2021

DETERMINATION OF MOLECULAR MARKERS OF VACCINIUM BERRY STANDARD REFERENCE MATERIALS THROUGH DIFFERENTIAL ANALYSIS WITH ULTRAHIGH RESOLUTION LC/MS

Abby Mikolitis
Michigan Technological University, asmikoli@mtu.edu

DETERMINATION OF MOLECULAR MARKERS OF V*ACCINIUM* BERRY STANDARD REFERENCE MATERIALS THROUGH DIFFERENTIAL ANALYSIS WITH ULTRAHIGH RESOLUTION LC/MS

By

Abigale S. Mikolitis

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Chemistry

MICHIGAN TECHNOLOGICAL UNIVERSITY

2021

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Chemistry.

Department of Chemistry

Thesis Advisor:	*Dr. Lynn R. Mazzoleni*
Committee Member:	*Dr. Sarah Green*
Committee Member:	*Dr. Laura Brown*
Committee Member:	*Dr. Momoko Tajiri*
Department Chair:	*Dr. Sarah Green*

Table of Contents

Acknowledgements

I would like to thank my advisor, Dr. Mazzoleni for allowing me the opportunity to join her research group and offering her support and guidance. I am grateful to my committee members Dr. Green, Dr. Tijari, and Dr. Brown for taking the time to review and critique this thesis and would like to thank them for their encouragement as I worked through this program. I would especially like to thank Dr. Tajiri for giving me the opportunity to participate in this project.

Additionally, I am thankful for my wonderful fellow group members, Dr. Simeon Schum, Thusitha Divisekara, and Amna Ijaz for all the support and assistance they gave as I continued to learn. I would like to thank the Department of Chemistry, who gave me financial support, and all the faculty and staff who taught and challenged me. I am thankful to all my friends and family who supported me through this journey. Finally, I would like to thank the Michigan Craft Beverage Council for their support and funding through this project.

Abstract

Blueberries and cranberries have vastly different flavor profiles despite both being a part of the *Vaccinium* genus. Their high content of phenolics makes them subjects of interest in preventative medicine, especially research in illnesses such as heart disease and cancer. Analysis has primarily focused on the identification and characterization of polyphenols, primarily anthocyanins which are a subclass of polyphenols, with ultra-high performance liquid chromatography coupled with electrospray ionization high resolution mass spectrometry (UHPLC-ES/HRMS). This study aims to develop a novel method of non-targeted UHPLC-ES/MS analysis with post processing and formula assignment performed with MZmine 2.53 and MFAssignR software tools. Pairing non-targeted data analysis and post data processing with assorted statistical analysis techniques assisted in identifying molecular markers that illustrate the flavor differences that contribute to the unique profiles of blueberries and cranberries. The National Institute of Standards and Technology (NIST) manufactures standard reference materials (SRM) that were used as the samples for the purpose of method development. The differential analysis produced more than 1000 statistically significant species between the two the SRMs. In comparing these species against a public database, Phenol Explorer 3.6, 67 statistically notable phenolic compounds were identified as potential markers that contribute to the unique flavor profiles of *Vaccinium* berries.

1 Introduction

Michigan is the third largest producer of blueberries in the United States. The fruit holds regional interest and is expected to have an important role in the increasing consumption and marketability of alcoholic and non-alcoholic fruit products (Michigan Ag Council, 2021; Abate and Peterson, 2005). There is no direct evidence demonstrating the relationship of Michigan blueberries and increased fruit-based product sales. However, there has been a recorded rise in the consumption of fruit-based products especially in the alcoholic beverage sector with the rise in craft breweries and wineries. These businesses may choose to take advantage of this local resource (Abate and Peterson, 2005).

Blueberries are part of the *Vaccinium* genus and have long been studied for their high concentration of phenolics and their use in preventing illnesses including heart disease and cancer. Phenolics are aromatic species containing at least one hydroxyl group and cover several subclasses including flavanols, anthocyanins, and phenolic acids (Neveu et al., 2010; Rothwell et al., 2013). The identification and characterization of these compounds within *Vaccinium* berries has been of high interest, especially anthocyanins (Kalt and Dufour, 1997; Holiman et. al., 1996; Katsube et al., 2003). Anthocyanins are a subclass of flavanols and are powerful antioxidants that are responsible for the red and blue pigments in berries. They serve several natural functions including pest repellant and UV protection (Nakajima et al., 2004; Montoro et al., 2006).

Studies using methods in ultrahigh pressure liquid chromatography paired with high resolution mass spectrometry with electrospray ionization (UHPLC-ES/HRMS) have primarily focused on the characterization of these compounds with positive mode

electrospray ionization (+ESI) due to the anthocyanins positive charge (NIST Certificate of Analysis, 2019a ; NIST Certificate of Analysis, 2019b; Montoro, et al., 2006). Additionally, post processing has relied on database matches to identify compounds of interest (Lowenthal et al., 2012). Methods using negative mode electrospray ionization (-ESI) have not been as extensively explored, changing the ionization mode will assist in characterizing additional subclasses of polyphenols and may cut down on noise (Liigand et al., 2017).

Optimizing the extraction method for the isolation of phenolic compounds prior to non-targeted analysis with UHPLC-ES/HRMS and -ESI could prove useful in identifying key differences in molecular markers across different species of *Vaccinium* berries. Standard reference materials (SRMs) were chosen for method validation because they offer a consistent, standardized samples with low environmental interferences. However, applying this method to fruit samples, specifically Michigan blueberries, could give insight into the compound classes that give these berries their distinct flavor.

1.1 *Vaccinium* Berry Standard Reference Material

SRMs of blueberry and cranberry were selected as a basis for comparison. Both berries are a part of the *Vaccinium* genus but are unique in habitat and flavor profile. Additionally, cranberries are widely available. This facilitates the possibility of expanding on this study with fresh or frozen berries and extracts. Both SRMs consist of freeze-dried and powdered fruits that have been homogenized (NIST Certificate of Analysis, 2019a ; NIST Certificate of Analysis, 2019b). These samples are optimal for developing the techniques outlined in this study by offering consistency and knowledge

of specific elemental and phenolic content as determined by the National Institute of Standards and Technology (NIST).

1.2 Analyte Isolation

Solid phase extraction (SPE) is a sample preparation technique meant to isolate select analytes, concentrate the sample, and remove matrix interferences such as salts, sugars, and fats. Most commonly, it is performed with cartridges or disks with a silica-based packing with low selectivity. It was developed as an easy, cost-effective replacement for liquid-liquid extraction by reducing the time and volume of materials needed to perform the extraction (Poole, 2003; Buszewski and Szultka, 2012). However, despite improvements, SPE has its own disadvantages. There is higher retention for more basic analytes and lower reproducibility (Poole, 2003).

Waters Oasis HLB (Hydrophillic-Lipophillic Balance) SPE cartridges are general use, reverse-phase cartridges with high retention capacity and a wide pH range (Poole, 2003; Waters, 2021b). The packing is extremely hydrophilic, allowing for more hydrophobic compounds to be retained by the surface of the silica while matrix interferences such as fats pass through. This is ideal for more aromatic compounds such as phenolics and antioxidants which may be eluted with a polar mobile phase (Waters, 2021b; Zwir-Ferenc and Biziuk, 2006).

1.3 UHPLC/HRMS

Ultrahigh performance liquid chromatography (UHPLC) utilizes a low diffusion, high pressure system with columns containing smaller, porous particles that allow for higher flowrates without band broadening. This is optimal for pairing with fast scanning mass analyzers for enhanced resolution and mass accuracy (Forcisi et al., 2013).

There are several classes of ultrahigh resolution mass spectrometry (HRMS). Fourier Transform Mass Spectrometry (FTICR) and the Orbitrap offer the highest mass accuracy and resolution (Scigelova et al., 2011). While FTICR offers the highest resolution, there has been considerable advancements in the resolving power of the Orbitrap, specifically with the development of the Orbitrap Elite which is capable of 240,000 resolving power at 400 m/z (Figure 1.1) (Michalski et al., 2011).

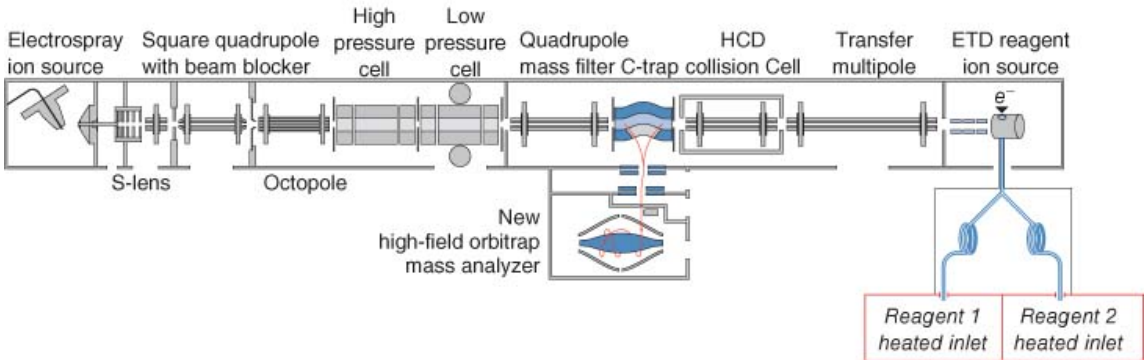

Figure 1.1 Orbitrap Elite Mass Spectrometer Thermo Scientific, (Hecht et al., 2019)

The combination of UHPLC with HRMS allows for the species to be pre-sorted by polarity. The most popular UHPLC technique is reverse phase which employs a polar mobile phase with a non-polar stationary phase allowing for more polar species to elute earlier and increasing the retention of hydrophobic species. The C18-bonded silica column is widely used for reverse phase chromatography, specifically for the analysis of alcohol and carboxylic acid containing compounds such as polyphenols (Waters, 2021a).

LC/ MS coupling can help differentiate between isotopic and enantiomeric species. For example, the three main sugar components of blueberries are sucrose, fructose, and galactose, all of which are structural isomers. The LC component allows for the isomers to be column separated to some degree prior to MS fragmentation. This allows us to explore these isomers in terms of intensity and possible structural differences.

The eluent is ionized through electrospray ionization (ESI), the mechanism will be discussed in the next section. The analytes are pulled through the chamber to the S lens which focuses the ion beam and pushes the beam to the beam blocker which removes any remaining neutral molecules preventing excessive noise (Hecht et al, 2019; Michalski, et al. 2011). From this point the charged analytes are fed to the high and low pressure cells through an octupole designed to filter out low stability ions. The high pressure cell forces the ions into packets by diminishing their kinetic energy through collision with dry nitrogen gas (N_2). The packets then move to the low pressure cell which isolate analytes withing the determined mass range which was 100 to 600 m/z for this method. After this, the analytes are transferred linear ion trap for collection. Once enough ions have been collected, they are transferred to the C-trap which is a curved

linear ion trap filled with N_2 causing the analytes to lose their kinetic energy before being fed into the Orbitrap (Michalski et al., 2011).

$$\omega_z = \sqrt{k\frac{z}{m}} \qquad (1) \qquad\qquad R = \frac{m}{\Delta m} = \frac{1}{2\Delta\omega}\sqrt{\frac{kz}{m}} \qquad (2)$$

The packet of analytes spirals around the central electrode of the Orbitrap with two directions of oscillation (axial and radial). The frequency of axial harmonic oscillation (ω_z) is solely reliant on mass (m_i) to charge (z) ratio as seen in Equation 1. Different masses will oscillate at different frequencies which also determines resolving power (R), as seen in Equation 2 where resolving power is equal to half the frequency. The unique spindle-shaped the inner electrode and the bell-shaped outer electrodes create an electrical potential difference at the poles of the Orbitrap analyzer (Gross, 2017, p. 248-249). This allows for higher resolution and mass accuracy between species with the same nominal mass with mDa differences and distinguishing between species such as CH_4 vs O or CH_2 vs N which have mass differences of 0.03639 and 0.01258 respectively.

1.4 Electrospray Ionization

The optimal fragmentation technique for UHPLC/HRMS coupling is ESI. There are several ionization methods for mass spectrometry, however, ESI is preferred for complex mixtures. ESI is known as a soft ionization method and results into little or no fragmentation. Additionally, ESI allows for the analysis of a wide variety of species and easily ionizes hydrophilic compounds such as phenolics (Kujawinski et al., 2002; Kujawininski and Behn, 2006).

Despite the popularity of ESI, the understanding of the mechanism remains limited. The eluent from the UPHLC ionized with a 2-5 kV power supply and is passed through

the capillary tubing to the spray needle where the analytes are either oxidated ($[M + H]^+$) or reduced ($[M - H]^-$). The far end of the chamber is charged opposite of the spray needle creating a potential difference. This pulls the solvated ions towards the S lens and creates a Taylor cone at the end of the needle and sprays a fine mist into the chamber (Kujawininski and Behn, 2006; Gross, 2017, p 721-723).

Organic solvents such as methanol (MeOH) and acetonitrile (ACN) are optimal for solvating analytes while allowing for easy evaporation. However, some water is necessary to create the surface tension at the tip of the spray needle to assist in the formation of the Taylor cone. Additionally, small amounts of weak acids, such as formic acid, added to the UHPLC method can increase ionization efficiency (Henriksen et al., 2005; Cech and Enke, 2001b; Lowenthal et al., 2012).

ESI creates charged droplets which are evaporated as the droplets move to the other end of the charged spray chamber. The evaporating solvent causes high coulombic repulsion forces between ions eventually breaking the surface tension causing the droplet to split. This continues until the droplets have been evaporated to dry single ions which are passed through a heated capillary to ensure dryness carried by N_2 before they enter the mass spectrometer (Talfin et al., 1989).

Analysis of *Vaccinium* berries with +ESI has been well documented, the technique is optimal for antioxidants such as anthocyanins (Nakajima et al., 2004; Montoro et al., 2006; Lowenthal et al., 2012). However, -ESI could help better characterize other phenolic compounds such as other subclasses of the flavonoids. -ESI offers lower background noise and could possibly show a wider range of species with better ionization efficiencies (Liigand et al., 2017).

1.5 Data Post Processing

1.5.1 MZmine 2.53

MZmine is one of the most widely used UHPLC/HRMS post data processing software for the construction of extracted ion chromatograms (EIC) and chromatographic peak detection (Katajamaa and Orešič, 2005). The software employs the point distribution pattern from the collected m/z and retention time data to build chromatograms. Mass extraction uses the extracted centroid mass from each scan and the EIC is built with Automated Data Analysis Pipeline workflow (ADAP) which orders the extracted masses by intensity and points below a determined noise threshold are removed (Katajamaa et al, 2006; Pluskal et al., 2010).

The EIC is then built-in order of intensity of the peaks in which peaks within a specified mass error range, 3 parts per million (ppm) for this study, are determined and grouped to continue to build the EIC. Once the peak of the next highest intensity falls outside of this pre-determined mass range, a new EIC is created, and this is repeated until all peaks have been processed (Myers et al., 2017). This is followed by chromatographic peak detection using ADAP peak detection algorithms which create peak lists that can be exported to MFAssignR for formula assignment.

There are many other features within the MZmine 2.53 that include alignment protocols and statistical analyses (Katajamaa et al., 2006; Pluska et al., 2010). However, these tools are limited and the processes for these analyses are not transparent. For the purposes of this study these tools were not pursued in favor of performing them with new code written in RStudio.

1.5.2 Molecular Formula Assignment

There are several methods for molecular formula (MF) assignment. If there are only a few analytes, the assignments could feasibly be done by hand. However, for complex mixtures this is unreasonable. Targeted studies have focused on using database matching (Lowenthal et al., 2012) which is not as effective for non-targeted studies.

MFAssignR uses a data-science driven approach for MF assignment to ensure accurate assignments. The MF assignments with MFAssignR builds off the CHOFIT algorithm developed by Perdue and Green in 2015 (Perdue and Green, 2015) with additional formula extensions to account for heteroatoms, adducts, and odd electrons (Schum et al., 2020). The assignment algorithm utilizes several customizable quality assurances checks that include acceptable ranges for the oxygen to carbon ratio (O/C), hydrogen to carbon ratio (H/C), double bond equivalence minus oxygen (DBE-O), the nitrogen rule and more (Schum et al., 2020).

Higher mass analytes will tend to have higher error due to a larger number of possible assignments and decreasing resolution (Koch et al., 2007). This can be combated through internalized mass recalibrations that use CH_2 homologous series and formula extensions with H_2 and O to identify recalibrant ions to identify the most reliable ion masses. These are then used to recalibrate the masses, effectively removing the systematic error for the assignments that are often seen with higher mass species (Schum et al., 2020). This recalibration process produces higher confidence MF assignments with lower ambiguity which is imperative for the statistical analysis method.

1.5.3 Differential analysis

Statistical differential analysis is a common tool in genomics, proteomics, and metabolomics. It can quickly compare two conditions and determine where the significant differences lie. The study of differential expressions of genes relies heavily on this technique and it has been well developed for this purpose (Cui and Churchill, 2003; Zhang and Cao, 2009; Vaes et al., 2009). Commercial packages for differential analysis in metabolomics and proteomics have been developed through existing software such as MZmine (Katajamaa and Orešič, 2005).

Volcano plots are a beneficial tool for visualizing these differences by plotting the $-\log_{10}$(p-value) against the foldchange (Li, 2012). This technique is also used by commercial software such as Compound Discoverer which compares data with existing databases to identify known compounds within the raw Xcalibur data. The software is able assign these compounds and then compare between samples with replicate data (Scarpone et al., 2020). However, software like Compound Discoverer is costly and non-transparent with the post processing method. A post processing method after molecular formula assignment with be ideal for rapid alignment and comparisons.

1.5.3.1 t-Tests

t-Tests are valuable tools for determining significant differences between two variables through defining error variance (Zhang and Cao, 2009; Cui and Churchill, 2003). The universally accepted p-value to prove significance is ≤ 0.05. There are three assumptions that are made when preforming a t-test: 1) the sample is random and representative of the population, 2) the data values are normally distributed in a bell-shaped curve, and 3) there is homogeneity of sample variances. Additionally, it is

generally agreed that a sample size of greater than 30 is generally ideal for an unbiased t-test, however, some statisticians recommend a sample size of at least 100 (Boneau, 1960; Kim and Park, 2019).

Traditionally the *t*-test, Equation 3, follows the format of mean ratio (y_i and x_i) over the square root of pooled variance which assumes equal variance between samples. However, with the small sampling sizes the test needs to be regularized as a global test statistic as seen in Equation 4 where the log of the mean ratio (R_g) over the standard error (*SE*). This version modified *t*-test was performed as part of the data analysis.

$$t = \frac{y_i - x_i}{\sqrt{s_i^2\left(\frac{1}{m_0} + \frac{1}{m_1}\right)}} \qquad (3) \qquad\qquad t = \frac{R_g}{SE} \qquad (4)$$

However, these assumptions are hard to keep when it comes to the modified t-tests used in this study for the differential analysis. In gene expression studies, the gene expressions are measured (i.e. florescence) between the control and test groups (Cui and Churchill, 2003; Noel et al., 2008). In the case of the comparison of blueberry and cranberry SRMs, triplicate species of the same formula and retention time are compared between the two groups resulting in three observations per group. Additionally, the complex mixture of the SRMs results in ~1500 total species that the *t*-test calculation would be applied to. Applying and checking the assumptions of all 1500 *t*-tests would be unreasonable. Verifying the variance between injections would allow for observations with low confidence to be identified. This could be achieved by finding an average abundance across the three injections and calculating the relative standard deviation

(RSD) to prefilter the data set prior to analysis. This would automatically remove species that do not meet the homogeneity requirement, allowing for confidence in the *t*-test results.

1.5.3.2 Fold Change

The fold change (FC) is the difference between the average log ratios of each observation (Equation 3). Contrasting to other statistical analyses, this calculation is not normalized. This may lead to the introduction of bias into the variable; formulas that have smaller abundancies may have a falsely large variance as opposed to formulas with larger abundancies. While this does not make FC an ideal test to determine statistical significance on its own, it is still a valid option for evaluating the confidence in the *t*-test results (Zhang and Cao, 2009; Cui and Churchill, 2003).Additionally, remaining bias could be combated by preselecting the data to exclude smaller abundancies. Visualizing the abundancies in a histogram will give an ideal cut off. Furthermore, there is not a universally accepted result that implies significance Determining the cut-off is up to the discretion of the researcher. A cutoff of two will allow for all observations with a mean difference that is at least two-fold (Cui and Churchill, 2003). For the purposes of this study, a FC of two was selected to filter for significant differences between formulas.

$$FC = \left(\frac{Log_2(C_1) + Log_2(C_2) + Log_2(C_3)}{3}\right) - \left(\frac{Log_2(T_1) + Log_2(T_2) + Log_2(T_3)}{3}\right) \quad (5)$$

2 Materials and Methods

2.1 Sample Preparation

Standard reference materials (SRM) were acquired from the NIST Standard Reference Materials program. SRM 3287 blueberry is comprised of a blend of Tifblue and Rubel variety of blueberries at an approximate ratio of 1:1. This mix was attained as homogenized, freeze-dried, 40-mesh ground powder (NIST, Certificate of Analysis, 2019b). SRM 3281 cranberry consists of frozen berries that were freeze-dried, ground, and homogenized to an 80-mesh powder (NIST, Certificate of Analysis, 2019a). The blueberries and cranberries were obtained by NIST from the U.S. Highbush Blueberry council and Van Drunen Farms respectively (NIST, Certificate of Analysis, 2019a; NIST, Certificate of Analysis, 2019b).

In preparation for analysis, the SRMs were prepared as a 30 g/L solution in 0.1 M hydrochloric acid. The samples were vortexed and then sonicated at room temperature for 1 hour. After sonication, the samples were centrifuged at 12,000 rpm for 20 minutes at 4°C (Lowenthal et.al, 2012). The supernatant was removed for solid phase extraction (SPE).

The SPE extraction was performed under vacuum with Water's Oasis HLB 6cc (200 mg) cartridges at a rate of ~1 mL/min. The cartridges were conditioned with ~6 mL each of Optima LCMS grade methanol, acetonitrile (ACN), and water (H_2O). After conditioning 10 mL of each sample and blank solution was loaded onto the column in 5 mL aliquots and washed with 10 mL of Optima LCMS grade H_2O. The cartridges were extracted with 90% ACN in H_2O and the effluents collected for analysis.

2.2 UHPLC-ES/HRMS Analysis

Data collection was performed in triplicate with a Thermo Fisher Scientific UltiMate 3000 UHPLC. Samples and the corresponding blank were injected onto a C18 column and were separated via gradient off 0.1% formic acid (FA) in H_2O (v/v) (mobile phase A) and ACN (mobile phase B) at a flow rate of 200 μL/min. The gradient started with 95% mobile phase A/5% mobile phase B held for 5 minutes. Mobile phase B was increased to 50% over 30 minutes and then increased to 75% over 2 minutes for a column wash. The gradient was then brought back to 95% mobile phase A/5% mobile phase B over 1 minuet and held until 43 minutes for re-equilibration.

Post column separation, the samples were run through an Ultrahigh resolution Orbitrap Elite mass spectrometer with a range of 100 – 600 Da under negative mode ESI. The data was recorded by Thermo Scientific Xcalibuur software as raw files. The raw files were converted to mzML files via ProteoWizard software. MZmine 2.53 was used to extract the ions for formula assignment.

2.3 Data Post Processing Procedure

2.3.1 Ion Extraction with MZmine 2.53

The mzML files were imported into MZmine 2.53. Mass detection was performed with the Exact Mass detector at a noise intensity level of 250. Next, the Lorenzian Extended Peak Model Function was applied with 240,000 mass resolution to filter FTMS shoulder peaks. Chromatograms were constructed using the ADAP Chromatogram builder. The set parameters allowed for a minimum group size 3, a minimum group intensity of 300, and a minimum highest intensity 350 with a mass error tolerance of 0.001 Da. Chromatogram deconvolution was performed using the ADAP Wavelet

Algorithm with a minimum feature height of 300, a signal to noise ratio of 10, a coefficient/area threshold of 1, a peak duration of 0.00 to 1.50 minutes and a RT wavelet range of 0.00 to 0.10. After chromatogram deconvolution, the data sets were exported as csv files for further analysis in R 4.0.0.

2.3.2 Molecular Formula Assignment with MFAssignR

Formula assignment was performed for the samples and corresponding blanks via MFAssignR 1.0.1. Formula assignment protocol allowed for $^{12}C_{1\text{-}c}$,$^{1}H_{0\text{-}h}$, $^{16}O_{1\text{-}o}$, $^{14}N_{0\text{-}3,}$ and $^{32}S_{0\text{-}1}$ element assignments, a double bond equivalence minus the number of oxygen atoms (DBE-O) of -15 to +15, an H/C ratio between 0.3 to 2.5, an O/C ratio of 0.1 to 2, a de novo cut-off of 300 m/z, and an allowed error of ≤ 3 ppm.

One of the SRM 3287 blueberry injections was discarded due to questionable error trends. It did not follow the trend of the other injections of SRM 3287 blueberry and lead to a lack of confidence in the assignments and the following statistical analysis, this will be discussed further in Chapter 3. Abundances for a third data set were fabricated by averaging the aligned abundances of the remaining SRM 3287 blueberry injections, the process will be defined in the next section.

2.3.3 Data Alignment

Following formula assignment, duplicate peaks from peak splitting were identified and the abundancies were summed together. The formulas and retention times of each observation were compared and peaks with the same formula and a retention time within a 30 second window were tagged as split peaks. Then the rounded and ceiling values for each retention time were identified and the sum (RT_Sum) and difference (RT_Diff) of these values were calculated. If there was a RT_Diff of 1, 1 was added to the RT_Sum. The

abundancies of peaks with the same formula and RT_Sum were summed. Finally, duplicate peaks were identified through a duplicate tag. A new variable was created through combining the formula and RT_Sum, Form_check, and all distinct values were kept.

The injections of SRM 3287 blueberry, SRM 3281 cranberry, and the blank were aligned into their own respective lists allowing for a retention time difference of up to 30 seconds. The protocol was modified from the split peak protocol, the Form_check variable was generated and used for the alignment and removal duplicate peaks. The data sets were then filtered further to only include peaks that appear in at least two of the three injections and to exclude any peaks with an abundance under 1500. Following the alignment between injections, the samples and blanks were aligned into a master list using the same conditions as above. Any missing abundance values remaining after filtering were filled with 1500 for the SRM samples and 500 for the blank.

2.3.4 Statistical Differential Analysis

After alignment, a blank subtraction was performed by averaging the blank abundances and subtracting this from the sample abundances. The RSD of each observation was calculated and used to filter out species with high abundance variance. Assigned formulas with an RSD of $\leq 10\%$ were selected. These new values were used to perform a t-test and obtain a p-value and foldchange for each observation. The data was filtered for species that had a p-value of ≤ 0.05 and a foldchange of $-2 \geq x \geq 2$ to determine species that were significantly different in abundance.

2.3.5 Determination of Polyphenols

The list of significantly different assigned formulas was run against Phenol Explorer Database 3.6 to tentatively identify known phenolic compounds. This was compared against known species for Highbush blueberry, Rabbiteye blueberry, and American cranberry (Neveu et al., 2010; Rothwell et al., 2013). Highbush and Rabbiteye are alternative names for Rubel and Tifblue berries respectively.

3 Results and Discussion

The primary focus of this section is to discuss the method development process and the results of the statistical analyses performed for SRM 3287 blueberry and SRM 3281 cranberry. The sections are ordered as described in Chapter 1 and Chapter 2 as they describe the data analysis and post processing method.

3.1 UHPLC-ES/HRMS Ion Extraction

Initial UHPLC-ES/HRMS scans resulted in vastly different spectra for the two SRM samples. As seen in Figure 3.1 there are areas of broad peaks registering as a single mass value. These areas are most likely due to solvent gradient changes and flatten out with baseline correction. The total ion chromatogram (TIC) directly compares SRM 3287 blueberry and SRM 3281 cranberry features. Though the features of the chromatograms are different, the retention time ranges with the most activity are similar in both samples and correspond with the mobile phase gradient changes outlined in Chapter 2. The ratio of organic to water mobile phase steadily increases after the five-minute mark allowing for the elution of analytes in decreasing polarity prior to the column wash that pushes off the remaining, least polar species.

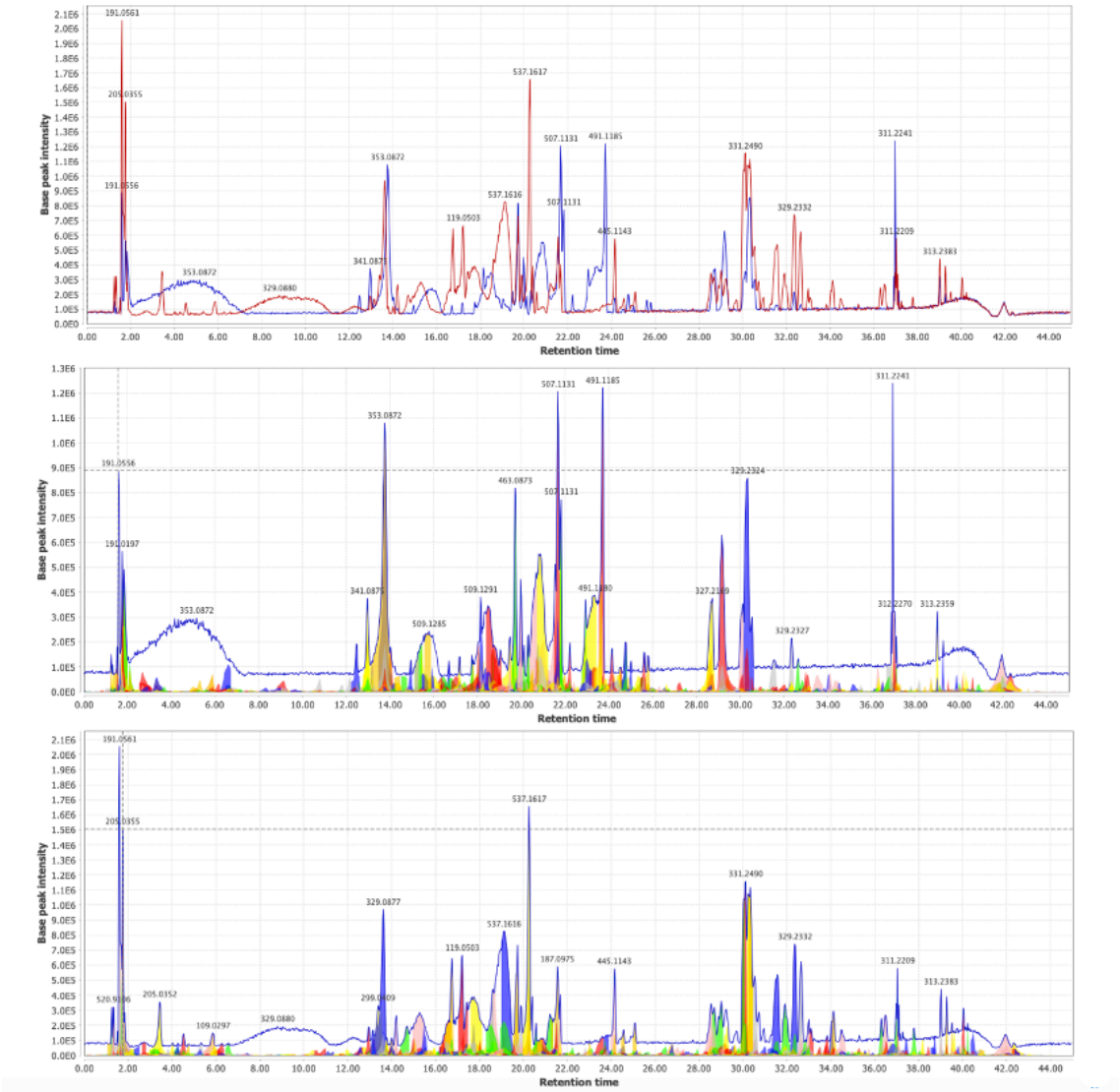

Figure 3.1 A) TIC of SRM 3287 blueberry (blue) and SRM 3281 cranberry (red); B) EIC of SRM 3287 blueberry; C) EIC of SRM 3281 cranberry

After post data processing with MZmine 2.53, approximately 11,000 and 12,000 ion masses were extracted for SRM 3287 blueberry and SRM 3281 cranberry respectively. MZmine 2.53 offers an array of post data processing options including alignment, gap filling, and data filtering. Following alignment with the gap filler method allows for peaks that may have been mis-aligned due insufficient ion extraction to be remedied by estimating peak heights and areas of these missing peaks. Filtering then removes duplicate and triplicate species from the aligned master list (Katajamaa et al., 2006). While these tools are useful for the creation of peak lists for formula assignment and performing subsequent data analysis, in this method, the process doubled the number of ion masses extracted for each sample even after gap filling and filtering. The cause for the increase in extracted ions is unknown. For this reason, it was determined that manually performing these processes in R is optimal for this method. Following ion extraction and chromatogram deconvolution, the feature lists are exported as CSV files to MFAssignR.

3.2 MFAssignR Formula Assignment

This section will focus on the formula assignment and initial comparisons of the SRM 3287 blueberry and 3281 cranberry samples. Formula assignment was performed with MFAssignR allowing for three nitrogen (N) atoms, one sulfur (S) atom, and oxygen (O) within an O/C ration range of 0.1 to 2 for heteroatom assignments. Recalibrant series within 3 ppm mass error were selected to ensure mass accuracy. Final assignments were withing the 3 ppm mass accuracy without allowing for full ambiguity.

3.2.1 Post Processing Data Comparison

Figure 3.2 displays the final error plot for the three SRM 3281 cranberry injections. The error spread increases with mass with the few ambiguous assignments occurring at 300 Da and above. This trend is expected, generally as molecular mass increases, so does the number of possible formulas with a decline in resolution, this gives more opportunities for higher error (Koch et al., 2007).

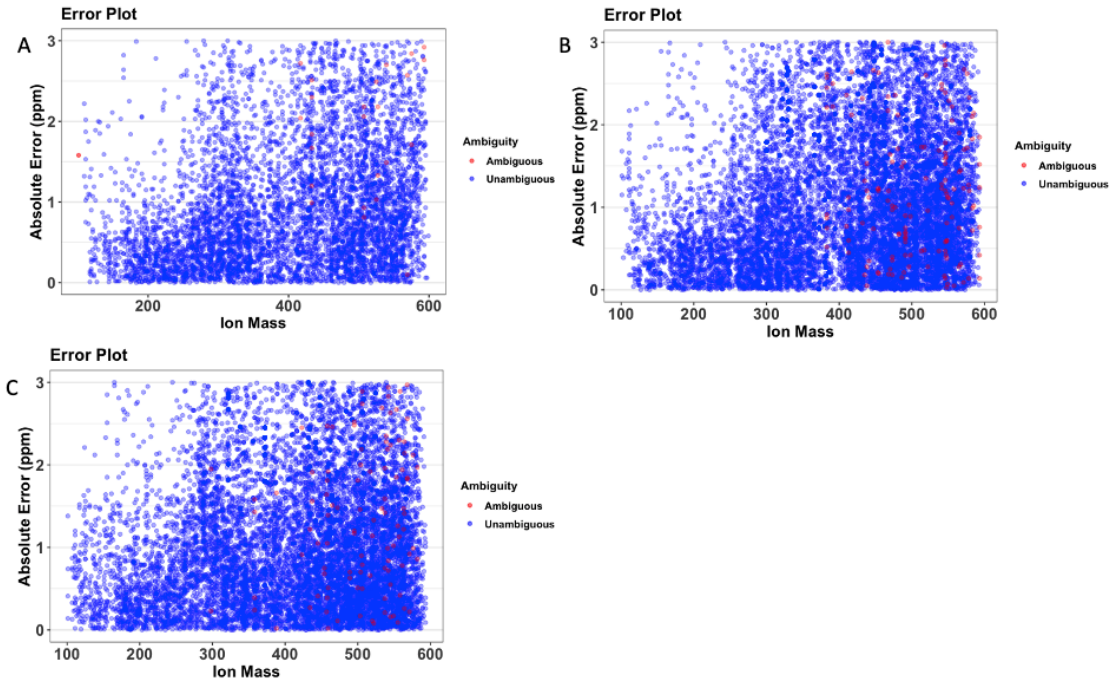

Figure 3.2 Error trend plots for the three injections of SRM 3281 cranberry: A) injection 1, B) injection 2, and C) injection 3. The color represents the ambiguity status, blue is unambiguous, and red is ambiguous.

Final Formula assignment resulted in ~5,500 and ~6,000 unambiguous formula assignments for SRM 3287 blueberry and SRM 3281 cranberry respectively. As seen in Figure 3.3, the two SRM samples are visually different with the highest abundance peaks occurring at different experimental masses. SRM 3287 blueberry had an overall higher number of species with a greater number of nitrogen and sulfur containing species (Table 3.1).

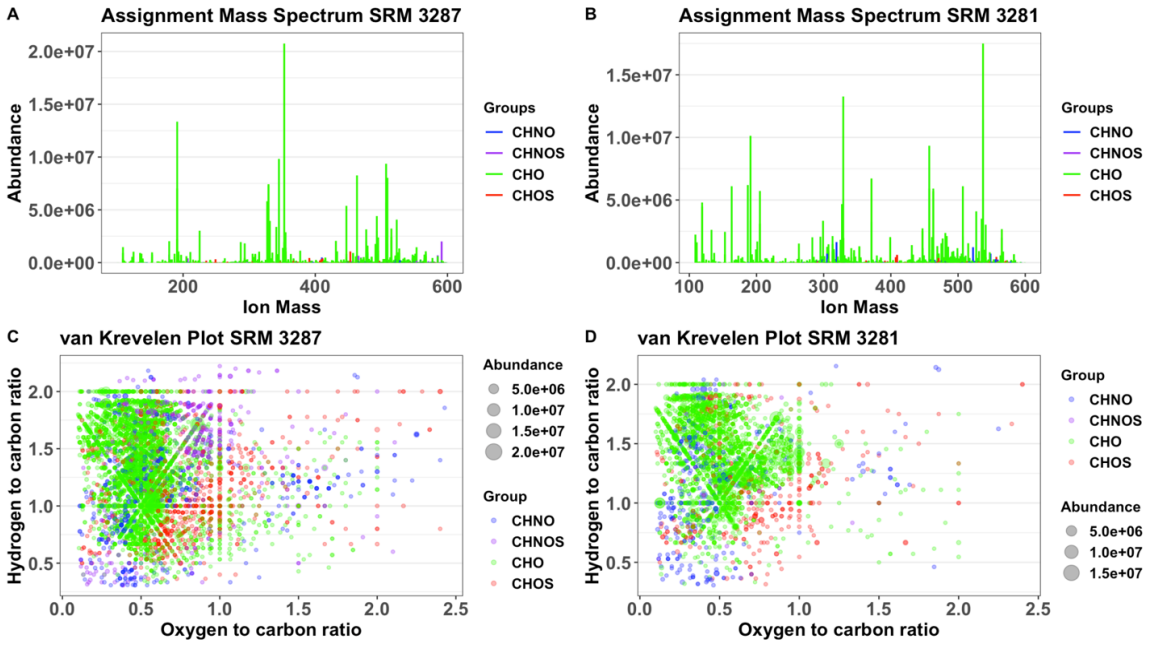

Figure 3.3 SRM samples have been averaged between injections and blank subtracted. Color corresponds to formula assignment group. A) Reconstructed mass spectra of SRM 3287 blueberry; B) Reconstructed mass spectra of SRM 3281 cranberry; C) van Krevelen plot of SRM 3287 blueberry; D) van Krevelen plot of SRM 3281 cranberry.

Table 3.1 number of formulas for SRM 3287 blueberry, SRM 3281 cranberry, common formulas common to both SRM 3287 blueberry and SRM 3281 Cranberry, formulas unique to SRM 3287 blueberry, and formulas unique to SRM 3281 cranberry.

	All SRM 3287 Blueberry Formulas	All SRM 3287 Cranberry Formulas	Formulas Common to SRM 3287 and SRM 3281	Formulas Unique to SRM 3287 Blueberry	Formulas Unique to SRM 3281 Cranberry
CHO	3663	2128	818	2807	1272
CHNO	499	351	57	437	289
CHNOS	424	9	2	422	7
CHOS	729	360	78	708	276
Total	5378	2848	955	4374	1844

van Krevelen plots visualize molecular classes in terms of saturation and polarity through oxygenation. The distinct trends detail the loss of volatile compounds typically lost during fragmentation such as water and hydrogen gas (Perdue and Green, 2015). Figure 3.4 shows van Krevelen plots of assignments that are common and unique to the SRMs. It is noteworthy that while there are some species with an O/C ratio of greater than 1.0, a vast majority fall below that.

Species common between SRM 3287 blueberry and SRM 3281 cranberry are fairly saturated with a majority of the species falling withing the CHO group (86 %). These assignments are clustered in the upper left-hand corner of the plot indicating that they are more saturated with lower oxygen content while N and S containing assignments are more scattered. Species unique to SRM 3287 blueberry are 64% CHO assignments that are clustered to the left side of the plot with more varying degrees of unsaturation. The sulfur containing assignments (CHOS and CHNOS) also have distinct cluster slightly to the right of the CHO cluster and stacked. The trend shows that the CHNOS assignments are trending in higher degrees of saturation than the CHOS assignments. Species unique

to SRM 3281 cranberry are 69% CHO assignments that follows the trends seen in the common species. However, the CHNO assignments are more clustered to the lower left corner indicating high degrees of unsaturation and low oxygenation. This could indicate various aromatic species unique to either SRM 3287 blueberry or SRM 3281 cranberry.

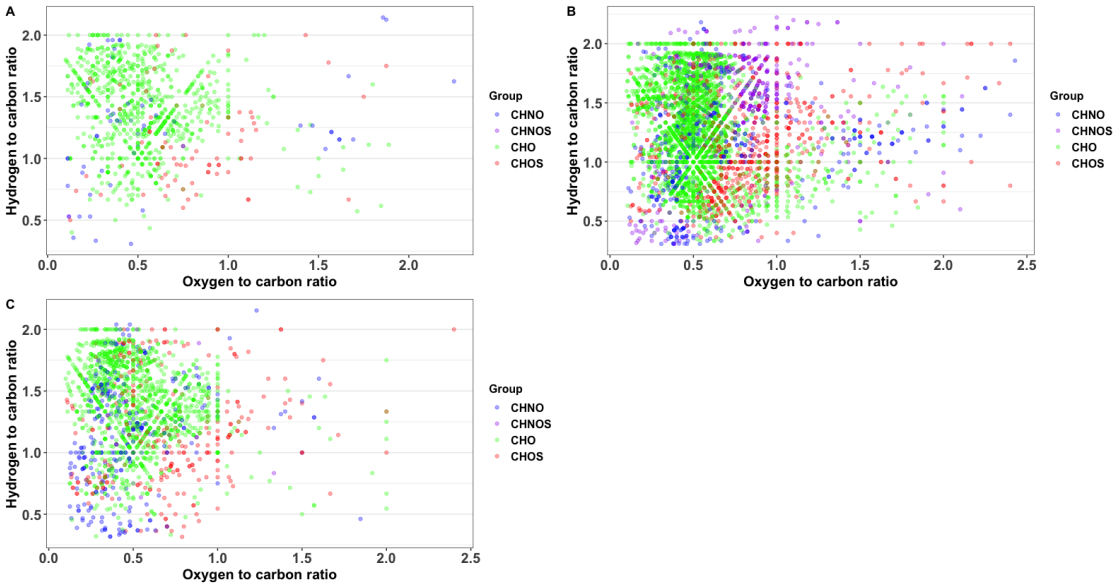

Figure 3.4 van Krevelen plots of A) common species between SRM 3287 blueberry and SRM 3281 cranberry; B) species unique to SRM 3287 blueberry; C) species unique to SRM 3281 cranberry. Color indicates elemental groups.

3.3 Post Processing Data Alignment

Preliminary attempts for alignment did not prove robust enough. Initially, the injections and samples had been aligned by formula and retention time rounded to the nearest tenth. This caused issues with initial attempts for alignment. Assigned peaks were being incorrectly aligned and subsequently filtered out without those abundance differences being accounted for; possibly causing questionable results in the statistical analysis. This was occurring with alignment between injections and samples. To remedy this, two issues had to be addressed: 1) split peaks and 2) retention time windows. To achieve the EICs seen in Figure 3.1, the RT wavelet range had to be extremely low (0-0.1 per Chapter 2.3.1). The RT wavelet range determines the scale of the wavelets that build the EICs. It is suggested that the wavelet range matched the peak duration (Katajamaa et al., 2006; Myers et al., 2017). However, for these data sets, having a smaller peak duration and larger wavelet range did not give good ion extraction and missed prominent peaks in the chromatograms. Conversely, the small wavelet range properly extracted prominent peaks but also resulted in assignments within the same injection that had the identical formulas with slightly different retention times. Usually several within a 30 second retention time window. These duplicate formulas were likely isomeric species that would have been lumped together with a wider RT wavelet range. Tough these species were likely isomers and not the exact same species, having the same formula multiple times within a short time range would lead to complications in the alignment process and misalignments. To simplify the process and ensure proper alignment, these were labeled

as split peaks. The summation of split peaks resulted in ~4,000 unique species for SRM 3287 blueberry and ~4,500 unique species for SRM 3281 cranberry.

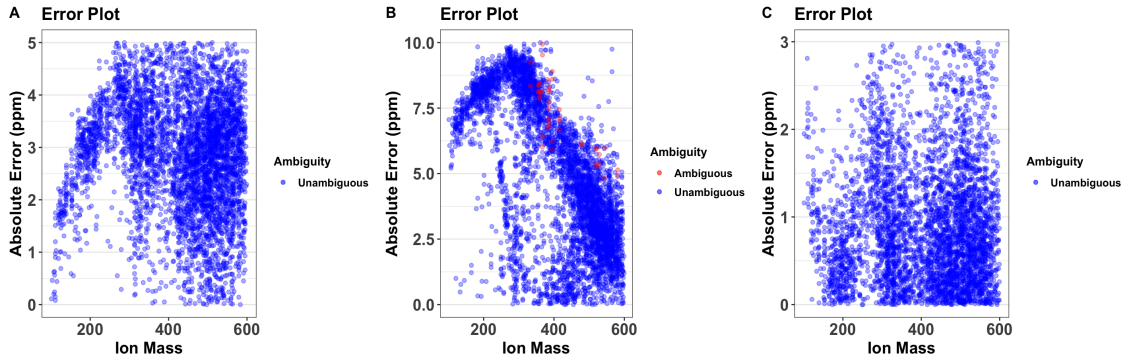

Figure 3.5 Error Trend plots for the three SRM 3287 blueberry injections, A) Injection 1; B) Injection 2; C) Injection 3. Note y-axis scale differences.

As mentioned in Chapter 2, an injection of SRM 3287 blueberry was discarded and a 3rd injection was fabricated through averaging the abundances for each species in the two remaining injections. This fabrication method was performed because triplicate injections are necessary for the differential analysis protocol and there was low confidence in the MF assignments obtained. The injection, plot B in Figure 3.5**,** had an elevated initial error which led to higher final error with fewer assignments. Additionally, looking at the distinct error trend, the initial error was much higher. The error for the injection peaks at 10 ppm in the lower mass range. This is not a typical trend that is expected and is reason to question the validity of the data set and subsequent formula assignments. This is reinforced in the comparison analysis with the other injections, there were only ~400 common injections between the discarded injection and the other SRM 3287 blueberry injections. Comparing the two-lower error SRM 3287 blueberry injections resulted to ~1500 common species. This is sufficient evidence to justify the lack of confidence in the

assignments and warrant excluding this injection from further analysis. While the remaining injections have their own distinct trends, the error peaks at a much lower ppm at a higher mass range, leading to more confidence in the results.

Alignment between the triplicate injections resulted in ~1200 and ~3000 assignments for SRM 3287 blueberry and SRM 3281 cranberry, respectively. The drop in assignments was caused by data filtering. After the injections for each sample were aligned, the data set was filtered to only include species that were found in at least 2/3 of the injections before the blank subtraction occurred. Following this, negative abundances from the blank subtraction and species with a RSD value > 10% were removed. The large drop in the number of SRM 3287 blueberry species was, in part, a result of the data fabrication for the 3rd injection. Due to there being only two viable injections for SRM 3287 blueberry, the data set did not go through the same filtering process before blank subtraction. If a species was present in one injection but not the other, the absentee value (1,500) was averaged with recorded abundance of the other injection. Depending on the magnitude of this abundance, the variance across the three values could be inflated resulting in a large RSD and causing the assignment to be filtered out

3.4 Differential Analysis Data Comparison

In developing the differential analysis method, it is not a viable option to only select common formulas. There are several factors that could contribute to the absence of a peak between samples such as the species being below the detection limit. Additionally, the differential analysis method is meant to determine differences in abundancies between samples, if a species does not exist in one sample, an abundance difference between the samples exists. However, filling missing values with zero is not sufficient for

this method. As seen in Equation 3, the log2(FC) is used to determine the significance of a point and the log2(0) is not a viable number.

To combat this, any positive integer would allow for a real number to be calculated. An abundance of 1500 was filled in for missing SRM species values prior to blank subtraction because pre-filtering removed species with an abundance lower than this. Previously, values of 501 and 500 for SRM and blank species were used respectively. This resulted in many species having a final abundance of 1.0, this led to unusual patterns in the subsequent volcano and variance plots.

3.4.1 *t*-Test and Fold Change Evaluation

After alignment, a tag was applied to the data to determine which sample had a higher overall abundance of the compound. This was calculated through taking the difference of the averaged abundances of each sample.

The modified *t*-test calculations seen in Equation 4 normalizes the standard error variance by removing the assumption that the variances are equal. This allows confidence in the *t*-test procedure despite the small sample sizes (Zhang and Cao, 2009; Cui and Churchill, 2003). A majority of the raw p-values calculated were under 0.6. Figure 3.6 details a histogram of the p-values, it shows as skewed extremely left which is, partly, a product of pre-filtering the data for RSD. Placing a constraint on the variance between injections allowed for more confident p-values and automatically removes misaligned formulas. Additionally, the species higher abundances in SRM 3287 blueberry are more concentrated at the lower end of the p-value range while there is more spread in the SRM 3281 cranberry values.

FC cannot be normalized in the same manner as the *t*-test and therefore on its own cannot be used as a test to determine significance. However, it can act as a verification of significance that has been determined by the *t*-test (Zhang and Cao, 2009; Cui and Churchill, 2003). The FC histogram in Figure 3.6 has more of a normalized bell-curve shape. There is a distinct pattern in this histogram in terms of the higher abundance tag. Species that are higher in abundance in SRM 3287 blueberry fall above zero while species higher in abundance in 3281 cranberry fall below zero with little overlap in between.

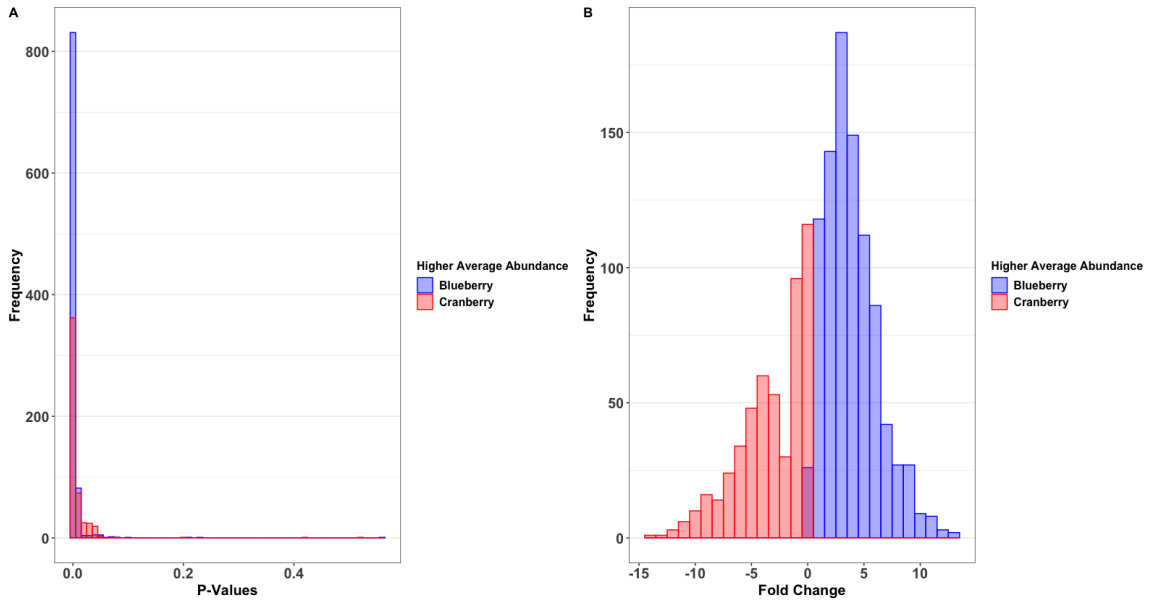

Figure 3.6 Histogram plots of A) p-values and B) fold change values. The color indicates in which SRM the compound is of higher abundance.

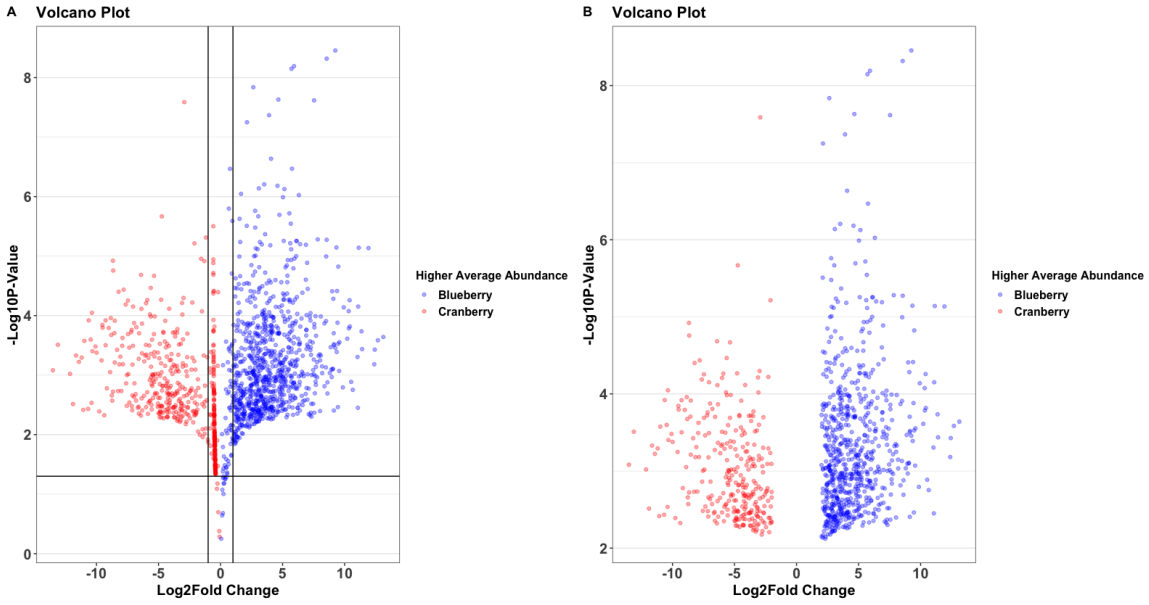

Figure 3.7 Volcano plot comparing SRM 3287 blueberry and SRM 3281 cranberry. The color indicates in which SRM the compound is of higher abundance. A) Initial volcano plot mapping areas of interest; B) Volcano plot of filtered results according to the mapped areas.

Applying the filters for significance in p-value ≤ 0.05 and FC $2 \geq x \geq 2$ resulted in ~1000 statistically significant species. SRM 3281 cranberry had ~300 significantly more abundant species while 3287 blueberry had ~700. Figure 3.7 shows the generated volcano plots pre and post filtering. The species that are more abundant in the SRM 3287 blueberry sample fall on the right side of the graph and the SRM 3281 cranberry on the left. Due to the prefiltering with RSD, there are few species who do not meet the designated filtering parameters, mostly in terms of FC.

The majority of the species have relatively low abundances. Figure 3.8 illustrates this. Additionally, while all these points are statistically significant, all the highest p-values are concentrated at the lower abundances while the species with the highest abundances have extremely low p-values. Though the statistic has been normalized, it can still be expected that the magnitude of the abundance would influence the p-value calculation, though the RSD is below 10%, that 10% variance will have a greater impact on smaller values (Cui and Churchill, 2003).

As seen in Equation 3 and Equation 4, standard deviation has relationship with the p-value which can be described as an inverse square root (Zhang and Cao, 2009; Cui and Churhill, 2003). This is also outlined in Figure 3.8. With the exception of a few outliers, the points follow a distinct trend of RSD increasing with the p-value. These outliers could be the result of issues in the alignment. This is not to say that there is low confidence in the alignment protocol. With large datasets of complex mixtures, it is unreasonable to expect every assignment to be perfect. Additionally, as RSD increases the spread of the p-value ranges increases. This is a reasonable observation, the higher the variance is

between injections, the less confidence there is in the significance (Cui and Churhill, 2003).

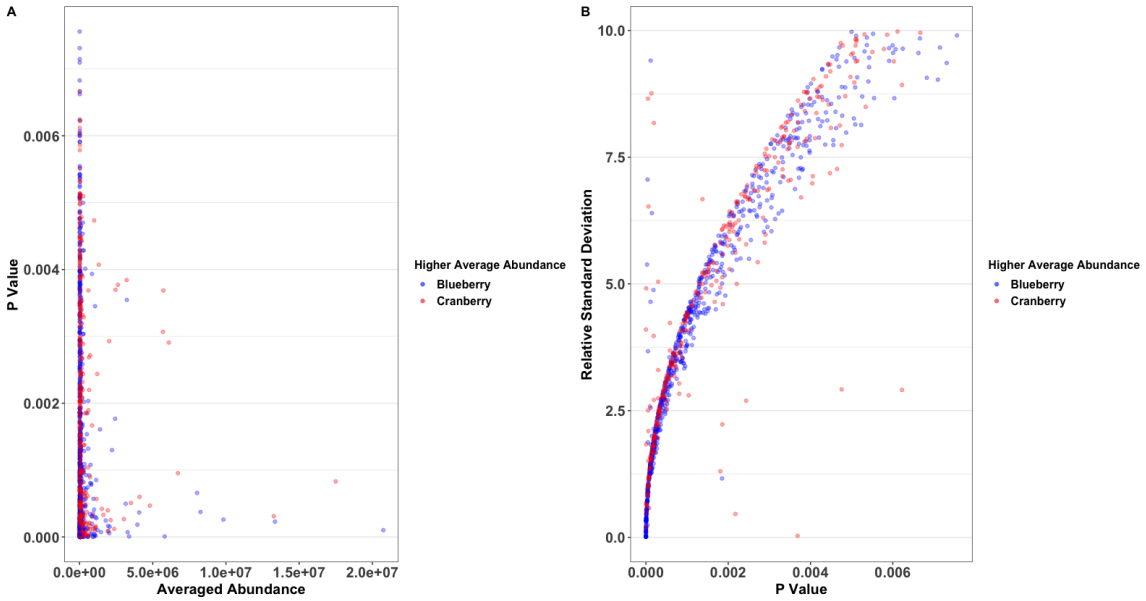

Figure 3.8 Plots visualizing the relationship of the p value with A) average sample abundance; B) relative standard deviation. The color indicates the SRM that has the higher abundance of the species.

3.5 Characterization of Significant Compounds

Figure 3.9 plots the statistically significant species as a Kendrick mass defect (KMD) plot and a van Krevelen plot to identify distinct patterns that differentiate between SRM 3287 blueberry and SRM 3281 cranberry. KMD plots help visualize homologous series through normalizing the masses to a determined molecular fragment, most typically CH_2 (Hughey et al., 2001). However, it can be normalized to other heteroatoms and fragments which rotates the plot and makes it easier to identify what fragments contribute to the series (Chevalier et al., 2019). The KMD plot in Figure 3.9 displays a noticeable pattern of diagonal lines possibly indicating homologous series outside of the typical CH_2 patterns. Another observation that can be drawn from this plot is the spread of masses for

the statistically significant species, especially in terms of the abundance tag mentioned above. The spread of species that are significantly more abundant in either SRM is uniform across the entire mass range showing that mass is not a differentiating factor characterizing the molecular markers for either SRM.

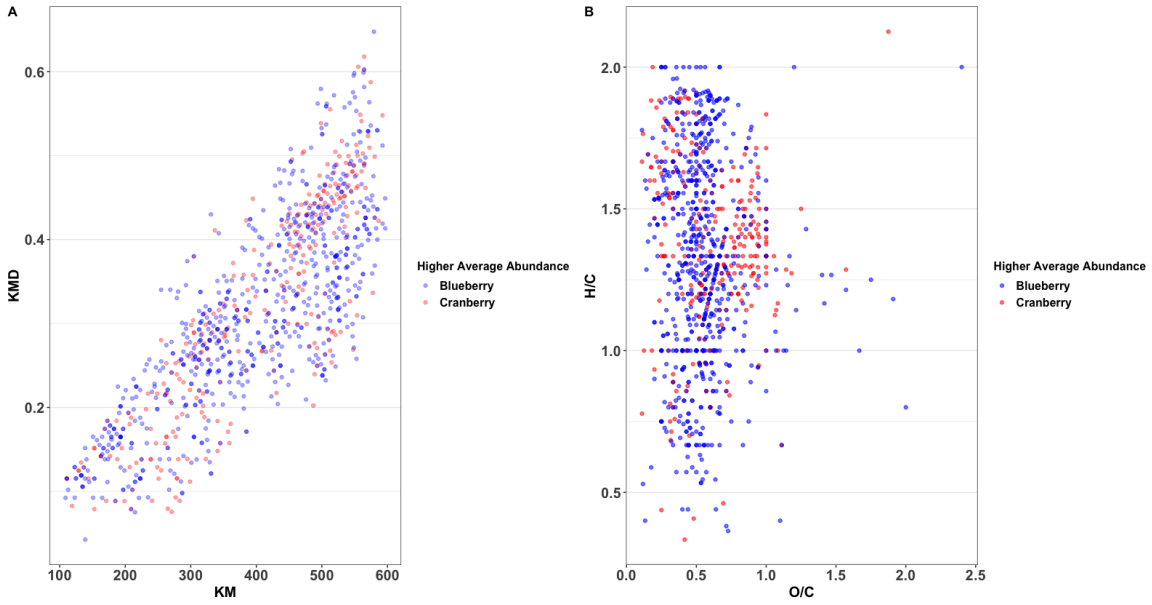

Figure 3.9 A) Kendrick mass defect plot of SRM 3287 blueberry and SRM 3287 cranberry; B) van Krevelen plot of SRM 3287 blueberry and SRM 3287 cranberry. The color indicates the SRM that has the higher abundance of the species.

The van Krevelen plot of Figure 3.9 shows a majority of the significant species with an O/C ratio of less than one and a cluster of species centered on a H/C ratio of 1.5 indicating that a majority of species are more saturated with low oxygenation. There are no discernable patterns in terms of higher average abundance that differentiate between SRM 3287 blueberry and SRM 3281 cranberry.

Figure 3.10 shows the reconstructed mass spectra of SRM 3287 blueberry against the mass spectra for SRM 3281 cranberry. As seen by the comparison analysis performed in 3.2, most significant compounds were CHO assignments with the highest abundance species being flagged as significant.

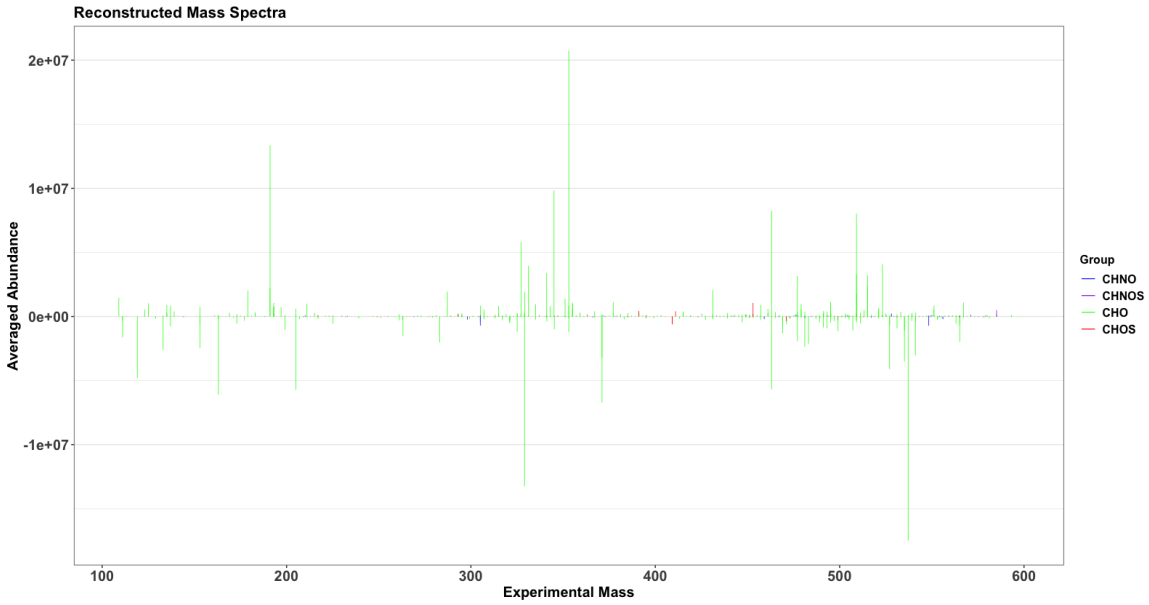

Figure 3.10 Reconstructed mass spectra comparing SRM 3287 blueberry (top) and SRM 3281 cranberry (bottom). Colors are representative of which elemental group the species belong to.

In plotting retention time against abundance as seen in Figure 3.11, most of the high abundance significant species eluted in the 10 to 25 minute retention time range. Additionally the highest abundance species are CHO assignments with an experimental mass ≥300 Da, high oxygenation, and RSD ≤5%. The O/C ratios of the significant species was filtered to exclude and values over 1.0. The few species with an O/C ratio >1.0 were low in abundance and removing these species allowed for better data visualization of oxygen content. As visualized in Figure 3.11 plot C, the most oxygenated species have the best ionization efficiencies due to the polar nature of oxygen atoms. The

most abundant significant species in SRM 3287 blueberry are more oxygenated than the highest abundances species in SRM 3281 cranberry. This also serves as proof that the ion extraction and formula assignment methods were sufficient. The more polar species eluted earlier as expected for a reverse phase column. Figure 3.11 plot D demonstrates the relationship of abundance and retention time in terms of relative standard deviation. The highest abundance species have the lowest RSD values which compliments expectations with variance calculations. Furthermore, This plot visibly illustrates the precision in mass measurements between injections that we would expect from HRMS.

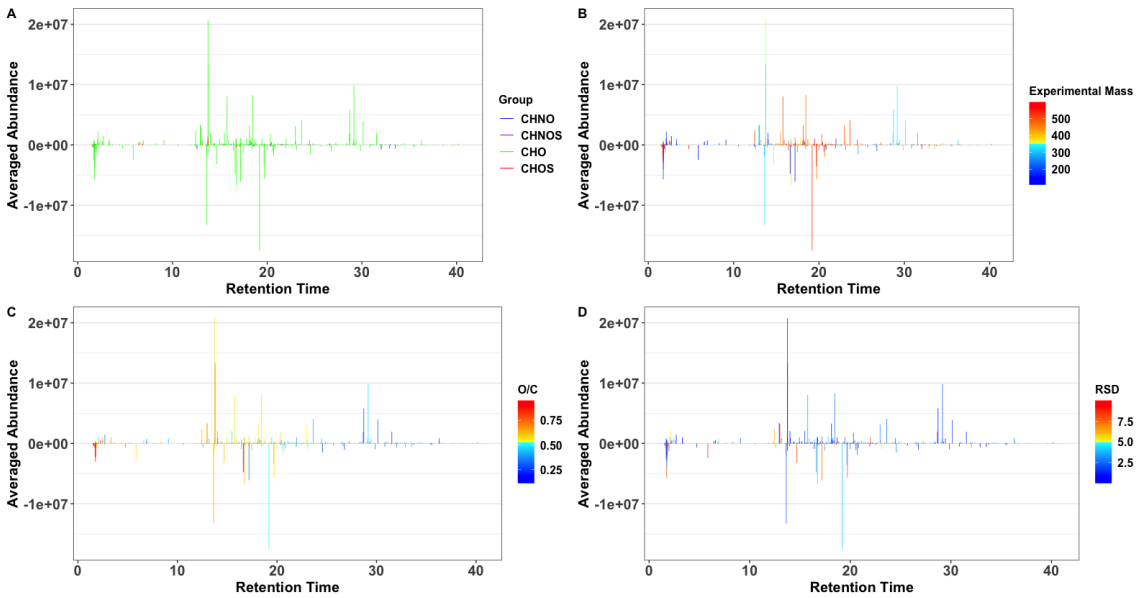

Figure 3.11 Reconstructed tics of SRM 3287 blueberry (top) and SRM 3281 cranberry (bottom) with color indicating A) elemental content; B) experimental mass; C) O/C ratio, filtering out species with an O/C ratio greater than 1.0; D) relative standard deviation.

3.6 Comparisons with Phenol Explorer 3.6

Comparisons with Phenol Explorer 3.6 were carried out through formula matching. These are only tentative matches due to the formula being the only alignment factor. Additional work would have to be performed with supplementary alignment factors and MS/MS analysis to confidently match the formulas with database structures. With this method, 67 tentative database matches were identified with 13 formulas matching with what has been reported in literature for American cranberries, highbush blueberries, and rabbiteye blue berries. A high number of matches were within the flavonoids and phenolic acids compound classes, specifically in the flavanols, hydroxybenzoic acids and hydroxycinnamic acids. Figure 3.12 details the volcano plot and reconstructed mass spectra of the phenolic matches. Similar to the volcano plots above, the points where FC ≥ 2 represent species that are significantly more abundant in SRM 3287 blueberry. The same principle is applied when FC ≤ -2 with SRM 3281 cranberry.

The tallest peaks for SRM 3287 blueberry were identified as 3-Caffeoylquinic acid, Quercetin 3-O-galactoside, and Caffeoyl glucose. The Caffeoyl glucose does not have any literature to support existing in Phenol Explorer 3.6. The highest intensity peak for SRM 3281 cranberry is p-Coumaric acid, which is supported by literature findings. However, the following highest peaks are 4-Vinylphenol and Protocatechuic acid which do not have available literature for the specific berry samples. Table 3.2 outlines the database matches that align with available reports (Neveu et al., 2010; Rothwell et al., 2013) and the level of intensity of each species in terms on SRM 3287 blueberry and SRM 3281 cranberry. As supported in Figure 3.12, a majority of the assignments appear in highest abundance in the SRM 3287 blueberry sample.

Table 3.2 gives a summary of the 67 Phenol Explorer 3.6 matches with the compound class and p-value of each compound. Additionally it gives the relationship between each SRM in terms of low, medium, or high abundance. All the raw p-values for all the compounds are tenfold lower than the 0.05 limit set by the filtering parameters.

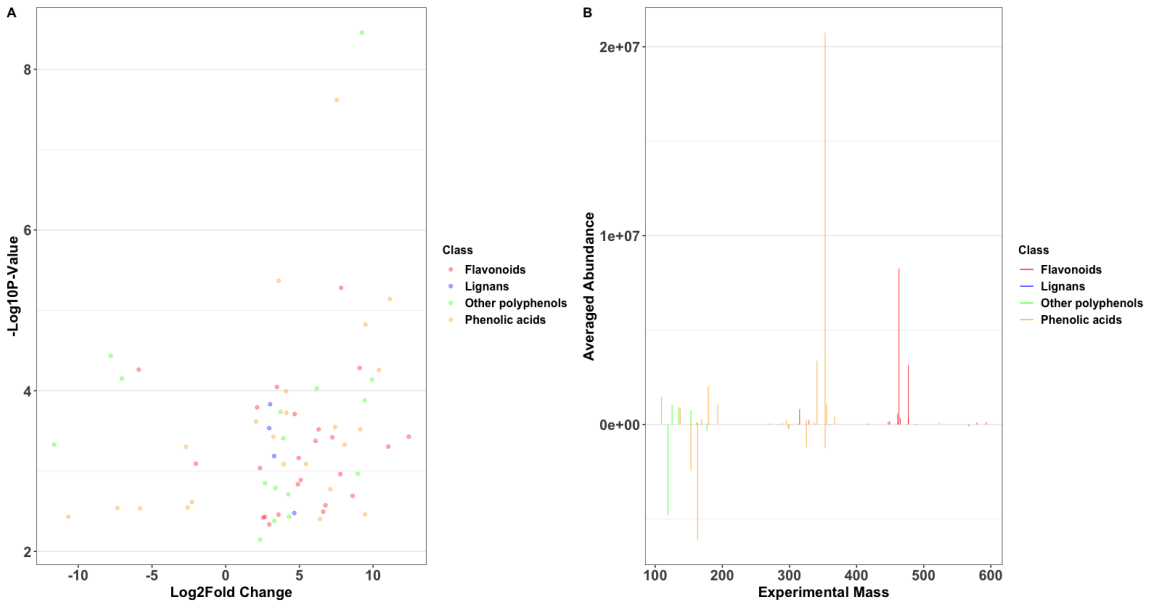

Figure 3.12 A) Volcano plot of tentative Polyphenol Explorer Database 3.6 matches; B) reconstructed mass spectra showing the experimental mass and averaged abundance of tentative Polyphenol Explorer Database 3.6 matches of SRM 3287 blueberry (top) and SRM 3281 cranberry.

Table 3.2 Tentative Phenol Explorer 3.6 database matches. The table includes the name of the compound, what class it belongs to, the raw p-value from the t-test, abundance class, and which SRM had the higher average abundance. Abundance class was classified as: Low abundance as $0 \leq X < 100{,}000$; Medium abundance as $100{,}00 \leq X < 1{,}000{,}000$; and High abundance as $\geq 1{,}000{,}000$.

Name	Compound Class	P-Value	SRM 3287 Blueberry Abundance	SRM 3281 Cranberry Abundance	Higher Average Abundance
Geraldone	Flavonoids	8.09E-04	Low	Low	Cranberry
Phloretin 2'-O-xylosyl-glucoside	Flavonoids	5.46E-05	Low	Low	Cranberry
6''-O-Malonylgenistin	Flavonoids	1.61E-04	Low	Low	Blueberry
Hesperetin	Flavonoids	9.18E-04	Low	Low	Blueberry
Quercetin 3-O-xyloside	Flavonoids	3.79E-03	Low	Low	Blueberry
Hispidulin	Flavonoids	3.70E-03	Low	Low	Blueberry
Quercetin 3-O-(6''-malonyl-glucoside)	Flavonoids	4.61E-03	Low	Low	Blueberry
6''-O-Acetylglycitin	Flavonoids	9.00E-05	Low	Low	Blueberry
Quercetin 3-O-acetyl-rhamnoside	Flavonoids	3.48E-03	Low	Low	Blueberry
Cirsimaritin	Flavonoids	1.95E-04	Low	Low	Blueberry
Luteolin 7-O-glucuronide	Flavonoids	1.46E-03	Low	Low	Blueberry
Butein	Flavonoids	6.88E-04	Low	Low	Blueberry
(+)-Catechin	Flavonoids	1.29E-03	Low	Low	Blueberry
Naringin	Flavonoids	4.21E-04	Medium	Low	Blueberry
Luteolin 7-O-rutinoside	Flavonoids	3.03E-04	Medium	Low	Blueberry
Luteolin 7-O-glucoside	Flavonoids	3.19E-03	Medium	Low	Blueberry
Myricetin 3-O-arabinoside	Flavonoids	2.66E-03	Medium	Low	Blueberry
Jaceosidin	Flavonoids	3.80E-04	Medium	Low	Blueberry
Quercetin 3-O-glucuronide	Flavonoids	1.09E-03	Medium	Low	Blueberry
Dihydromyricetin 3-O-rhamnoside	Flavonoids	5.24E-06	Medium	Low	Blueberry
Chrysoeriol 7-O-glucoside	Flavonoids	2.03E-03	Medium	Low	Blueberry
Nepetin	Flavonoids	5.22E-05	Medium	Low	Blueberry
Isorhamnetin 7-O-rhamnoside	Flavonoids	4.96E-04	High	Low	Blueberry
Quercetin 3-O-galactoside	Flavonoids	3.74E-04	High	Low	Blueberry
Secoisolariciresinol	Lignans	2.92E-04	Low	Low	Blueberry
Conidendrin	Lignans	1.47E-04	Low	Low	Blueberry
Lariciresinol	Lignans	6.51E-04	Low	Low	Blueberry
Syringaresinol	Lignans	3.33E-03	Low	Low	Blueberry
[6]-Gingerol	Other polyphenols	7.06E-05	Low	Medium	Cranberry
Ferulaldehyde	Other polyphenols	3.67E-05	Low	Medium	Cranberry
4-Vinylphenol	Other polyphenols	4.68E-04	Low	High	Cranberry
Ligstroside	Other polyphenols	1.82E-04	Medium	Low	Blueberry

Name	Compound Class	P-Value	SRM 3287 Blueberry Abundance	SRM 3281 Cranberry Abundance	Higher Average Abundance
Isopimpinellin	Other polyphenols	7.15E-03	Low	Low	Blueberry
Guaiacol	Other polyphenols	1.42E-03	Low	Low	Blueberry
3,4-Dihydroxyphenylglycol	Other polyphenols	4.18E-03	Low	Low	Blueberry
Esculin	Other polyphenols	1.62E-03	Low	Low	Blueberry
4-Ethylguaiacol	Other polyphenols	3.91E-04	Low	Low	Blueberry
Esculetin	Other polyphenols	1.95E-03	Low	Low	Blueberry
Curcumin	Other polyphenols	3.70E-03	Low	Low	Blueberry
Umbelliferone	Other polyphenols	9.35E-05	Medium	Low	Blueberry
Hydroxytyrosol	Other polyphenols	1.08E-03	Medium	Low	Blueberry
p-Anisaldehyde	Other polyphenols	3.51E-09	Medium	Low	Blueberry
Pyrogallol	Other polyphenols	1.32E-04	High	Low	Blueberry
Catechol	Other polyphenols	7.27E-05	High	Low	Blueberry
5-O-Galloylquinic acid	Phenolic acids	2.85E-03	Low	Low	Cranberry
Homoveratric acid	Phenolic acids	4.99E-04	Low	Low	Cranberry
Avenanthramide 2p	Phenolic acids	2.89E-03	Low	Medium	Cranberry
p-Coumaroyl glucose	Phenolic acids	2.44E-03	Medium	High	Cranberry
Protocatechuic acid	Phenolic acids	3.70E-03	Low	High	Cranberry
p-Coumaric acid	Phenolic acids	2.91E-03	Medium	High	Cranberry
3-Caffeoylquinic acid	Phenolic acids	1.01E-04	High	High	Blueberry
Cinnamoyl glucose	Phenolic acids	2.42E-04	Low	Low	Blueberry
Sinapic acid	Phenolic acids	3.74E-04	Low	Low	Blueberry
Syringic acid	Phenolic acids	4.29E-06	Low	Low	Blueberry
Feruloyl tartaric acid	Phenolic acids	8.22E-04	Low	Low	Blueberry
Vanillic acid	Phenolic acids	1.89E-04	Low	Low	Blueberry
Homovanillic acid	Phenolic acids	8.18E-04	Low	Low	Blueberry
5-p-Coumaroylquinic acid	Phenolic acids	3.93E-03	Medium	Low	Blueberry
Protocatechuic acid 4-O-glucoside	Phenolic acids	1.68E-03	Medium	Low	Blueberry
p-Coumaroyl tartaric acid	Phenolic acids	2.83E-04	Medium	Low	Blueberry
Gallic acid	Phenolic acids	2.42E-08	Medium	Low	Blueberry
5-Feruloylquinic acid	Phenolic acids	4.68E-04	Medium	Low	Blueberry
4-Hydroxybenzoic acid	Phenolic acids	3.02E-04	Medium	Low	Blueberry
Ferulic acid	Phenolic acids	3.45E-03	High	Low	Blueberry
Feruloyl glucose	Phenolic acids	1.50E-05	High	Low	Blueberry
Caffeic acid	Phenolic acids	5.53E-05	High	Low	Blueberry
Caffeoyl glucose	Phenolic acids	7.24E-06	High	Low	Blueberry

Plotting the phenolic compounds in Phenol Explorer 3.6 in a van Krevelen, Figure 3.13 plot A, shows all the structures clustered around a H/C of 1.0 and O/C of 0.5 which is what would be expected for aromatic, alcohol containing compounds. Figure 3.13 plot B shows the van Krevelen for the 67 phenolic matches from Phenol explorer 3.6, with identified compounds grouped at H/C 1.0. In comparing this to all the statistically significant species, Figure 3.13 plot C, shows several species also in this phenolic region that were not identified with Phenol Explorer. There is a visible difference in the density of points in the specified phenolic region between plots B and C indicating that there are several phenolic species that were not identified in comparisons with phenol explorer 3.6. Furthermore, a large portion of the species that were not matched in phenol explorer appearing to be largely aliphatic or alicyclic. Identification of these species would require other databases.

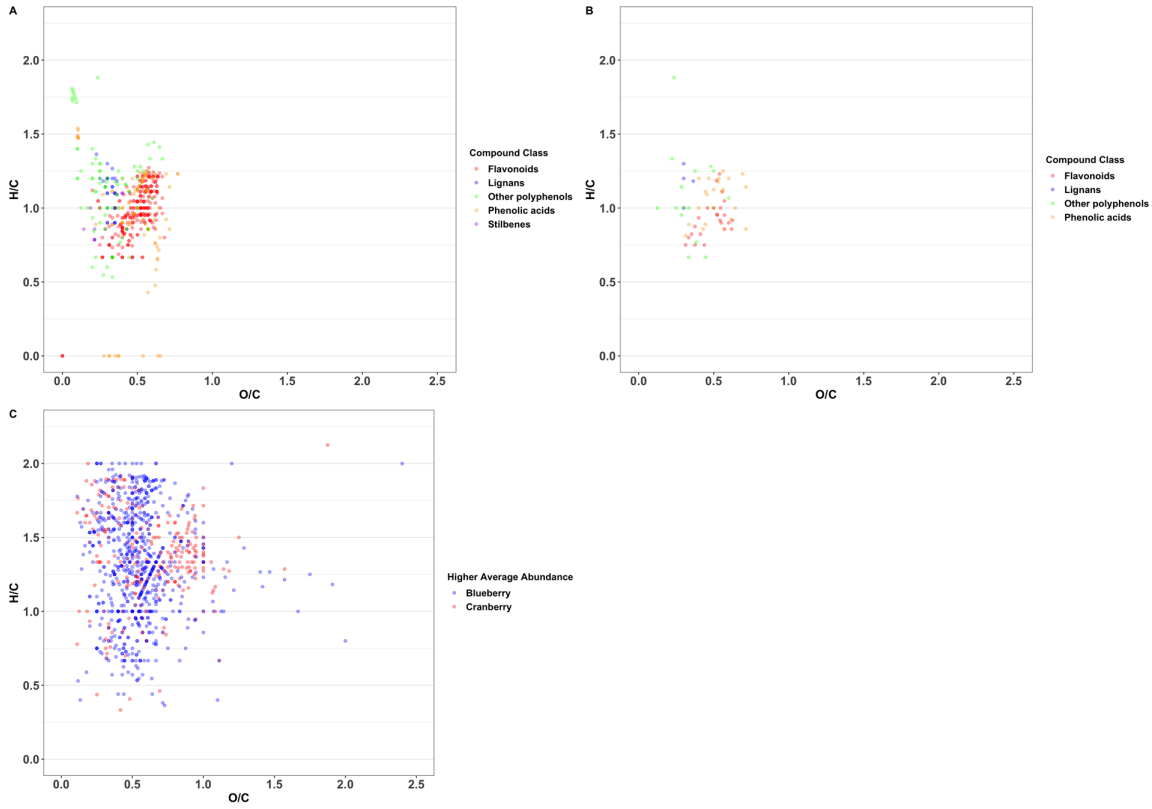

Figure 3.13 van Krevelen plots of A) all compounds in the Phenol explorer 3.6 database; B) Tentative matches from the significant results with the Phenol explorer 3.6 database; C) All formula assignments deemed as significant. The color of plots A and B correspond to the polyphenol class that of the compound. The color of plot C corresponds to which molecule had the higher abundance of the species

4 Conclusions and Recommendations for Future Works

Blueberries and cranberries are widely available species of *Vaccinium* berries with unique flavor profiles. Berries within the *Vaccinium* genus have been studied for their phenolic content and their prevention of cardiovascular illnesses and cancer. Through enhancing extraction methods for phenolic compounds, this study aimed to develop a novel method of non-targeted screening through UHPLC-ES/HRMS coupled with post data processing in MZmine 2.53 and MFAssignR. Investigating the molecular formula assigned species with a modified statistical differential analysis method allowed for the differences in molecular markers of the two NIST SRM samples to be examined.

A major goal of this method development was to be able to enhance the extraction methods to be optimal for the analysis of phenolic compounds as they are of high interest in terms of flavoring components. With this, a robust, transparent alignment and statistical differential analysis method was developed that identified over 1000 statistically significant species between SRM 3287 blueberry and SRM 3281 cranberry.

Of the over 1000 statistically significant species identified through the statistical analysis method, 67 were tentatively matched to the Phenol Explorer 3.6 database. The highest abundance compounds for SRM 3287 blueberry were identified as 3-Caffeoylquinic acid, Quercetin 3-O-galactoside, and Caffeoyl glucose which, like a majority of these assignments, were in the flavonoid and phenolic acid subclasses.

However, this leaves well over 900 assignments that were not database matched, a majority of which are lower abundance species. Through data visualization and analysis, some conclusions can be drawn about these species. Predominantly, the large number of CHO assignments.

In searching for discernable patterns that highlight key differences in the statistically significant species for each SRM, it was found that SRM 3287 blueberry had more statistically significant species in higher abundance. Additionally, the significant SRM 3287 blueberry species had slightly elevated O/C ratios indicating greater polarity over the SRM 3281 cranberry species. These discoveries could be valuable in the future analysis of Michigan blueberries. Especially if these trends are observed across other species of blueberries and varying storage conditions. Developing a deeper understanding the commonalities in the array of flavoring components could determine the impact of these compounds on flavoring. This could influence how Michigan blueberries are currently being used in the market, especially in the alcoholic beverage sector.

In continuing the expansion of this method, validating a non-targeted analysis with positive mode ionizations UHPLC-ES/HRMS method would be an advantageous addition to allow for the statistical analysis of additional phenolic subclasses that are not partial to negative mode ionization. Furthermore, it would be worthwhile to explore p-value adjustment methods such as the Bonferroni correction and the Benjamini-Hochberg procedure. This could decrease the possibility of errors, especially with the extremely small sample size used for the statistical analysis. Additionally, fruit and fruit extract samples have been analyzed and are awaiting post processing. Performing a comprehensive statistical analysis between varying combinations of these samples would be valuable. Especially in the differences of fruit under diverse storage conditions as well as a direct comparison fruit samples against extracts for food additives.

5 Reference List

Abate, G.; Peterson, C. H. rep.; Michigan State University Product Center, 2005.

Blueberries. https://michigangrown.org/blue-berries/ (accessed Jul 27, 2021).

Boneau, C. A. The Effects of Violations of Assumptions Underlying the t Test. *Psychological Bulletin* **1960**, *57* (1), 49–64.

Buszewski, B.; Szultka, M. Past, Present, and Future of Solid Phase Extraction: A Review. *Critical Reviews in Analytical Chemistry* **2012**, *42* (3), 198–213.

Chevalier, M.; Ricart, E.; Hanozin, E.; Pupin, M.; Jacques, P.; Smargiasso, N.; De Pauw, E.; Lisacek, F.; Leclère, V.; Flahaut, C. Kendrick Mass DEFECT Approach Combined to NORINE Database for Molecular FORMULA Assignment Of NONRIBOSOMAL PEPTIDES. *Journal of The American Society for Mass Spectrometry* **2019**, *30*(12), 2608–2616.

Cui, X.; Churchill, G. A. Statistical Tests for Differential Expression in cDNA Microarray Experiments. *Genome Biology***2003**, *4* (4), 210.2–210.10.

Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray IONIZATION-PRINCIPLES and Practice. *Mass Spectrometry Reviews* **1990**, *9* (1), 37–70.

Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt-Kopplin, P. Liquid Chromatography–Mass Spectrometry in Metabolomics Research: Mass Analyzers in Ultra High Pressure Liquid Chromatography Coupling. *Journal of Chromatography A* **2013**, *1292*, 51–65.

Gavrilova, V.; Kajdžanoska, M.; Gjamovski, V.; Stefova, M. Separation, Characterization and Quantification of Phenolic Compounds in Blueberries and Red and Black Currants by Hplc−Dad−Esi-Msn. *Journal of Agricultural and Food Chemistry* **2011**, *59* (8), 4009–4018.

Gilbert, J. L.; Guthart, M. J.; Gezan, S. A.; Pisaroglo de Carvalho, M.; Schwieterman, M. L.; Colquhoun, T. A.; Bartoshuk, L. M.; Sims, C. A.; Clark, D. G.; Olmstead, J. W. Identifying Breeding Priorities for Blueberry Flavor Using Biochemical, Sensory, and Genotype by Environment Analyses. *PLOS ONE* **2015**, *10* (9), 1–21.

Gross, J. H. In *Mass spectrometry: A textbook*; Springer, 2018; pp 246–253.

Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. *Encyclopedia of Analytical Chemistry* **2019**, 1–40.

Holiman, P. C. H.; Hertog, M. G. L.; Katan, M. B. Analysis and Health Effects of Flavonoids. *Food Chemistry* **1996**, *57*(1), 43–46.

Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Analytical Chemistry* **2001**, *73* (19), 4676–4681.

Kalt, W.; Dufour, D. Health Functionality of Blueberries. *HortTechnology* **1997**, *7* (3), 216–221.

Katajamaa, M.; Miettinen, J.; Oresic, M. MZmine: Toolbox for Processing and Visualization of Mass Spectrometry Based Molecular Profile Data. *Bioinformatics* **2006**, *22* (5), 634–636.

Katajamaa, M.; Orešič, M. Processing Methods for Differential Analysis of LC/MS Profile Data. *BMC Bioinformatics***2005**, *6* (1).

Katsube, N.; Iwashita, K.; Tsushida, T.; Yamaki, K.; Kobori, M. Induction of Apoptosis in Cancer Cells BY Bilberry (Vaccinium Myrtillus) and the Anthocyanins. *Journal of Agricultural and Food Chemistry* **2003**, *51* (1), 68–75.

Kim, S.; Kramer, R. W.; Hatcher, P. G. Graphical Method for Analysis OF Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Analytical Chemistry* **2003**, *75* (20), 5336–5344.

Kim, T. K.; Park, J. H. More about the Basic Assumptions of t-Test: Normality and Sample Size. *Korean Journal of Anesthesiology* **2019**, *72* (4), 331–335.

Koch, B. P.; Dittmar, T.; Witt, M.; Kattner, G. Fundamentals of Molecular FORMULA Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter. *Analytical Chemistry* **2007**, *79* (4), 1758–1763.

Kujawinski, E. B.; Behn, M. D. Automated Analysis of Electrospray Ionization Fourier TRANSFORM Ion Cyclotron RESONANCE Mass Spectra of Natural Organic Matter. *Analytical Chemistry* **2006**, *78* (13), 4363–4373.

Kujawinski, E. Electrospray Ionization Fourier TRANSFORM Ion Cyclotron RESONANCE Mass Spectrometry (ESI FT-ICR Ms): Characterization of Complex Environmental MIXTURES. *Environmental Forensics* **2002**, *3* (3-4), 207–216.

Li, W. Volcano Plots in Analyzing Differential Expressions with Mrna Microarrays. *Journal of Bioinformatics and Computational Biology* **2012**, *10* (06), 1231003.

Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A. Think Negative: Finding the Best Electrospray Ionization/Ms Mode for Your Analyte. *Analytical Chemistry* **2017**, *89* (11), 5665–5668.

Lowenthal, M. S.; Phillips, M. M.; Rimmer, C. A.; Rudnick, P. A.; Simón-Manso, Y.; Stein, S. E.; Tchekhovskoi, D.; Phinney, K. W. Developing Qualitative Lc-Ms Methods for Characterization Of Vaccinium Berry Standard Reference Materials. *Analytical and Bioanalytical Chemistry* **2012**, *405* (13), 4451–4465.

Lucci, P.; Saurina, J.; Núñez, O. Trends in LC-MS and LC-HRMS Analysis and Characterization of Polyphenols in Food. *TrAC Trends in Analytical Chemistry* **2016**, *88*, 1–24.

Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Müller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. Ultra High Resolution Linear Ion TRAP Orbitrap Mass Spectrometer (ORBITRAP Elite) Facilitates Top Down LC MS/MS and Versatile Peptide FRAGMENTATION MODES. *Molecular & Cellular Proteomics* **2012**, *11* (3).

Montoro, P.; Tuberoso, C. I. G.; Perrone, A.; Piacente, S.; Cabras, P.; Pizza, C. Characterisation by Liquid CHROMATOGRAPHY-ELECTROSPRAY Tandem Mass Spectrometry of Anthocyanins in Extracts Of MYRTUS Communis L. BERRIES Used for the Preparation OF Myrtle Liqueur. *Journal of Chromatography A* **2006**, *1112* (1-2), 232–240.

Motilva, M.-J.; Serra, A.; Macià, A. Analysis of Food Polyphenols by Ultra High-Performance Liquid Chromatography Coupled to Mass Spectrometry: An Overview. *Journal of Chromatography A* **2013**, *1292*, 66–82.

Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X. One Step Forward for Reducing False Positive and False NEGATIVE Compound Identifications from Mass Spectrometry Metabolomics DATA: New Algorithms for Constructing Extracted Ion CHROMATOGRAMS and DETECTING Chromatographic Peaks. *Analytical Chemistry***2017**, *89* (17), 8696–8703.

Nakajima, J.-ichiro; Tanaka, I.; Seo, S.; Yamazaki, M.; Saito, K. LC/PDA/ESI-MS Profiling and Radical Scavenging Activity OF Anthocyanins in Various Berries. *Journal of Biomedicine and Biotechnology* **2004**, *2004* (5), 241–247.

National Institute of Standards & Technology Certificate of Analysis Standard Reference Material 3281 Cranberry (Fruit), 2019a.

National Institute of Standards & Technology Certificate of Analysis Standard Reference Material 3287 Blueberry (Fruit), 2019b.

Neveu, V.; Perez-Jimenez, J.; Vos, F.; Crespy, V.; du Chaffaut, L.; Mennen, L.; Knox, C.; Eisner, R.; Cruz, J.; Wishart, D.; Scalbert, A. Phenol-Explorer: An Online Comprehensive Database on Polyphenol Contents in Foods. *Database***2010**, *2010*.

Perdue, E. M.; Green, N. W. Isobaric Molecular Formulae of C, H, and O: A View from the NEGATIVE Quadrants of Van KREVELEN SPACE. *Analytical Chemistry* **2015**, *87* (10), 5079–5085.

Perestrelo, R.; Silva, P.; Porto-Figueira, P.; Pereira, J. A. M.; Silva, C.; Medina, S.; Câmara, J. S. QuEChERS - Fundamentals, RELEVANT IMPROVEMENTS, Applications and Future Trends. *Analytica Chimica Acta* **2019**, *1070*, 1–28.

Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass SPECTROMETRY-BASED Molecular Profile Data. *BMC Bioinformatics* **2010**, *11* (1).

Poole, C. F. New Trends in Solid-Phase Extraction. *TrAC Trends in Analytical Chemistry* **2003**, *22* (6), 362–373.

Poole, C.; Fanali, S.; Haddad, P. R.; Poole, C. F.; Riekkola, M.; Nováková, L.; Svoboda, P.; Pavilk, J. In *Liquid chromatography (second edition)*; Elsevier: Amsterdam; Vol. 1, pp 719–769.

Rothwell, J. A.; Perez-Jimenez, J.; Neveu, V.; Medina-Remon, A.; M'Hiri, N.; Garcia-Lobato, P.; Manach, C.; Knox, C.; Eisner, R.; Wishart, D. S.; Scalbert, A. Phenol-Explorer 3.0: A Major Update of The Phenol-Explorer Database to Incorporate Data on the Effects of Food Processing on POLYPHENOL CONTENT. *Database* **2013**, *2013*.

Scarpone, R.; Rosato, R.; Chiumiento, F.; Cipolletti, C.; Sergi, M.; Compagnone, D. Preliminary Study to Develop an Alternative Method for the Non-Targeted Determination of Xenobiotics in Food by Means of Ultra High Performance Liquid Chromatography Coupled to High Resolution and Accuracy Mass Spectrometry. *Food Analytical Methods* **2020**, *13* (5), 1099–1110.

Schum, S. K.; Brown, L. E.; Mazzoleni, L. R. MFAssignR: Molecular Formula Assignment Software for Ultrahigh Resolution Mass Spectrometry Analysis of Environmental Complex Mixtures. *Environmental Research* **2020**, *191*, 110114.

Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A. Fourier Transform Mass Spectrometry. *Molecular & Cellular Proteomics* **2011**, *10* (7).

Taruscio, T. G.; Barney, D. L.; Exon, J. Content and Profile of FLAVANOID and Phenolic Acid Compounds in Conjunction with the Antioxidant Capacity for a

Variety of NorthwestVacciniumBerries. *Journal of Agricultural and Food Chemistry* **2004**, *52* (10), 3169–3176.

Vaes, E.; Khan, M.; Mombaerts, P. Statistical Analysis of DIFFERENTIAL Gene Expression Relative to a Fold Change Threshold on NanoString Data of Mouse Odorant Receptor Genes. *BMC Bioinformatics* **2014**, *15* (1), 39.

Waters. Quechers dispersive solid phase extraction. https://www.waters.com/waters/en_US/QuEChERS-Dispersive-Solid-Phase-Extraction/nav.htm?cid=10072671 (accessed Jul 27, 2021c).

Waters. HPLC separation MODES. https://www.waters.com/waters/en_US/HPLC-Separation-Modes/nav.htm?cid=10049076&locale=en_US (accessed Jul 15, 2021).

Waters. Oasis sample Extraction Products. https://www.waters.com/waters/en_US/Waters-Oasis-Sample-Extraction-SPE-Products/nav.htm?cid=513209&locale=en_US (accessed Jul 27, 2021).

Zhang, S.; Cao, J. A Close Examination of Double Filtering with Fold Change and t Test in Microarray Analysis. *BMC Bioinformatics* **2009**, *10* (1).

Zwir-Ferenc, A.; Biziuk, M. Solid Phase Extraction Technique - Trends, Opportunities and Applications. *Polish Journal of Environmental Studies* **2006**, *15* (5), 677–690.

A Method Development

This appendix will cover experiments that contributed to the method development outlined in this paper. There was a total of six sample preparations and seven sample analyses that were performed to create this method with ample room for future expansions on sample analysis.

A.1 UHPLC-ES/HRMS without Analyte Isolation

Method development had initially used Bberri® 100% blueberry fruit juice and Brewers Best natural flavoring brew syrup without analyte isolation. The UHPLC-ES/MS method employed +ESI and the mobile phase did not contain any acid caused poor ionization. The method was adapted from a 2011 study by Gavrilova, et al. There were prevalent issues with interferences and ambiguity as seen by Figure A.1. A large portion of the peaks were not assignable including the most abundant species for both samples.

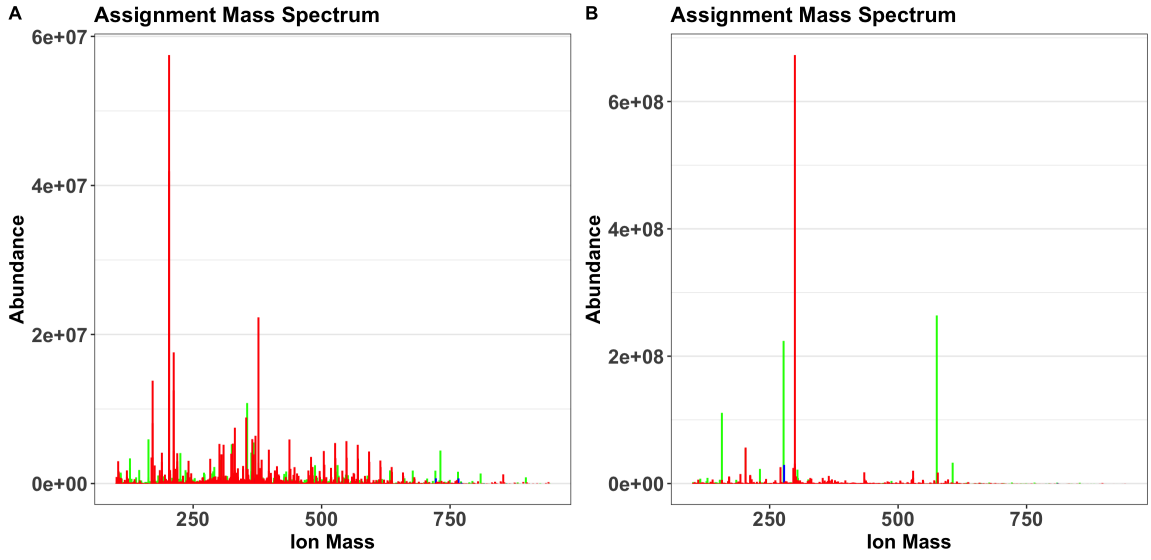

Figure A.1 Assignment mass spectra of A) Bberri® 100% blueberry fruit juice and B) Brewers Best natural flavoring brew syrup. The colors reference the status of assignment: Green is assigned, red is unassigned, and blue is ambiguous.

A.2 UHPLC-ES/HRMS with QuEChERS Analyte Isolation

Quick, easy, cheap, effective, rugged, and safe (QuEChERS) is an analyte isolation method most typically used for analyzing pesticide residues on fruits and vegetables (Waters, 2021c). However, there have been advancements in the method for targeting other matrixes such as phenolic compounds. DisQuE^{TM} pouches for 50 mL solid phase extraction with 0.5 g sodium sesquihydrate, 1 g sodium citrate, 1g sodium chloride, and 4 g magnesium sulfate ($MgSO_4$) paired with 2mL tubes with 150 mg magnesium sulfate, 25 mg polar surface area (PSA), and 2.5 mg graphitized black carbon (GBC) were chosen with a solvent system of 1:1 ethyl acetate:ACN (Waters, 2021c; Perestrelo et al., 2019). This method was performed with Bberri® 100% blueberry fruit juice and Brewers Best natural flavoring brew syrup at two concentrations, 15 g and 1.5 g of sample. Additionally, a blueberry fruit sample (stored at -20°C, thawed, and filtered through cheese cloth) was prepared at 1.5 g of sample. However, as observed in Figure A.2, no meaningful data can be extracted from the collected data sets. All sample runs were comparable to the blank. The LC-ES/HRMS with +ESI is the same method as used above (Gavrilova et al., 2011).

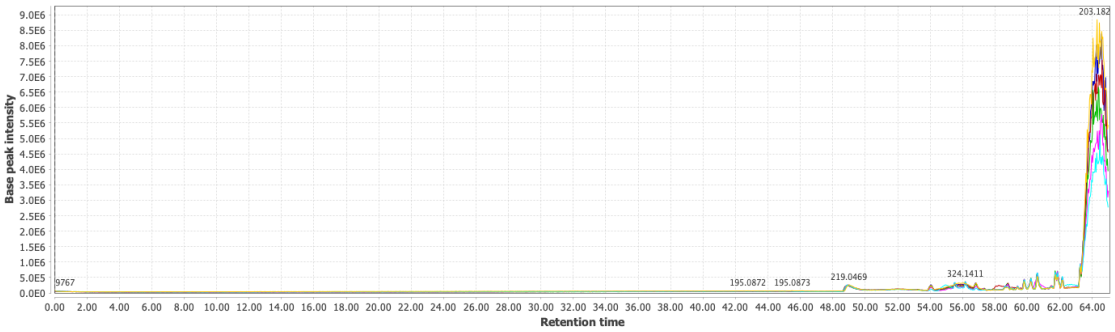

Figure A.2 TIC of QuEChERS sample preparations with UHPLC-ES/HRMS, color indicates the sample. Blue represents the preparation of 15 g of Bberri® 100% blueberry fruit juice; red represents the preparation of 15 g of Brewers Best natural blueberry flavoring; green represents the preparation of 1.5 g of Bberri® 100% blueberry fruit juice; pink represents the preparation of 1.5 g of Brewers Best natural blueberry flavoring; teal represents the preparation of 1.5 g of frozen (-20°C) blueberry paste strained through cheese cloth. Yellow represents the solvent blank (1:1 ethyl acetate:ACN) preparation.

A.3 DI-HRMS with SPE Analyte Isolation

SPE analyte isolation was performed with Bberri® 100% blueberry fruit juice and Brewers Best natural flavoring brew syrup according to the guidelines set out in (Curtis, 2013). Effluents were collected with two solvent schemes: 100% ACN and 90% ACN in water. Samples were analyzed using direct injection (DI) HRMS to ensure the analyte isolation method was sufficient. Figure A.3 shows assignment mass spectra of first extract (100% ACN) of the Bberri® 100% blueberry fruit juice sample. While the figure shows evidence of analyte isolation and is visually unique from Figure A.1, there are still parallel issues with increased noise levels and unassignable peaks as seen in section 5.1.

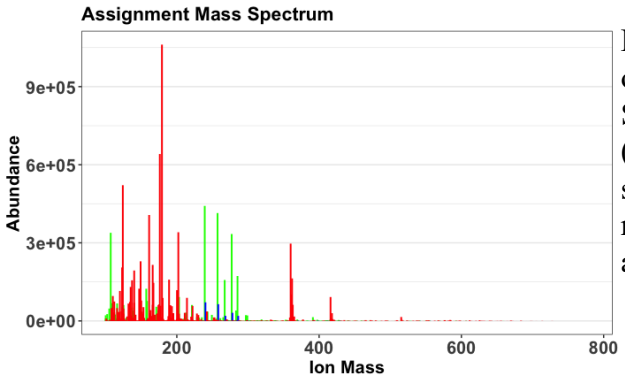

Figure A.3 Reconstructed mass spectra of Bberri® 100% blueberry juice with SPE analyte isolation, first extraction (100% ACN). The colors reference the status of assignment: Green is assigned, red is unassigned, and blue is ambiguous.

A.4 UHPLC-ES/HRMS with SPE Analyte Isolation

The samples prepared in section 5.3 were re-used to perform UHPLC-ES/HRMS analysis with positive mode ionization. The TICs in Figure A.4 correspond to the Bberri® 100% blueberry fruit juice and Brewers Best natural flavoring brew syrup. Though analytes were isolated, there needs to be additional extraction steps to improve the method. The two solvent extraction methods used (100% ACN and 90% ACN in water) show similar results, however, the second extraction with 90% ACN extracted a

large portion of analytes seen from the first extraction as well as improved isolation of water-soluble analytes. However, the peaks from the TICs were unresolved with poor peak shape. This caused doubts for using 100% blueberry fruit juice as the best sample for method development and suggested more sample preparation was necessary prior to SPE.

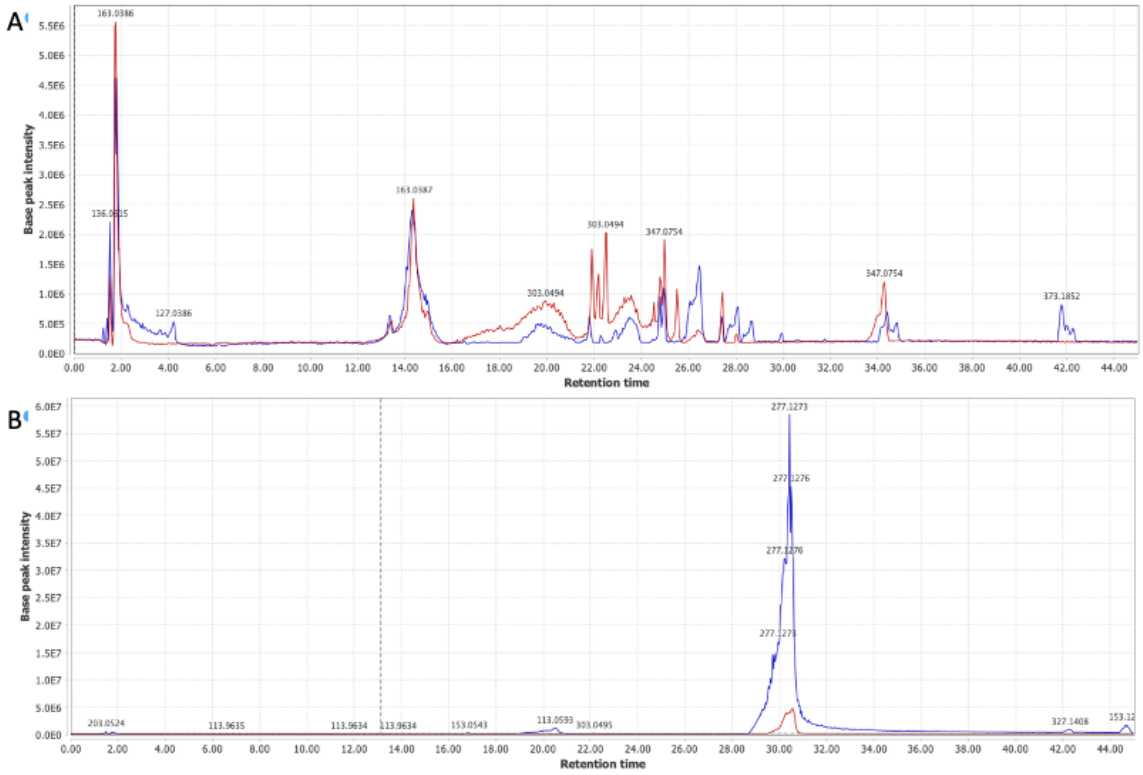

Figure A.4 TIC of A) Bberri® 100% blueberry juice and B) Brewers Best natural blueberry flavoring brew syrup. Color indicated extraction solvent, blue is the first extraction with 100% ACN and red is the second extraction with 90% ACN in water.

A.5 Low Resolution MS with Adjusted SPE Analyte Isolation

SPE analyte isolation was performed with the NIST SRM 3287 blueberry and SRM 3281 cranberry following the preparation method outlined in Lowenthal, et al. from 2012. Analysis was performed with low-resolution MS to look for chlorine interference from the sample preparation with both positive and negative ionization modes. Figure A.5 shows the low-resolution mass spectra with obvious chlorine interferences present in the blank. From this, the analyte isolation method was improved by increasing the amount of water used in washing the samples before the extraction step in sample preparation.

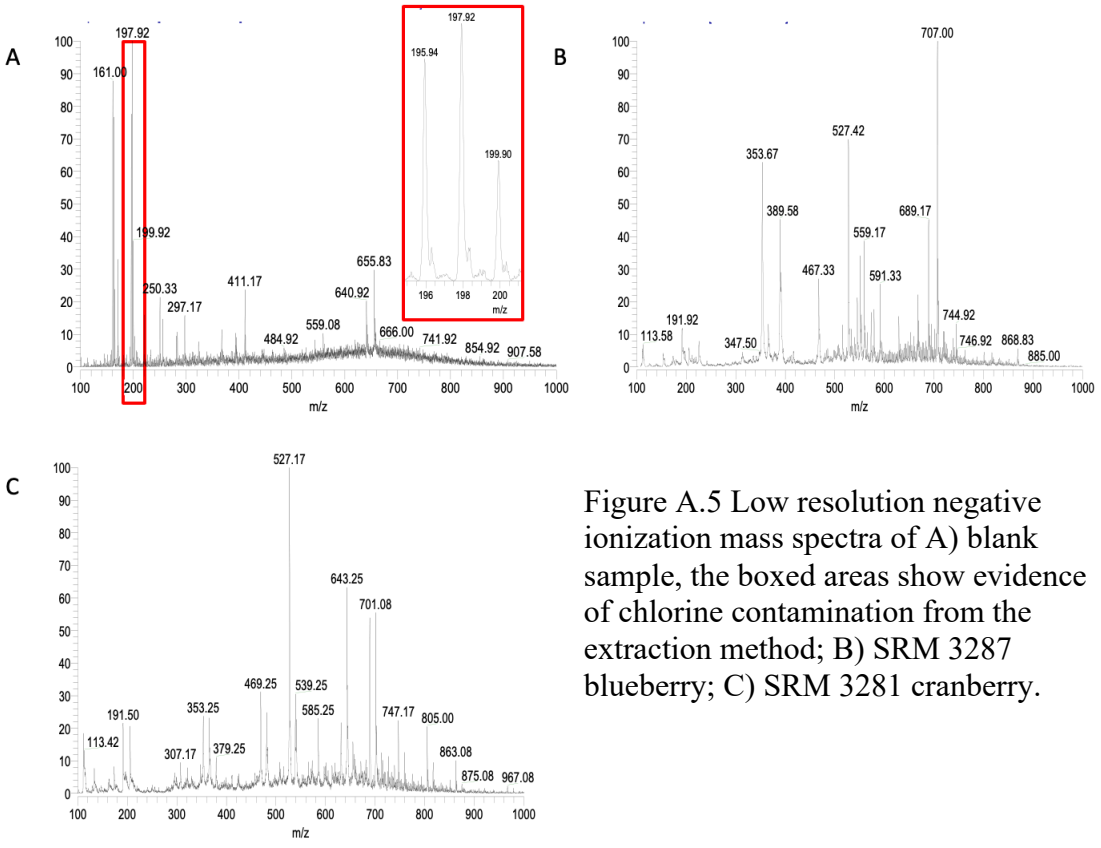

Figure A.5 Low resolution negative ionization mass spectra of A) blank sample, the boxed areas show evidence of chlorine contamination from the extraction method; B) SRM 3287 blueberry; C) SRM 3281 cranberry.

A.6 UHPLC-ES/HRMS with Adjusted SPE Analyte Isolation

The NIST SRM samples used in this study were prepared with blueberry fruit, cranberry fruit, Brewer's Best natural blueberry flavoring, and Brewer's Best natural cranberry flavoring at concentration of 30 g/L (Lowenthal et al., 2012). The UHPLC-ES/HRMS method used negative mode ionization in hopes of reducing noise interferences and increasing ionization efficiency with the phenolic compounds. This method is outlined in Chapter 2. The UPHPLC method had been changed to include FA in the mobile phase and adjusted the gradient to be more efficient with a shorter run time. While SRM 3287 blueberry and SRM 3281 cranberry were analyzed for this study, the fruit and flavoring samples were too dilute to draw any meaningful data from (Figure A.6).

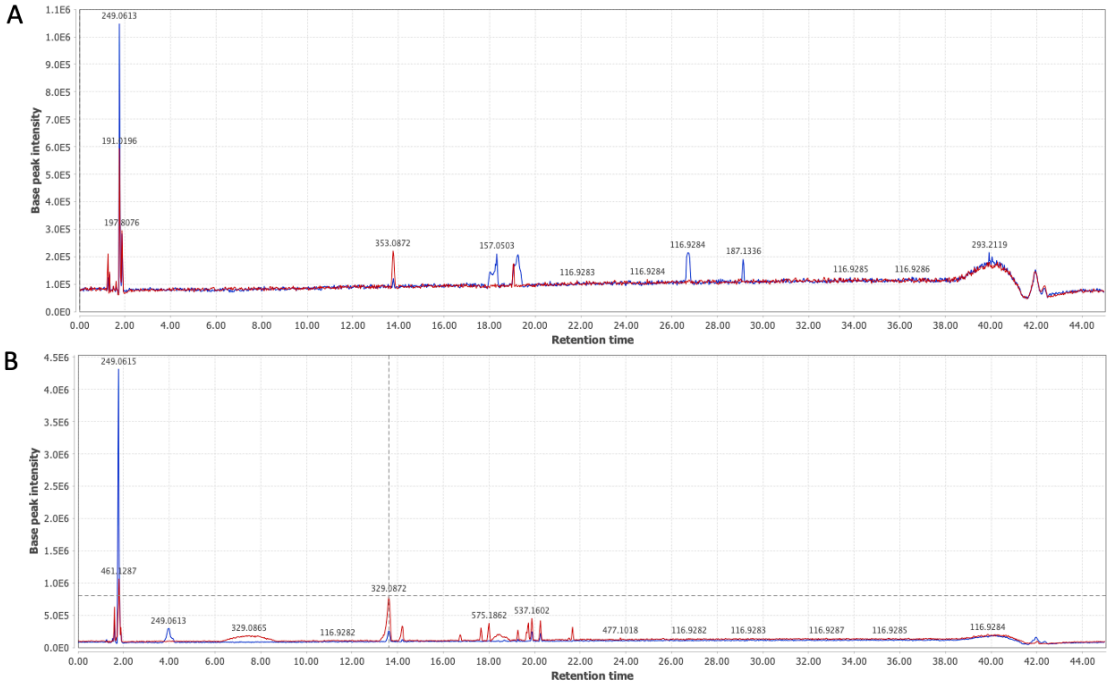

Figure A.6 TIC of A) blueberry-based samples, blue indicates Brewers Best natural blueberry flavoring prepared at 30 g/L. Red represents blueberry fruit paste stored at -20°C and strained through cheese cloth prepared at 30g/L; B) cranberry-based samples, blue indicates Brewers Best natural cranberry flavoring prepared at 30 g/L. Red represents cranberry fruit paste stored at -20°C and strained through cheese cloth prepared at 30g/L.

A.7 Re-Preparation of Non-SRM Samples

The blueberry fruit, cranberry fruit, Brewer's Best natural blueberry flavoring, and Brewer's Best natural cranberry flavoring were re-prepared by increasing the concentration of the samples by a factor of five (150 g/L). The same UHPLC-ES/MS method in Section 5.7 and Chapter 2 was used. Full post data processing has yet to be performed on these samples, however from the TICs shown in Figure A.7 illustrate the differences between the SRM, fruit paste, and Brewer's Best natural flavorings. The TICs show some analogous features between the samples but specifically between the SRMs and the fruit paste samples. This is to be expected because the SRMs are comprised of freeze-dried berries, the differences could be accounted for by variances between species of blueberries and cranberries along with storage conditions.

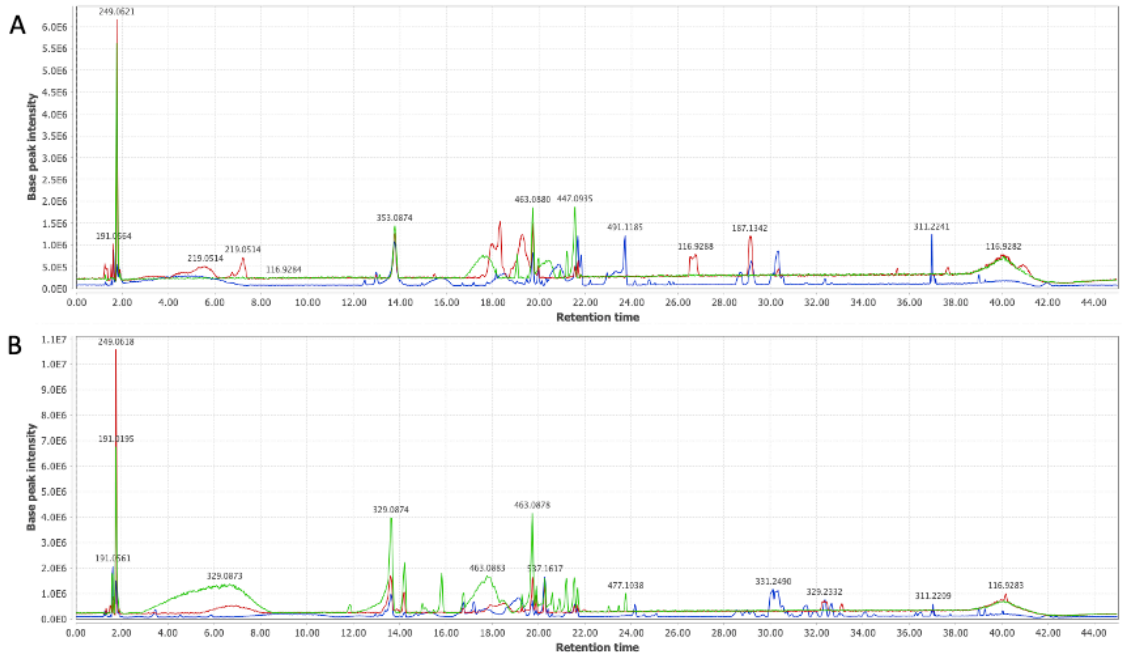

Figure A.7 TIC of A) blueberry-based samples, blue represents SRM 3287 blueberry prepared at 30 g /L, red represents Brewers Best natural blueberry flavoring prepared at 150 g/L. Green represents blueberry fruit paste stored at -20°C and strained through cheese cloth prepared at 150 g/L; B) Cranberry-based samples, blue represents SRM 3281 cranberry prepared at 30 g /L, red represents Brewers Best natural cranberry flavoring prepared at 150 g/L. Green represents cranberry fruit paste stored at -20°C and strained through cheese cloth prepared at 150 g/Land strained through cheese cloth prepared at 30g/L.

B Copyright documentation

Images used in the figures for Chapter 1 come from copy written material and have been licensed for use in this thesis.

Figure 1.1 Orbitrap Elite Mass Spectrometer Thermo Scientific, (Hecht et al., 2019)

Licensed CC-BY-Copyright Clearance Center's RightsLink via John Wiley and Sons- https://onlinelibrary.wiley.com/doi/10.1002/9780470027318.a9309.pub2. Accessed July 2020.