data. *Nat. Protoc*. **2022**, 17 (8), 1735–1761. https://doi.org/10.1038/s41596-022-00710-w.

Pellicer-Castell, E.; Belenguer-Sapiña, C.; Amorós, P.; El Haskouri, J.; Herrero-Martínez, J. M.; Mauri-Aucejo, A. R. Mesoporous Silica Sorbent with Gold Nanoparticles for Solid-Phase

Extraction of Organochlorine Pesticides in Water Samples. *J. Chromatogr. A* **2022**, 1662, 462729. https://doi.org/10.1016/j.chroma.2021.462729.

Peng, J.; Huang, H.; Zhong, Y.; Yin, R.; Wu, Q.; Shang, C.; Yang, X. Transformation of Dissolved Organic Matter during Biological Wastewater Treatment and Relationships with the Formation of Nitrogenous Disinfection Byproducts. *Water Res*. **2022**, 222 (7), 118870. https://doi.org/10.1016/j.watres.2022.118870.

Qiu, J.; Craven, C.; Wawryk, N.; Carroll, K.; Li, X.-F. Integration of Solid Phase Extraction with HILIC-MS/MS for Analysis of Free Amino Acids in Source Water. *J. Environ. Sci. (China)* **2022**, 117, 190-196. https://doi.org/10.1016/j.jes.2022.04.025.

Qiu, J.; Huang, Y.; Wu, Y.; Shi, P.; Xu, B.; Chu, W.; Pan, Y. Detection, Transformation, and Toxicity of Indole-Derivative Nonsteroidal Anti-Inflammatory Drugs during Chlorine Disinfection. *Chemosphere* **2020**, 260, 127579. https://doi.org/10.1016/j.chemosphere.2020.127579.

Richardson, S. D.; Plewa, M. J.; Wagner, E. D.; Schoeny, R.; Demarini, D. M. Occurrence , Genotoxicity , and Carcinogenicity of Regulated and Emerging Disinfection by-Products in Drinking Water : A Review and Roadmap for Research. *Mutat. Res. - Rev. Mutat*. **2007**. 636, 178–242. https://doi.org/10.1016/j.mrrev.2007.09.001.

Richardson, S. D.; Ternes, T. A. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem*. **2022**, 94 (1), 382–416. https://doi.org/10.1021/acs.analchem.1c04640.

Richardson, S. D.; Kimura, S. Y. Emerging Environmental Contaminants: Challenges Facing Our next Generation and Potential Engineering Solutions. *Environ. Technol. Innov*. **2017**, 8, 40–56. https://doi.org/10.1016/j.eti.2017.04.002.

Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H. C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat. Methods* **2016**, 13 (9), 741–748. https://doi.org/10.1038/nmeth.3959.

Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag Relaunched: Incorporating Strategies beyond in Silico Fragmentation. *J. Cheminform*. **2016**, 8 (3), 1–16. https://doi.org/10.1186/s13321-016-0115-9.

Scheurer, M.; Brauch, H. J.; Lange, F. T. Analysis and Occurrence of Seven Artificial Sweeteners in German Waste Water and Surface Water and in Soil Aquifer Treatment (SAT). *Anal. Bioanal. Chem*. **2009**, 394 (6), 1585–1594. https://doi.org/10.1007/s00216-009-2881-y.

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.

Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N. S.; Bletsou, A.; Zwiener, C.; Ibáñez, M.;

Portolés, T.; De Boer, R.; Reid, M. J.; Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.; Bogialli, S.; Stipaničev, D.; Rostkowski, P.; Hollender, J. Non-Target Screening with High-Resolution Mass Spectrometry: Critical Review Using a Collaborative Trial on Water Analysis. *Anal. Bioanal. Chem*. **2015**, 407 (21), 6237–6255. https://doi.org/10.1007/s00216-015-8681-7.

Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol*. **2014**, 48 (4), 2097–2098. https://doi.org/10.1021/es5002105.

Sebestyen, S. D.; Boyer, E. W.; Shanley, J. B.; Kendall, C.; Doctor, D. H.; Aiken, G. R.; Ohte, N. Sources, Transformations, and Hydrological Processes That Control Stream Nitrate and Dissolved Organic Matter Concentrations during Snowmelt in an Upland Forest. *Water Resources Research* **2008**, 44, 1–14. https://doi.org/10.1029/2008WR006983.

Seiwert, B.; Nihemaiti, M.; Troussier, M.; Weyrauch, S.; Reemtsma, T. Abiotic Oxidative Transformation of 6-PPD and 6-PPD Quinone from Tires and Occurrence of Their Products in Snow from Urban Roads and in Municipal Wastewater. *Water Research*. **2022**, 212, 118122. https://doi.org/10.1016/j.watres.2022.118122.

Smith, R.A.; Zhang, Q. Peak Pair Pruner: a post-processing software to MS-DIAL for peak pair validation and ratio quantification of isotopic labeling LC-MS(/MS) data. *Adv. Bioinformatics* **2023**. 3(1), 1-3. https://doi.org/10.1093/bioadv/vbad044.

Stats Canada: Survey of Drinking Water Plants, 2019. https://www150.statcan.gc.ca/n1/daily-quotidien/210817/dq210817c-eng.htm (Accessed July 26, 2023).

Tang, Y.; Xu, Y.; Li, F.; Jmaiff, L.; Hrudey, S. E.; Li, X. F. Nontargeted Identification of Peptides and Disinfection Byproducts in Water. *J. Environ. Sci. (China)* **2016,** 42, 259–266. https://doi.org/10.1016/j.jes.2015.08.007.

Tautenhahn, R.; Bottcher, C.; Neumann, S. Highly Sensitive Feature Detection for High Resolution LC/MS. *BMC Bioinformatics* **2008**, 9, 1–16. https://doi.org/10.1186/1471-2105-9-504.

Telang, S. A. Effects of Reservoir-Dam, Urban, Industrial, and Sewage Treatment Run-off on the Presence of Oxygen and Organic Compounds in the Bow River. *Water. Air. Soil Pollut*. **1990**, 50 (1–2), 77–90. https://doi.org/10.1007/BF00284785.

Tong, Z.; Linge, K. L.; Busetti, F.; Joll, C. A. Formation of Odorous and Hazardous By-Products from the Chlorination of Amino Acids. *Water Res*. **2018**, 146, 10–18. https://doi.org/10.1016/j.watres.2018.08.072.

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.

Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E. R. Deuterium Isotope Effects on Hydrophobic Interactions: The Importance of Dispersion Interactions in the Hydrophobic Phase. *J. Am. Chem. Soc*. **2003**, 125 (45), 13836–13849. https://doi.org/10.1021/ja036006g.

United Nations: Goal 6: Ensure access to water and sanitation for all https://www.un.org/sustainabledevelopment/water-and-sanitation/ (Accessed Jul 26, 2023).

USEPA: National Primary Drinking Water Regulations https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations#Byproducts (Accessed Jul 7, 2023).

USEPA. *Fact Sheet Fifth Contaminant Candidate List (CCL 5)*, 2022. https://www.epa.gov/system/files/documents/2022-10/Fact%20Sheet%20Final%20Fifth%20Contaminant%20Candidate%20List%20%28CCL%205%29.pdf (Accessed July 26, 2023)

Walker, G. S.; Lee, F. P.; Aieta, E. M. Chlorine Dioxide for Taste and Odor Control. *J. Am. Water Work. Assoc*. **1986**, 78 (3), 84–93. https://doi.org/10.1002/j.1551-8833.1986.tb05719.x.

Wallis, P. M.; Hynes, H. B. N.; Telang, S. A. The Importance of Groundwater in the Transportation of Allochthonous Dissolved Organic Matter to the Streams Draining a Small Mountain Basin. *Hydrobiologia* **1981**, 79 (1), 77–90. https://doi.org/10.1007/BF00005821.

Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal. Chem*. **2021**, 93 (34), 11692–11700. https://doi.org/10.1021/acs.analchem.1c01465.

Wawryk, N. J. P.; Craven, C. B.; Blackstock, L. K. J.; Li, X. F. New Methods for Identification of Disinfection Byproducts of Toxicological Relevance: Progress and Future Directions. *J. Environ. Sci. (China)* **2021**, 99, 151–159. https://doi.org/10.1016/j.jes.2020.06.020.

Wawryk, N. J. P.; Huang, G.; Craven, C.; Qiu, J.; Jmaiff Blackstock, L. K.; Li, X. F. Aspartame-Sweetened Tap Water: Transformation Products and 2,6-Dichloro-1,4-Benzoquinone Formation. *Environ. Sci. Technol*. **2022**, 2–11. https://doi.org/10.1021/acs.est.2c07156.

WHO: Drinking-water https://www.who.int/news-room/fact-sheets/detail/drinking-water

(Accessed Jul 26, 2023).

Xia, D.; Liu, H.; Lu, Y.; Liu, Y.; Liang, J.; Xie, D.; Lu, G.; Qiu, J.; Wang, R. Utility of a Non-Target Screening Method to Explore the Chlorination of Similar Sulfonamide Antibiotics: Pathways and NCl Intermediates. *Sci. Total Environ*. **2023**, 858, 160042. https://doi.org/10.1016/J.SCITOTENV.2022.160042.

Xing, S.; Shen, S.; Xu, B.; Huan, T. Molecular Formula Discovery via Bottom-up MS/MS Interrogation. *Nat. Methods* **2023**, 20 (6), 881-890. https://doi.org/10.1038/s41592-023-01850-x.

Yang, M.; Liberatore, H. K.; Zhang, X. Current Methods for Analyzing Drinking Water Disinfection Byproducts. *Curr. Opin. Environ. Sci. Heal*. **2019**, 7, 98–107. https://doi.org/10.1016/j.coesh.2018.12.006.

Yu, H.; Huan, T. Systems Biology MAFFIN : Metabolomics Sample Normalization Using Maximal Density Fold Change with High-Quality Metabolic Features and Corrected Signal Intensities. *Bioinformatics* **2022**, 38 (13), 3429–3437. https://doi.org/10.1093/bioinformatics/btac355.

Yu, Y.; Reckhow, D. A. Formation and Occurrence of N-Chloro-2,2-Dichloroacetamide, a Previously Overlooked Nitrogenous Disinfection Byproduct in Chlorinated Drinking Waters. *Environ. Sci. Technol*. **2017**, 51 (3), 1488–1497. https://doi.org/10.1021/acs.est.6b04218.

Yuan, M.; Breitkopf, S. B.; Yang, X.; Asara, J. M. A Positive/Negative Ion-Switching, Targeted Mass Spectrometry-Based Metabolomics Platform for Bodily Fluids, Cells, and Fresh and Fixed Tissue. *Nat. Protoc*. **2012**, 7 (5), 872–881. https://doi.org/10.1038/nprot.2012.024.

Zhang, D.; Chu, W.; Yu, Y.; Krasner, S. W.; Pan, Y.; Shi, J.; Yin, D.; Gao, N. Occurrence and Stability of Chlorophenylacetonitriles: A New Class of Nitrogenous Aromatic DBPs in Chlorinated and Chloraminated Drinking Waters. *Environ. Sci. Technol. Lett*. **2018**, 5 (6), 394–399. https://doi.org/10.1021/acs.estlett.8b00220.

Zhang, H.; Gao, P.; Liu, Y.; Du, Z.; Feng, L.; Zhang, L. Effects of Different Types of Nitrogen Sources in Water on the Formation Potentials of Nitrogenous Disinfection By-Products in Chloramine Disinfection Process Based on Isotope Labeling. *Sci. Total Environ*. **2022**, 842, 156692. https://doi.org/10.1016/J.SCITOTENV.2022.156692.

Zhang, H.; Yang, M. Characterization of Brominated Disinfection Byproducts Formed during Chloramination of Fulvic Acid in the Presence of Bromide. *Sci. Total Environ*. **2018**, 627, 118–124. https://doi.org/10.1016/j.scitotenv.2018.01.215.

Zhao, T.; Carroll, K.; Craven, C. B.; Wawryk, N. J. P.; Xing, S.; Guo, J.; Li, X.; Huan, T. HDPairFinder : A Data Processing Platform for Hydrogen / Deuterium Isotopic Labeling-Based Nontargeted Analysis of Trace-Level Amino-Containing Chemicals in Environmental Water. *J. Environ. Sci*. **2024**, 136, 583–593. https://doi.org/10.1016/j.jes.2023.02.033.

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

Zhou, R.; Xu, Z.; Zhu, J.; Liu, W.; Meng, Y.; Zhu, P.; Zhou, W.; Huang, C.; Ding, X. Determination of 10 Haloacetamides in Drinking Water by Gas Chromatography with Automated Solid Phase Extraction. *J. Chromatogr. B*. **2020**, 1150 (10), 122191. https://doi.org/10.1016/j.jchromb.2020.122191.