LC/MS OPTIMIZATION AND IMPLEMENTATION FOR METABOLOMIC APPLICATIONS

By

NOELLE MARIE ELLIOTT

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

To my loving and supportive family

ACKNOWLEDGMENTS

I would like to thank those people who have supported me both academically and emotionally throughout my graduate studies. First, I would like to thank my advisors. I am thankful to Dr. David Powell for the patience shown to me during my time in his lab, the insightful discussions, and the knowledge that was shared. I am grateful to Dr. Richard Yost for the academic advice that was offered to me and the interesting ideas that were offered. I also appreciate the time spent discussing metabolomics with Dr. Peggy Borum and the samples provided to me by her. I would also like to thank Dr. Ben Smith and Dr. Nick Polfer for their helpful comments while serving on my committee. I am grateful for the samples provided to me by Dr. John Koomen and Dr. Erin Siegel at Moffitt Cancer Center, as well as the opportunity to work in John's lab for several days.

The friends that I have made in graduate school deserve to be acknowledged for their support both in the lab and outside of it. The Powell and Yost groups have been a supportive second family to me during my time at UF. I am grateful for the thorough training given to me by Dr. Soledad Cerutti. Dr. Julia Rummel should also be thanked for her scientific insight and help early in my graduate studies. The Yost group was particularly helpful with editing of documents and presentations. I would especially like to thank Dr. Kyle Ann Mino and Robert Menger for their assistance with editing, and more importantly, their friendships.

Lastly, I would like to thank my family for their love and never ending faith in me. Even if I didn't believe in myself, my parents always believed in me, and I will always be grateful for that. My husband has also been very supportive of me; this is most apparent in his agreement to relocate from Pennsylvania to Florida so that I could

continue my education. I cannot imagine being here without him next to me. I look forward to starting the next chapter of our lives together.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

LC/MS OPTIMIZATION AND IMPLEMENTATION FOR METABOLOMIC APPLICATIONS

By

Noelle Marie Elliott

August 2012

Chair: Richard A. Yost
Major: Chemistry

The study of metabolomics aims to detect and identify as many compounds as possible in biofluids or tissues in order to determine the changes occurring within an organism in response to some perturbance. This is commonly performed by liquid chromatography/mass spectrometry (LC/MS) or nuclear magnetic resonance spectroscopy (NMR) followed by statistical analysis of the data. The field of metabolomics has grown drastically over the last decade, including the formation of the Metabolomics Society in 2004 and the start of the *Metabolomics* journal in March of 2005.

This work employed LC/MS to investigate the metabolomics of several different biological studies on biofluids. The metabolomics of piglet plasma and red blood cells were compared over the first seven days of life and plasma was indicated to be a better sampling technique, depending on the application. Another study compared the plasma from human patients with colon cancer to healthy controls and five potential biomarkers were identified including dodecenoylcarnitine, myristoleylcarnitine, myristoylcarnitine, palmitoleylcarnitine, and linoleoylcarnitine. Manatee urine was also examined to look for potential biomarkers in estrous urine that may cause the male manatee to easily find

a female manatee in estrus. In manatee urine three potential biomarkers were identified as creatine, glutamic acid, and cytidine. LC with MS/MS instrumentation was used in conjunction with accurate mass to identify metabolites determined to be significant in the colon cancer and manatee urine projects.

In addition to the metabolomics applications, some LC/MS optimization was performed. The main goal of the optimization was to increase the number of detectable molecular features, while maintaining an efficient chromatographic run time. Particularly, the positive mode electrospray settings were optimized and the LC run time was shortened.

CHAPTER 1
INTRODUCTION

Metabolomics

Background

A biological marker, or biomarker, is defined as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to administered therapeutics that are objectively measured and evaluated.[1] Biomarker discovery has changed over the last decade due to the introduction of high-throughput technologies.[2] The metabolomics workflow has become much more automated in sampling, analysis, and data processing. Despite this change, the overall purpose remains to identify compounds directly associated with diseases, or compounds that are responses to different stimuli. Biomarkers may offer a compound for 'screening' or give new insight into various diseases. Furthermore, biomarkers may not appear to be directly associated with the disease; thus, new biological pathways may need to be considered.[3] There are many categories of biomarkers, as well as identification techniques (Table 1-1). The biomarker category that will be focused on in this research is metabolomics, as it is the most predictive of phenotype.[4]

Metabolomics is a relatively new field of science, which has grown significantly over the last 10 years (Figure 1-1) and only recently formed a Metabolomics Society (2004) and a *Metabolomics* journal (2005). Metabolomic studies evaluate and compare changes in biological systems due to disease, diet, or other conditions to an unchanged system (i.e., the control).[5] Metabolomics is one of several omic approaches but is the most predictive of phenotype.[6] Typically, small endogenous compounds (<1000 Daltons) present in tissues or biofluids are monitored for change. For example, a recent

study determined metabolite changes between healthy individuals and patients with either methylmalonic acidemia (MMA) or propionic acidemia (PA), two inborn errors in metabolism.[7] Precipitated plasma was used in this example, but other biological fluids can be used, such as urine, saliva, tissue biopsies, and cerebral spinal fluid (CSF).[3] After choosing a biological sample, an analytical technique must also be chosen for analyses. There are various techniques used to study metabolomics, but two common techniques are mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR).

Approaches

Two general approaches are used in metabolomics: profiling and fingerprinting. Metabolic profiling is a targeted analysis of a limited set of metabolites, whereas fingerprinting is an unbiased global approach.[4] Global approaches in metabolomics are truly 'omic'; the general approach is discussed below. A global fingerprinting approach begins with sample collection and preparation. Sample preparation is important for the removal of proteins, as will be discussed later in this chapter. Next, an analytical technique is implemented (e.g., high-performance liquid chromatography (HPLC or LC) coupled to MS) in order to separate and detect the large number of compounds present within biological samples. LC/MS generates a large amount of data, which can then be extracted and exported for statistical multivariate data analyses. The statistical analyses are utilized to determine significant mass-to-charge ratios (*m/z*) from the raw data set. Once significant *m/z*'s are determined, database searches and other complementary techniques, such as MS/MS or NMR, can also be employed to confirm the identity of the compounds.[4]

Platforms

MS and NMR offer complementary information in metabolomics; MS is used to determine accurate masses and NMR provides structural information based on hydrogen or carbon locations. Each platform has its strengths and weaknesses, but MS has the advantage of being highly sensitive.[8] In contrast, NMR performed as an initial acquisition tool risks potentially significant, low-concentration metabolites going unmeasured.[9] Therefore, it is often beneficial to begin a metabolomics workflow with MS, since MS is capable of detecting compounds in a larger dynamic range than NMR. Following MS analysis, NMR can be used to analyze concentrated fractions collected within the LC run to assist in compound identification.[10-11]

Both platforms generate large datasets that can be analyzed with multivariate pattern recognition techniques to determine potential biomarkers. Similar to LC being used to decrease the complexity of the mass spectra, LC fractions can be collected to simplify the data collected by NMR.[12] Bioinformatics technology continues to grow in the field of metabolomics, improving the software available to handle the large datasets generated in MS or NMR based metabolomic experiments.

Biological Samples

Blood and urine are common biofluids used for metabolomic experiments, especially since collecting CSF is an invasive procedure.[13-15] Although sampling urine is less labor-intensive, blood contains metabolites that travel from organ to organ; therefore, blood provides information on fluxes caused by diseases or other perturbations that urine cannot. Metabolites are maintained in the blood until their concentrations exceed the relevant renal threshold (at which point they become present in urine), so blood is a more favorable sampling medium than urine.[16]

Due to the complexity of whole blood, plasma is often used in metabolomics. Plasma, composed of mostly water that contains metabolites and proteins, requires less sample preparation than blood components containing cells, and is less complex than using whole blood. For example, erythrocytes, or red blood cells (RBC), require lysing in order to empty their contents.[17] For these reasons, plasma was used in most of our experiments.

Plasma is separated from whole blood by centrifugation. The separation, depicted in Figure 1-2, causes plasma to separate to the top, followed by a buffy coat, and the red blood cells on the bottom. The buffy coat contains platelets and white blood cells and is also called platelet rich plasma (PRP). Proteins are commonly precipitated from the plasma by organic solvent, such as acetonitrile, methanol, acetone, or a mixture of similar solvents. After precipitation, the supernatant is removed for direct analysis or is dried down and reconstituted in a solvent of choice.[18] Removing proteins from plasma can also remove smaller molecules; however, it is a necessary step if HPLC is to be performed. This is due to the increased back pressure that can be experienced, in addition to the potential deterioration of the HPLC column due to protein precipitation from organic mobile phases and buffers.[19] Additionally, proteins may interfere with the ion signal of smaller metabolites if present in the samples.

Additionally, plasma should not be confused with serum. While the two biofluids are similar, plasma is prepared by the addition of an anticoagulant, such as heparin or ethylenediaminetetraacetic acid (EDTA). After the addition of the anticoagulant, the blood is centrifuged as described above and the plasma is removed from the top layer. Serum, on the other hand, is blood plasma that has had the clotting factors, such as

fibrinogen-a glycoprotein, removed.[20] Once the serum or plasma is removed from whole blood the precipitation process can be performed. Also, it was recently suggested that plasma and serum are equally useful biofluids for use in a metabolomics study.[21]

High-Performance Liquid Chromatography

LC Theory

To understand liquid chromatography a basic knowledge of analytical separations is needed. In a liquid-liquid extraction a solute will spend a certain amount of time in each solvent or phase. This relationship can be described by the partition coefficient (K), described in Eq. 1-1 below. In the equation A_{s1} and A_{s2} are the activity of the solute in phase 1 and phase 2, respectively.[22] More commonly the concentration of the solute in each phase is used ($[S]_1$ and $[S]_2$). This extraction theory is directly applicable to HPLC separations. Similar to liquid-liquid extractions, LC operates based on the partitioning of a solute between two phases, one phase being the stationary phase and the other a liquid mobile phase. The mobile phase continually flows through the stationary phase, if a solute has a higher affinity for the mobile phase it will spend less time on the stationary phase and elute earlier. Similarly, if a solute has a higher affinity for the stationary phase it will spend more time there and be less affected by the mobile phase causing later elution.[22]

$$K = \frac{A_{s2}}{A_{s1}} \approx [S]_2 / [S]_1 \qquad (1\text{-}1)$$

When performing LC, the data output is in the form of a chromatogram, which is simply the detector signal as a function of time (Figure 1-4). Once the sample is injected, the amount of time it takes the mobile phase (or unretained analyte) to reach

the detector is called the dead time (labeled t_0 in Figure 1-4). The time it takes for an individual retained analyte to reach the detector is its retention time or t_r. In Figure 1-4, two different compounds are shown, A and B, that have two different retention times (t_A and t_B). The chromatographic peaks for compounds A and B have a certain peak width (w) that can be measured at the baseline or at half of the maximum width ($w_{0.5}$).

Once the retention times and peak widths are determined other calculations can be performed relevant to the separating power of the column. In Chapter 3, numbers of theoretical plates (N) and plate heights (H) will be examined for different LC methods to determine the optimal method. The numbers of theoretical plates are often used to determine the chromatographic efficiency of a column. The number of theoretical plates (N), which is related to the retention time and width of a chromatographic peak, can be calculated by using Eq. 1-2. After the number of theoretical plates is determined, the plate height, which is dependent upon column length (L), can also be determined using Eq. 1-3.[23]

$$N=16*(t_r/w)^2 \quad (1\text{-}2)$$

$$H=L/N \quad (1\text{-}3)$$

The efficiency of chromatography methods utilizing the same column can be compared by calculating N for individual analytes. If different columns are used, then H allows for a better comparison as the column length is considered.

Another way to estimate the quality of separation, which is argued to be more versatile than the number of theoretical plates (N), is to calculate the peak capacity.[24] Peak capacity is defined as the maximum number of resolvable peaks.[25] In gradient elution, the peak capacity (n_c) is defined by Eq. 1-4 and Eq. 1-5. Equation 1-4

considers the entire gradient run time (t_g) and Eq. 1-4 removes unused space at the beginning or end of a chromatogram. When determining the peak capacity from Eq. 1-4 and Eq. 1-5, w is the average peak width and t_1 and t_f are the retention times of the first and last eluting peaks, respectively.

$$n_c = 1 + \left(\frac{t_g}{w}\right) \quad (1\text{-}4)$$

$$n_c = (t_1 - t_f)/w \quad (1\text{-}5)$$

In LC, the resolution (R_s) is a measure of how well two analytes are separated. Resolution is defined below in Eq. 1-6, where t_a and t_b are the retention times of two peaks a and b, and w_a and w_b are their respective baseline widths.[23]

$$R_s = 2*(t_b - t_a)/(w_b + w_a) \quad (1\text{-}6)$$

Due to the complex nature of biological samples, LC separations are often performed prior to MS or NMR analyses. LC allows for direct injection of biofluids with minimal sample preparation and is capable of high-throughput analyses.[26] Furthermore, LC columns have a relatively long life span and are routinely used to analyze hundreds of samples daily.[23] Chromatographic methods allow optimization through varying different parameters, such as columns, solvents, and gradients. Although columns are chosen based on the analytes' physical properties, biofluids are composed of compounds that cover a wide range of polarities; thus, all compounds cannot be efficiently retained and separated with a single column. The most common column employed is a reverse-phase (RP) column, which retains more polar compounds poorly, but can separate compounds with low to moderate polarities with reproducibility. RP chromatography uses a non-polar column (stationary phase) and polar solvents (mobile phases) compatible with the biological samples and the mass spectrometer. Solvents

can be used in an isocratic (constant composition of solvents) or gradient elution (composition of solvents varies with time). Gradient elution is more common due to its ability to remove non-polar compounds more rapidly, as well as its self-cleaning nature.

Monolithic and Small Particle HPLC Columns

Early chromatography columns were composed of larger packed particles, leading to longer diffusional path lengths. With time, smaller particles were made that allowed shorter diffusional path lengths at the expense of higher back pressures. Monolithic columns were developed with the intention of having a shorter diffusional path length as well as allowing column permeability.[27] A monolithic column is essentially a one-piece continuous phase of porous silica rods modified with various functional groups, such as C18.[28] Pressures observed with a Phenomenex C18 Onyx monolithic column (Phenomonex, Torrance, CA) are comparable to packed columns with 11 µm particles but have a chromatographic efficiency of packed columns with 3 µm particles.[29]

Small particle columns, as already mentioned, have the advantage of efficient chromatography at the expense of a high back pressure; however, back pressure may be decreased by applying heat, which in turn decrease the solvent viscosity. While the monolithic column is useful, small particle columns are more commonly implemented. In general, typical columns are approximately 100 mm x 2 mm with 3 µm particles.[20] The small particle column used in this work is packed with 1.8 µm particles and is half the length (50 mm) as what is commonly used. By heating the column, ultra high performance LC (UPLC) is mimicked without the ultra high pressure. Figure 1-3 visualizes the stationary phases in both monolithic and small particle columns.

Electrospray Ionization

The electrospray ionization (ESI) interface was developed in John Fenn's lab in 1985 and is now one of the most widely used ionization sources.[30] ESI is performed by passing a solution through a capillary, typically held at a high potential (2-4 kV). The high potential causes the solution to break into highly charged droplets that reduce in size due to evaporation and 'Coulomb explosions'.[31] ESI is performed at atmospheric pressure and the desolvated ions are transferred through ion optics by potential and pressure gradients. The ESI source used in this work has an orthogonal spray (Figure 1-5) with respect to the mass spectrometer capillary entrance. Additionally, the ESI capillary is held at ground and a non-zero potential is instead applied to the MS capillary (Vcap).

Time-of-Flight Mass Spectrometry

Time-of-flight mass spectrometry (TOF-MS) analysis can be explained mathematically by considering Equation 1-7:

$$\frac{mv^2}{2} = zeE \qquad (1\text{-}7)$$

where m is the mass of the ion, v is the velocity of the ion, z is the number of charges, e is the charge of an electron and E is the voltage applied in order to accelerate all the ions to the same kinetic energy.[32] Solving for v shows the inverse square root dependence on mass; thus the larger the mass, the slower the velocity of the ion. The flight time (t) can be determined using the flight tube length (D) (Eq. 1-8) and then the m/z value can be found by Eq. 1-9.

$$t = \frac{m}{2zeE}^{\frac{1}{2}} D \qquad (1\text{-}8)$$

$$\frac{m}{z} = 2eE\left(\frac{t}{D}\right)^2 \quad (1\text{-}9)$$

Commonly, TOF mass spectrometers are equipped with an ion mirror or reflectron. Reflectrons are beneficial because they allow for longer flight tube lengths by simply changing the direction of the ion path within the flight tube. For instance, a one meter flight tube becomes a flight path of two meters by adding a reflectron. Increasing the flight path becomes beneficial to mass resolution, as will be discussed in the next section. How the reflectron operates is relatively straight forward. Ions that arrive at the flight tube begin their journey at the ion pulser. The ion pulser is a stack of plates with a center hole, except for the back plate. A pulse is applied to the back plate of the pulser and the ions are accelerated through the flight tube. At the opposite end of the flight tube the ions reach the reflectron. There are two potential gradients applied to the ion mirror that allow improved second-order time focusing. The focusing is a result of removing excess kinetic energy both horizontally and vertically by "turning around" the ions' trajectory. The initial horizontal momentum of the ions allows the reflection path to reach the detector instead of directly back to the pulser.[33]

Resolution and Mass Accuracy

Mass spectrometers are often evaluated based on two figures of merit: resolution and mass accuracy. Resolution defines how well two mass spectral peaks are separated and mass accuracy defines how far from the true or theoretical mass an observed mass is. The equations for resolution (R) and mass accuracy (A) are shown below (Eq. 1-10 and Eq. 1-11): m_2 and m_1 are two adjacent masses and m_m and m_t are the measured mass and the theoretical mass. Although R is unitless, A is measured in parts per million (ppm). The flight tube length of the TOF-MS used in this work is

approximately two meters. This is a relatively large flight tube length, yielding improved resolution, as resolution is proportional to flight time, and the ions will maintain a constant time interval (Δt) between them (Eq. 1-12). Therefore, the resolution increases with the flight time, which is proportional to the flight tube length (D), (Eq. 1-8).[32]

$$R = \frac{m_1}{m_2 - m_1} \tag{1-10}$$

$$A = \frac{m_m - m_t}{m_t} * 1x10^6 \tag{1-11}$$

$$\frac{1}{R} = \frac{\Delta m}{m} = \frac{2\Delta t}{t} \tag{1-12}$$

TOF-MS Data Acquisition

When using TOF-MS, a 'transient' describes the spectral acquisition. A transient begins when the ions are pulsed through the flight tube and ends when all ions are detected. For the range of *m/z* 100-3200 (flight time of ~100 μs for the largest mass), a transient length is ~100,000 ns. The detector samples every nanosecond, thus one transient will have 100,000 sample intervals. About 10,000 transients are summed together for one spectrum, so at this mass range the sampling rate is one spectrum per second.[34] The mass range that was used in these metabolomic studies was *m/z* 112-1700. The transient length was ~72,640 ns and 9,652 transients were summed; therefore, the sampling rate was 1.41 spectra/s and one spectrum was generated every 0.7 s.

TOF-MS System

The TOF-MS system operated in this research is characterized by relatively high mass accuracy (typically less than 5 ppm) and resolution (5000-13000 over *m/z* 100-3200). Since thousands of compounds can be present in a single biological sample,

high resolving power and accurate mass are desirable to reduce the number of possible candidates for compounds of interest.

Most MS experiments in this work were performed on an Agilent 6210 orthogonal acceleration Time-of-Flight (oa-TOF) mass spectrometer (Agilent Technologies, Santa Clara, CA). As shown in Figure 1-6, ions generated by ESI travel through a capillary to the skimmer. Lighter ions and drying gas are then pumped away in vacuum stage 1. Ions that proceed through the skimmer are guided into a DC quadrupole, which shapes the beam of ions into a parallel packet. The ion pulser applies a DC bias that causes a packet of ions to accelerate towards the ion mirror with equal kinetic energies. At the ion mirror, the direction of the ions is reversed and the ions move toward the detector.

Data Analysis

Data analysis (extraction, alignment and statistics) can be performed with various software packages, some of which are freely available online. [35] One of the most popular methods implements XCMS (The Scripps Research Institute, La Jolla, California) for data processing in combination with SIMCA P (Umetrics).[36-38] XCMS is an alignment package that can be downloaded from The Scripps Research Institute's website and SIMCA P is a statistical software package.[39] While this combination is widely utilized, there are a few disadvantages. SIMCA P is rather expensive and XCMS requires extensive knowledge and understanding of the programming language R, making it more difficult to implement than other software provided by instrument companies. An online version of XCMS was recently launched (summer 2011), requiring no prior knowledge of R. The online program allows files to be uploaded and analyzed by an appropriate generic R script of your choosing. For example, ultra-high performance liquid chromatography (UPLC) or HPLC separations can be chosen in

conjunction with a quadrupole-TOF (Q-TOF), or orbitrap mass analyzer. Once the online process is started, the program proceeds to align all chromatograms and determine *p*-values for each feature. Box-and-whisker plots are created, as well as extracted ion chromatograms for each feature. When all the processing is finished an email is sent to the user. By logging into the Scripps website the results of the data set can be visualized. Unfortunately, the online version of XCMS does not perform any statistics other than the common Student's *t*-test.

In addition to the packages mentioned above, most mass spectrometer manufacturers have a program available for metabolomics data processing/analysis. For example, AB SCIEX has created Marker View and Waters uses MassLynx.[40] Agilent Technologies created MassHunter software that performs pretreatment (extraction of all *m/z* and retention time pairs over a certain abundance threshold) and an additional software package called Mass Profiler Professional (MPP), capable of multivariate statistical analysis. Some statistics MPP can perform include: Student's *t*-test, analysis of variance (ANOVA), principal component analysis (PCA), and partial least squares-discriminant analysis (PLS-DA). MPP was used frequently throughout this work and Qualitative Analysis was used for all pretreatment (part of Agilent MassHunter) of the LC/MS data collected. After extraction of the molecular features through Qualitative Analysis, files in the form of .CEF (formatted for MPP use) and .CSV (Excel format) can be exported.

Principal Component Analysis

PCA is an unsupervised statistical method, which reduces the dimensionality of a sample set by identifying the variance in data and projecting it into a lower-dimensionality.[41] In general, the results of a study can be considered as a matrix, and

the covariance matrix of the initial matrix can be calculated. After the covariance matrix is determined, the Eigenvectors and Eigenvalues of the matrix are calculated. The Eigenvalues are organized from largest to smallest and the first two or three corresponding Eigenvectors are plotted (an Eigenvector is termed a principal component).[42] If two components are plotted the outcome will be two-dimensional (2D), whearas if three components are plotted it will be three-dimensional (3D).

PCA plots are commonly used to determine if there are any obvious differences occurring within a sample set. For example, if a metabolomics study is composed of three groups, a control group and two different treated groups, one would expect to see the PCA plot resulting in three individual groups, such as that shown in Figure 1-7. In addition to being used as a data visualization tool in metabolomics, PCA can also be used to determine the features/masses (or loadings) that cause those differences observed in the PCA plot (or scores plot). Although PCA is popular within the field of metabolomics it is increasingly being replaced by partial least squares-discriminant analysis.

Partial Least Squares-Discriminant Analysis

Partial least squares-discriminant analysis is a statistical technique that has recently gained more attention in the field of metabolomics. In an ISI Web of Science database search of the terms “metabolomics” and “PLS-DA” for the years 2005-2008 only 34 results were found, while a search of those same terms for the years 2008-2011 yielded 142 results. This increase in metabolomics publications utilizing PLS-DA shows its usefulness as a discrimination and dimension reduction tool.

PLS-DA differs from PCA in the fact that it is a supervised method. This means it is directed by a response matrix Y, which is then used to derive components from the

data set matrix X that best describe the Y matrix.[43] PLS-DA and other supervised methods are commonly used for complex metabolomics studies, where important changes between study groups may be masked by other large changes (e.g. human urine can vary greatly between individuals, which may hinder elucidation of more subtle changes).[43] Additionally, PLS-DA can be used as a prediction model. Since the method is supervised, the conditions given can be used to determine how individual samples would be grouped.

While PCA is commonly used solely to determine how samples group, PLS-DA is often used to determine the masses/features that are able to describe the grouping. As PLS-DA is a supervised method, analysis of the loadings plot from PLS-DA is more straightforward than the loadings plot from PCA. Thus, PLS-DA can be used to model the sample separation as well as to easily determine significant influences on that separation.

Student's *t*-test

A Student's *t*-test is used when only two experimental groups exist (e.g. diseased and control). Significant masses are determined by comparing the means of the two groups for each individual feature. The *t* statistic is calculated by applying Eq. 1-13, where $\bar{x}$ is the average for the two groups (denoted by the subscript), *s* is the pooled standard deviation (Eq. 1-14), s_n is the standard deviation of a sample grouping and *n* is the number of samples in either group one or group two, depending on the subscript.[44]

$$t = \frac{(\overline{x}_1 - \overline{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{1-13}$$

$$s^2 = \frac{n_1 - 1\ s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \tag{1-14}$$

The *t* statistic is compared to a critical value to determine if a significant difference exists between sample groupings. If the *t* statistic exceeds the specified critical value within a certain probability (*p*-value), the null hypothesis (that the two averages are equal) can be rejected.[44] Once significant masses are determined by the *t*-test more analyses can be performed to either filter the list further, or identify those significant masses.

Analysis of Variance

The analysis of variance (ANOVA) is commonly used in metabolomics when there are more than two groups, making a traditional Student's *t*-test inefficient. ANOVA uses the sum of squares (SS), which is the variance after mean-centering, as well as *F*-tests to calculate *p*-values and determine significant features. The sum of squares is shown below in Eq. 1-15, where $\bar{x}$ is the average of a set of values and x_i is an individual value.[41]

$$SS = \sum_{i=1}^{n} (x_i - x)^2 \quad (1\text{-}15)$$

The general format for ANOVA is illustrated in Table 1-2. The among-group variation (SS_A) is calculated by determining the *SS* of the means multiplied by the number of samples for each group. The within-group variation (SS_W) is calculated using the *SS* for each group. Once the *SS* is determined for among and within the groups, the degrees of freedom are used to calculate the mean squares and the *F*-test value (Table 1-2).[41] The *p*-value can then be determined for each metabolite using Fisher's *F*-density plots (some of which are freely available online).

Benjamini and Hochberg False Discovery Rate

It should be noted that *p*-values can be calculated and corrected by various methods. Most commonly, *p*-values are asymptotically calculated (as opposed to permutative calculation) and a multiple testing correction is used in metabolomics. The testing correction that will be used in this work is the Benjamini and Hochberg false discovery rate (FDR). The goal of multiple testing corrections such as FDR is to prevent type I errors from occurring (the rejection of a true null hypothesis, more commonly known as a false positive). Unfortunately, controlling for the type I errors often leads to an increase in type II errors (the failure to accept a false null hypothesis, a false negative).[44] The relationship of type I and type II errors is often referred to as the familywise error rate (FWER). The possible outcomes of individual tests can be seen in Table 1-3. The procedure for the Benjamini and Hochberg FDR is as follows: the *p*-values (P) for m tests are ranked in ascending order ($P_1 \leq P_2 ... \leq P_m$) and each *p*-value, beginning with P_1, is checked to meet the requirement $P_i \leq i\alpha/m$, where i is the rank. The term α/m is the type I error rate. Eventually k, the largest i where $P_i \leq i\alpha/m$ is true, is determined and all the corresponding null hypotheses ($H_{(1)}$ to $H_{(k)}$) are rejected.[45] The ability of the Benjamini and Hochberg FDR to potentially reduce type II errors compared to other methods is shown in Figure 1-8. Because of the desire to have not even one type I error, FWER correction methods require a lower threshold setting (thus increasing the likelihood of type II errors) than the FDR method (Figure 1-8).

Table 1-1. Biomarker categories and examples of methods used within each category.[2-3]

Biomarker Category	Methods
Genomics	DNA arrays
DNA-Based- Copy number/loss of heterozygosity, sequence and epigenetic variations, and genome rearrangements	Sequencing
RNA-Based- mRNA and miRNA signatures	
Proteomics	Mass spectrometry
Proteins	Liquid chromatography
Peptides	Protein arrays
Metabolomics	Mass spectrometry
Metabolites	Liquid chromatography
Lipids	Nuclear magnetic resonance
Carbohydrates	

Table 1-2. Typical set-up for ANOVA.

Source of Variation	Sum	Degrees of Freedom	Mean Squares	F-Test
Among sample groups	SS_A	m-1	s_A^2	s_A^2/s_W^2
Within sample replicates	SS_W	N-m	s_W^2	
Total Variation	SS_T	N-1	s_T^2	

SS is the sum of squares, m is the number of groups, N is the number of samples, and s is the average (sum/degrees of freedom) for each source of variation.

Table 1-3. Possible outcomes of individual tests.[46]

Truth	Decision		Total
	Not Significant	Significant	
Null Hypothesis	U	V	m_0
Alternative Hypothesis	T	S	m-m_0
Total	m-r	r	m

V is the number of type I errors, T is the number of type II errors and the variables m, r, and m-r are observed. U, T, and S are unknown.

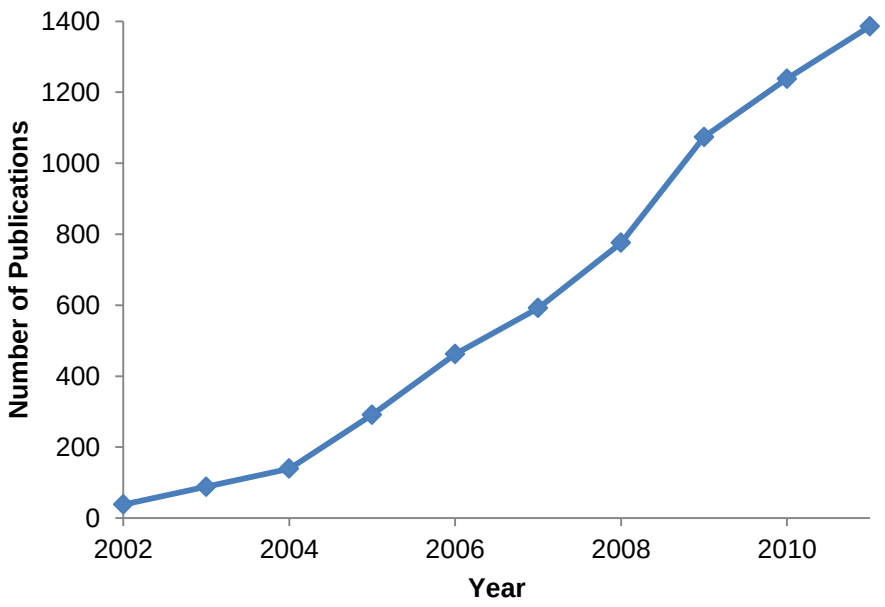


Figure 1-1. Depiction of the number of publications in metabolomics each year, for the past 10 years (database search of "metabolomics" in ISI Web of Science).

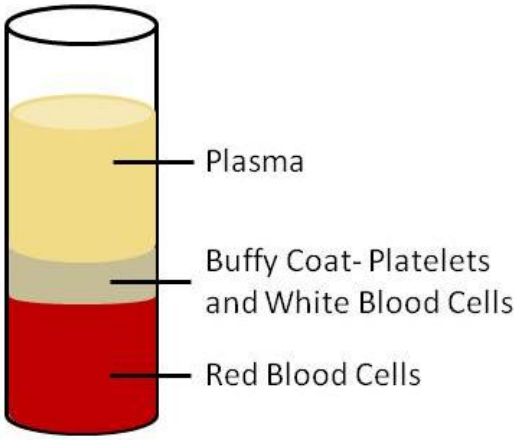


Figure 1-2. The separation of whole blood into its three components.[47]

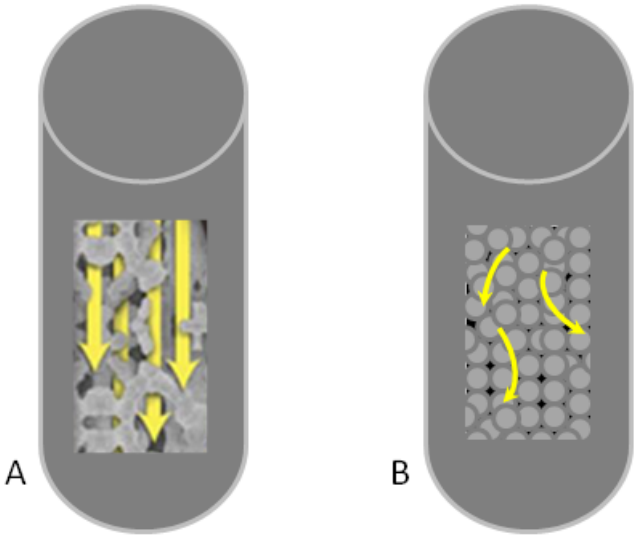


Figure 1-3. Comparison of the diffusional path of monolithic based columns (A) and small particle columns (B).[29]

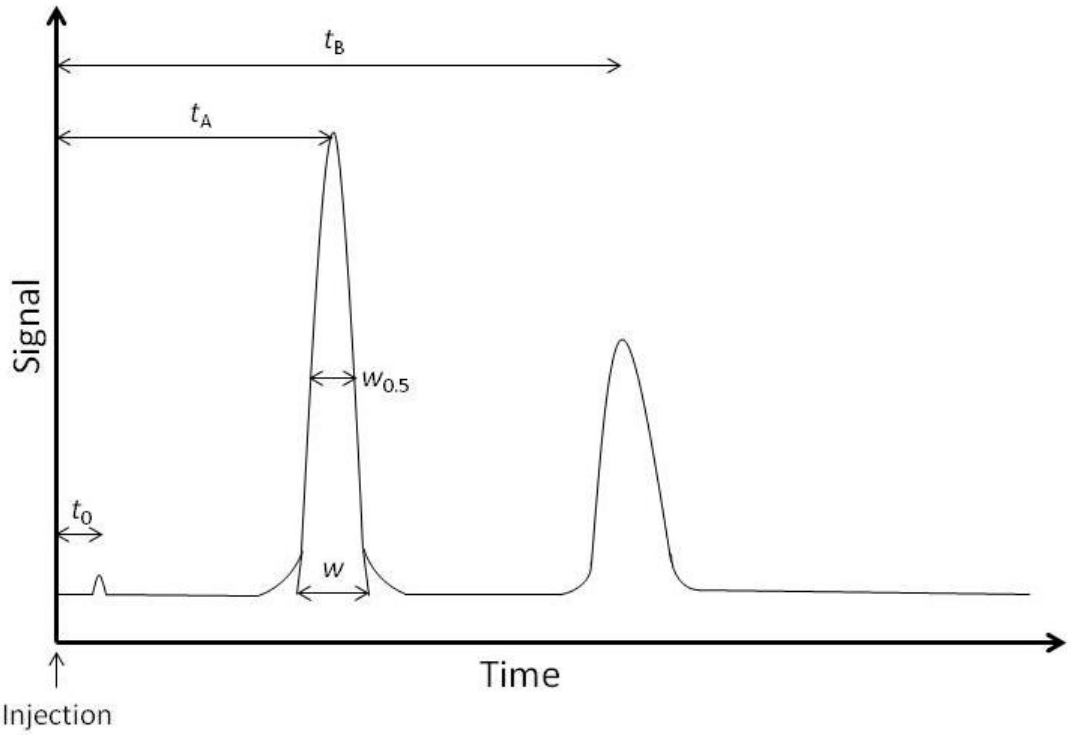


Figure 1-4. Depiction of dead time, retention time and peak widths in LC.[48]

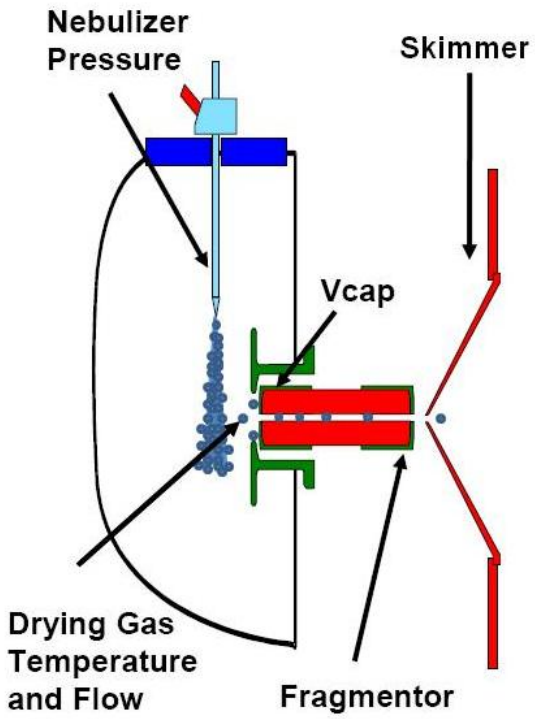


Figure 1-5. Schematic of the electrospray source. The LC flow enters the source through the nebulizer, where it is sprayed orthogonally with respect to the mass spectrometers capillary inlet (Vcap).[49]

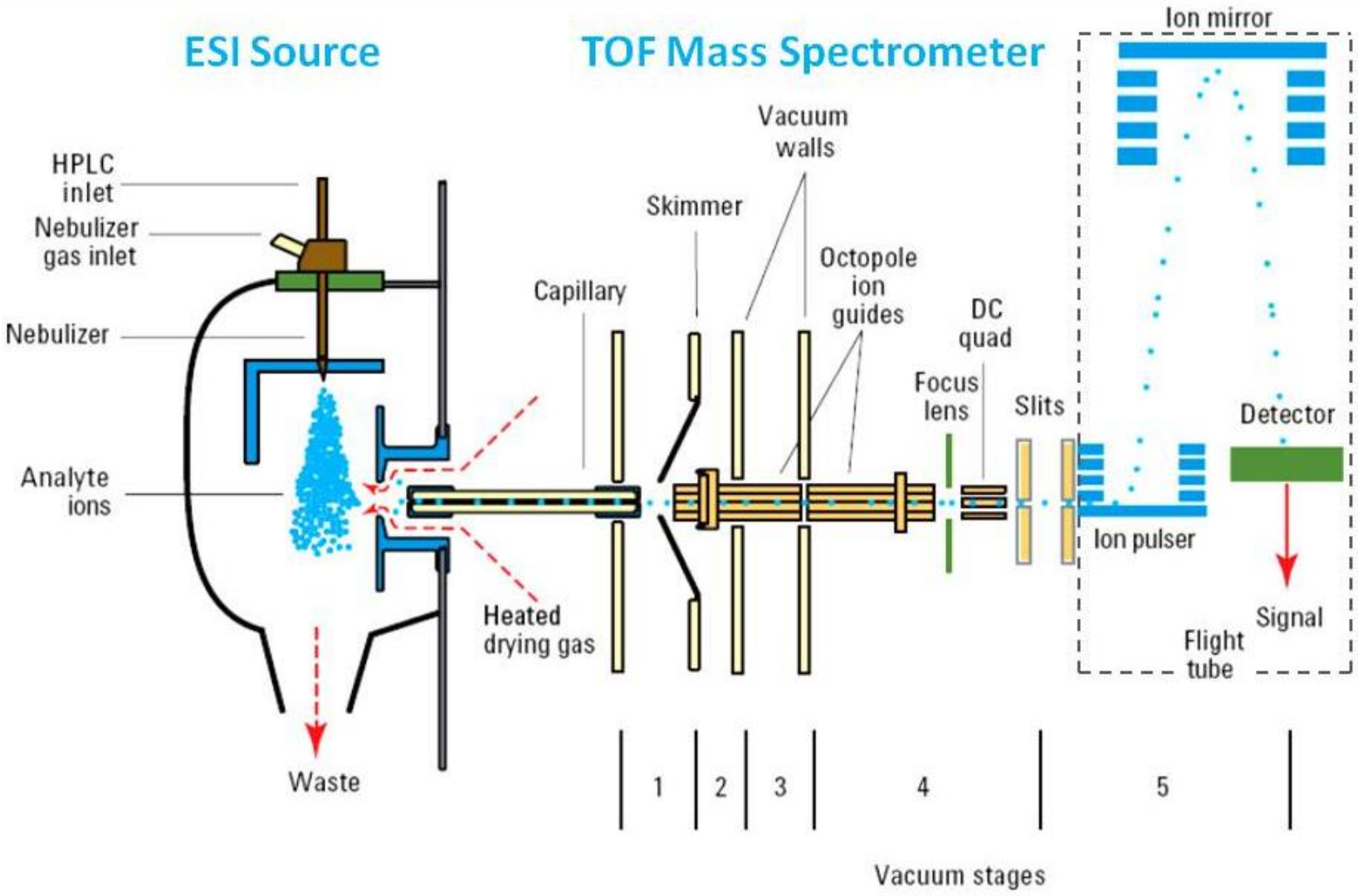


Figure 1-6. Schematic of the Agilent 6210 time-of-flight mass spectrometer.[50]

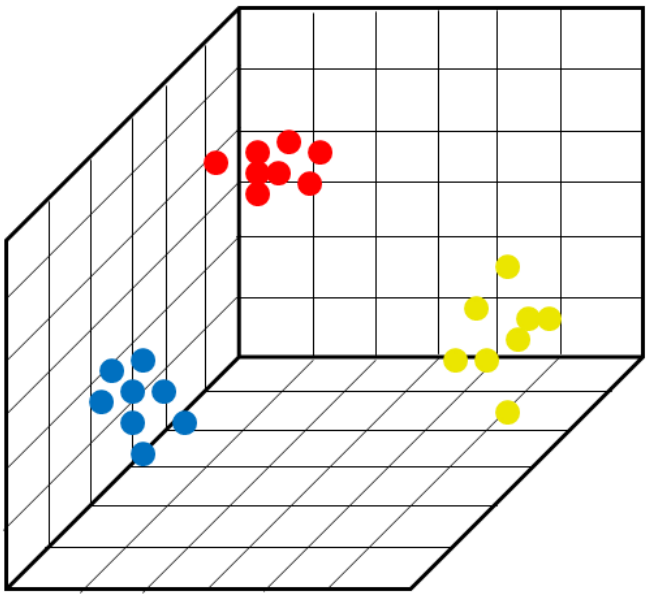
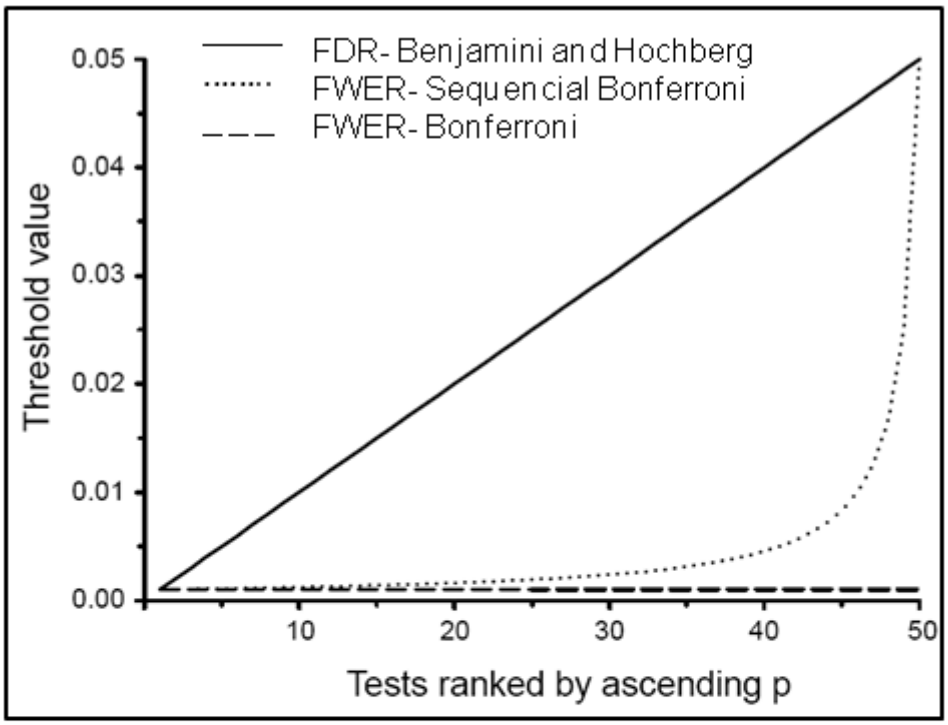


Figure 1-7. Generic 3D-PCA plot of a metabolomics experiment with three groups (indicated by the colors red, blue and yellow).[51]



Rank	Bonferroni	Sequencial Bonferroni	Benjamini and Hochberg
1	α/m	α/m	α/m
2	α/m	$\alpha/(m-1)$	$2\alpha/m$
3	α/m	$\alpha/(m-2)$	$3\alpha/m$
i	α/m	$\alpha/(m-i)$	$i\alpha/m$
m	α/m	α	α

Figure 1-8. Threshold *p*-values for three different multiple testing corrections. Conditions include 50 tests being performed and FWER = FDR = 0.05 (that is, m = 50 and α = 0.05). The Benjamini and Hochberg FDR tests from the largest to smallest *p*-value while the two Bonferroni methods test from the smallest to the largest *p*-value.[45]

CHAPTER 2
COMPARISON OF THE METABOLOMICS OF PLASMA AND RED BLOOD CELLS FROM PIGLETS IN THE FIRST SEVEN DAYS OF LIFE

Background

Piglets in the first seven days of life are excellent animal models for neonatal disorders.[52] While much is already known about neonatal disorders, additional knowledge could only improve our methods of testing for these diseases.[53] For instance, octanoylcarnitine (C8) is currently used as a biomarker to screen newborns for medium-chain acyl-CoA dehydrogenase deficiency (MCAD).[54] One problem with this method is the lack of specificity for MCAD, since C8 can also be elevated in other disorders. Global metabolomic profiling and targeted studies of carnitines and acylcarnitines in piglets are hypothesized to indicate how levels of metabolites change throughout the first week of life in humans; this could allow further insight into neonatal disorders. For these analyses, whole blood is separated into plasma, red blood cells, and a buffy layer, but typically only plasma is analyzed. In this work, we extend the study to both plasma and red blood cells (RBC).

It is common in metabolomic studies to utilize biofluids such as plasma or serum (as discussed in Chapter 1). In fact, one paper in the field discusses whether plasma or serum is a better choice for intersubject comparisons. The conclusion to this paper is that neither biofluid is better than the other; there are very little variations between plasma and serum allowing them to be compared relatively easily.[21] There is no comparison in the literature of RBCs to plasma. Therefore, this work is beneficial to determining if RBCs are a valid way of looking at the metabolome of piglets.

As mentioned, carnitines and acylcarnitines are of particular interest in this study due to their importance in early human development. This is because acylcarnitines are

the catabolic end products of fatty acids and some branched chain amino acids used to generate energy. Acylcarnitines are derived from the exchange of acyl groups between their corresponding acyl coenzyme A and L-carnitine. However, acylcarnitines are capable of crossing mitochondrial and cell membranes, while their corresponding acyl coenzyme A cannot. Additionally, the acylcarnitine pattern in a fasting individual, under normal circumstances, is relatively stable. Because of this, the plasma acylcarnitine profile can be used to detect many metabolic disorders. In fact, all babies born in the U.S., and most of Europe, are screened by a dried blood spot analysis for their acylcarnitine profile.[55] For these important reasons, some carnitines and acylcarnitines will be targeted in the piglet plasma and red blood cells in order to observe changes occurring over the first seven days of life.[55]

Experimental

Sample Collection and Preparation

Plasma and red blood cell samples were prepared from whole blood of piglets in the first seven days of life. The whole blood was separated into its components by centrifugation and the plasma and RBC were frozen. The frozen RBC and plasma were then prepared by two different procedures.

Frozen plasma was thawed and 100 µL portions were mixed with 1 mL of acetonitrile:methanol (3:1) and frozen overnight. After the overnight freeze, the samples were thawed, centrifuged, and the supernatants removed. The supernatants were then centrifuged two more times and evaporated to dryness under nitrogen.

Frozen RBCs were prepared by adding equal volumes of ice-cold de-ionized water to each sample. The RBC samples were then re-frozen overnight and later centrifuged. The supernatants were removed and the samples dried under nitrogen.

The dried products from both plasma and RBC samples were each dissolved in 100 µL of 40% acetonitrile, 40% methanol, 10% water and 10% isopropanol.[56] Three piglets per day of life were sampled and separated into two vials per piglet and run in duplicate for plasma analyses (12 runs per day, 84 files in total). For RBC analyses, three piglets were sampled per day and run in duplicate (6 runs per day, 42 files in total).

Liquid Chromatography Conditions

All chromatograms were acquired using an Agilent 1200 Series LC system (Agilent Technologies, Inc., Santa Clara, California). Two, in-series, Phenomenex C_{18} Onyx monolithic columns (Phenomenex, Torrance, CA), with individual dimensions of 4.6 x 100 mm, were used. Water was used for solvent A and acetonitrile for solvent B, both solvents contained of 1% (v/v) acetic acid. The gradient performed began with 95% A and 5% B and held for 6.5 minutes, followed by a ramp to 100% B at 25.5 minutes. Solvent B was held at 100% for 9.5 minutes and then returned to 95% A and 5% B over 10.0 minutes (Table 2-1). The column was then re-equilibrated with 95% A and 5% B for 5.0 minutes. The total gradient time was 50.0 minutes, with only the first 37.0 minutes being used for data analysis. The flow rate was set to 0.8 mL/min and an injection volume of 15.0 µL was used for all RBC and plasma samples.

Electrospray Source

Ionization of the LC flow was performed by an orthogonal dual-spray electrospray source operating in positive mode (Agilent Technologies, Inc., Santa Clara, California). The dual-spray is used to constantly deliver calibrant ions into the system. The dual-spray ESI source allows for easy post-analysis recalibration, and when used properly, no recalibration is necessary since a real-time calibration will be performed. A gas

temperature of 300˚C and a drying gas flow of 10 L/min was used. The nebulizer pressure was set to 50 psi.

Mass Spectrometer

Mass spectra were acquired using an Agilent 6210 Time-of-Flight mass spectrometer (Agilent Technologies, Inc., Santa Clara, California). The capillary, fragmentor, and skimmer voltage were set as 4000, 180, and 60 V, respectively. The instrument was operated in extended dynamic range (2 GHz), as opposed to high resolution mode (4 GHz) in order to increase the linear dynamic range and decrease saturation of the detector.

Data Analysis

All files were recalibrated in Agilent's Qualitative Analysis software (Agilent Technologies, Inc., Santa Clara, California) and molecular features were extracted. The abundance threshold for the feature extraction was set to 2,500 arbitrary units (a.u.). Only singly charged species were considered. To be considered a molecular feature at least two ions had to be present (this could be two ions of the isotope pattern). After feature extraction the files were exported in .CEF and .CSV formats. The .CEF files were used in Agilent's Mass Profiler Professional (MPP) software for statistical analysis (Agilent Technologies, Inc., Santa Clara, California). In MPP, features over the first 37 minutes were considered. Mass and retention time alignment was performed on features with thresholds 5 ppm and 0.5 minutes, respectively. Normalization of each file was done by baselining to the median.

Additionally, Quantitative Analysis software (Agilent Technologies, Inc., Santa Clara, California) was also used to extract 11 carnitines/acylcanitines (carnitine, acetylcarnitine, propionylcarnitine, butyrylcarnitine, hexanoylcarnitine, octanoylcarnitine,

decanoylcarnitine, lauroylcarnitine, myristoylcarnitine, palmitoylcarnitine, and stearoylcarnitine). This was done by entering the known retention times and accurate masses into the software so that a spreadsheet could be automatically generated, containing important information for those compounds. This information included, but was not limited to, retention time, chromatographic peak area and height, and peak width. From this spreadsheet the relative behavior of the different carnitines/acylcarnitines could be monitored over the seven days of life.

Results and Discussion

As can be seen by comparing the chromatograms of the piglet plasma and RBC samples (Figures 2-1 and 2-2), the plasma chromatograms are much more complex. This may mean there is more information to be found by analyzing plasma than RBCs. After visual comparison of the chromatograms, the extracted features from the plasma and RBC samples were considered.

Piglet Plasma

In the piglet plasma, 15,989 masses were determined to be present within the first seven days of life. These 15,989 masses were filtered on frequency of observation (80%) in order to retain the masses more frequently observed, as well as to decrease the number of compounds to a more manageable number. By doing this, an individual feature had to be present in 80% of the samples tested in an individual day, but not necessarily every day. For example, a feature would pass the filter if it was present in 80% of the samples from day 2 and day 5 but not in days 3, 4, 6, 7, or 8. A total of 568 masses passed this frequency cut-off. ANOVA was then performed to determine which of those 568 features were statistically significant. Using a *p*-value of 0.01, a post hoc test (Tukey HSD), and a multiple testing correction (Benjamini-Hochberg FDR), 82

masses were determined to be significant. Table 2-2 depicts the statistically significant masses between days. This number of features is not overwhelming and may make it easier to find identifications. The 3D-PCA plot is shown in Figure 2-3 for the 82 ANOVA masses. The x-axis (component 1) accounts for 25.15% of the variation, while the y- and z-axes (components 2 and 3) account for 11.93% and 7.53%, respectively. There does appear to be some separation in Figure 2-3. In particular it seems that the earlier days of life are on the top right side of the plot and the later days of life drift towards the bottom left. A two-dimensional plot is also shown in Figure 2-4, as some prefer this view.

Previous work in our lab, with only piglet plasma, found day 2 and days 7 and 8 grouping away from the middle days, with days 7 and 8 grouping together on the PCA plot.[56] These results are similar to what is shown here but with less distinct separation. Also, the data analysis process is not the same for this work and the previous work.

A PLS-DA was also performed and the plot can be seen in Figure 2-5 (3D view) and Figure 2-6 (2D view). The prediction model had an accuracy of 80.95%, most of the inaccuracies stemmed from day 6 as well as some other days of life in the middle of the time period tested (Table 2-3). The day of life that was most accurate (100%) was day 2. This means that significant changes are occurring from day two compared to the later days. The loadings plot from the PLS-DA (Figure 2-7) is often used to determine what masses are influencing the separation observed. This can be done by examining the masses that are farthest from WC0 = 0. These masses (30) will be discussed in more detail later.

Piglet Red Blood Cells

Piglet RBC samples were put through the same data analysis procedure as described above for piglet plasma. A total of 8,739 masses were determined to be present in the red blood cells of piglets during the first seven days of life. After filtering on the frequency of observation (80%, as described above) 879 masses passed the filter. The filtered masses were then subjected to ANOVA (*p*-value = 0.01) with a post hoc test (Tukey HSD) and a multiple testing correction (Benjamini-Hochberg FDR) to determine the statistically significant masses. Only 23 features were determined to be significant from day to day within the RBC samples (Table 2-4). A 3D and 2D-PCA plot of the 23 ANOVA masses is shown in Figure 2-8 and Figure 2-9, respectively. The x-axis (component 1) accounts for 37.87% of the variation, while the y- and z-axes (components 2 and 3) account for 14.16% and 7.95%, respectively. As can be seen in the PCA plots, the most separation is observed for days 7 and 8 of life.

The 3D-PLS-DA plot is shown in Figure 2-10, as well as the 2D-PLS-DA plot in Figure 2-11. The piglet RBC prediction model had the same accuracy as the plasma model, 80.95% (Table 2-5). Days 2, 7, and 8 had 100% prediction accuracy while days 4 and 6 were lacking in accuracy. It is important to stress that there are half as many samples in the RBC model as in the plasma model; therefore, the RBC model may be lacking the appropriate statistical power. The loadings plot (Figure 2-12) was used to examine the masses that were farthest from WC0 = 0. These masses (13) will be discussed below.

Database Search of Significant Piglet Plasma and RBC Masses

As already discussed, PLS-DA is a powerful and popular model. The loadings plot allows easy determination of masses that are influencing the observed separation. The

ANOVA masses were further filtered by considering masses in the PLS-DA plot that satisfied the conditions of WC0 < -0.1 or WCO > 0.1, for both the plasma and RBC data sets. The piglet plasma dataset resulted in 30 masses and the piglet RBC dataset resulted in 13 masses, which were then applied to an accurate mass database search. This was performed using METLIN.[57- 58] The mass of the positively charged ion was used in the database search and a mass accuracy of 10 ppm was applied. Two different ion options were searched, $[M+H]^+$ and $[M+Na]^+$. The positive ion that is outputted as significant in MPP is supposed to be an $[M+H]^+$ ion, however, the $[M+Na]^+$ ion was still searched as a precaution. Piglet plasma masses yielded a formula for 3 ions as $[M+Na]^+$ ions. Piglet RBC masses yielded only one formula as an $[M+H]^+$ ion.

Targeted Analysis

In this study, 11 carnitines/acylcarnitines were studied further to determine their relative change over the first seven days of life. This was done by determining the average abundance within the two replicates of three piglets (n = 6) for the RBC dataset, and for two replicates of thee piglets with two vials each (n = 12) for the plasma dataset. Of the 11 compounds all were found to be present in the plasma samples, but only 7 were present in the RBC samples. Those seven compounds were carnitine, octanoylcarnitine, decanoylcarnitine, lauroylcarnitine, myristoylcarnitine, palmitorylcarnitine and stearoylcarnitine. The fact that only 7 of the targeted compounds were found in the RBC samples (compared to the 11 found in plasma) further supports the appropriateness of plasma for an experiment of this type. Bar graphs of the average abundance of the 7 compounds found in both plasma and RBC are shown in Figures 2-13 to 2-19. In Figure 2-13 the abundance of carnitine is observed to be lowest on day 2 in both plasma and RBC samples. The abundance of

carnitine over the rest of the days overlapped when the standard deviation of the mean was considered, but showed a slight increase from day 2 thru day 8. This same trend is apparent in the other 6 compounds, with the exception of some of the RBC samples. The RBC trends from day 2 to day 8 seem to randomly go up and down, although day 2 was consistently the lowest abundance day for all 7 compounds except octanoylcarnitine (Figure 2-14).

There may be a few reasons the plasma and RBC samples do no show the same trends over the 7 days. One reason may be the extraction efficiency. If all the contents of the red blood cells are not efficiently emptied than there will be different results when compared to plasma. Since some of the lower molecular weight carnitines and aceylcarnitnes are missing from the RBC samples, this may indicate a lack of extraction of these from the red blood cells. Also, it may simply be that certain carnitines and acylcarnitines are present in the red blood cells at lower concentrations than in plasma.

Summary

Piglet plasma yielded 82 significant masses, while piglet red blood cells only resulted in 23 significant masses. This shows that there are more changes to compare in plasma than in RBCs of piglets. Only two statistically significant masses matched between the ANOVA results of the piglet plasma and RBC samples. Again, this may be the result of poor compound extraction from the RBC.

It would be beneficial to reproduce this experiment with the correct number of replicates (at least three) as well as an equal number of samples between the plasma and RBC groups. Unfortunately, at the time of analysis, the chromatographic method was still 50 minutes, requiring any long metabolomic experiment to be broken into two separate analyses, usually about a week apart. This allowed for the instrument to be

used for service samples and by other lab mates. It also led to irreproducibility for one set of RBC samples to the other, which is why there are only half as many as the plasma samples. This was amplified by the fact that the instrument received maintenance in between the two weeks that the experiment was broken into.

Table 2-1. Liquid chromatography gradient.

Time (minutes)	A:B
0.0-6.5	95:5
25.5-35.0	0:100
45.0-50.0	95:5

A:B represents the ratio of mobile phase A to mobile phase B at a given time.

Table 2-2. Results from ANOVA for piglet plasma, $p = 0.01$.

Day of Life	8	7	6	5	2	3	4
8	82	4	3	10	44	29	14
7	78	82	5	12	48	28	15
6	79	77	82	4	34	16	9
5	72	70	78	82	34	10	6
2	38	34	48	48	82	14	30
3	53	54	66	72	68	82	8
4	68	67	73	76	52	74	82

Values above the diagonal indicate those masses that are differentially expressed between the two indicated days. The values below the diagonal indicate the masses that are not differentially expressed. The diagonal values (82) show the total number of statistically significant masses between all the days of life.

Table 2-3. PLS-DA prediction results for piglet plasma.

			Predictions				
Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Accuracy (%)
12	0	0	0	0	0	0	100.00
1	11	0	0	0	0	0	91.67
0	2	10	0	0	0	0	83.33
0	0	0	10	0	2	0	83.33
0	0	2	1	5	1	3	41.67
0	0	0	0	0	11	1	91.67
0	0	0	1	0	2	9	75.00
							80.95

There are 12 files for each day of life, thus each should have a prediction of 12, those that vary cause the accuracy of the model to decrease.

Table 2-4. Results from ANOVA for piglet RBCs, $p = 0.01$.

Day of Life	8	7	6	5	2	3	4
8	23	4	5	4	8	9	8
7	19	23	8	9	15	14	10
6	18	15	23	0	2	4	2
5	19	14	23	23	4	4	4
2	15	8	21	19	23	4	3
3	14	9	19	19	19	23	1
4	15	13	21	19	20	22	23

Values above the diagonal indicate those masses that are differentially expressed between the two indicated days. The values below the diagonal indicate the masses that are not differentially expressed. The diagonal values (23) show the total number of statistically significant masses between all the days of life.

Table 2-5. PLS-DA prediction results for piglet RBCs.

			Predictions				
Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Accuracy (%)
6	0	0	0	0	0	0	100
0	5	0	1	0	0	0	83.33
0	4	2	0	0	0	0	33.33
0	0	0	6	0	0	0	100
0	0	0	0	3	3	0	50
0	0	0	0	0	6	0	100
0	0	0	0	0	0	6	100
							80.95

There are 6 files for each day of life, thus each should have a prediction of 6, those that vary cause the accuracy of the model to decrease.

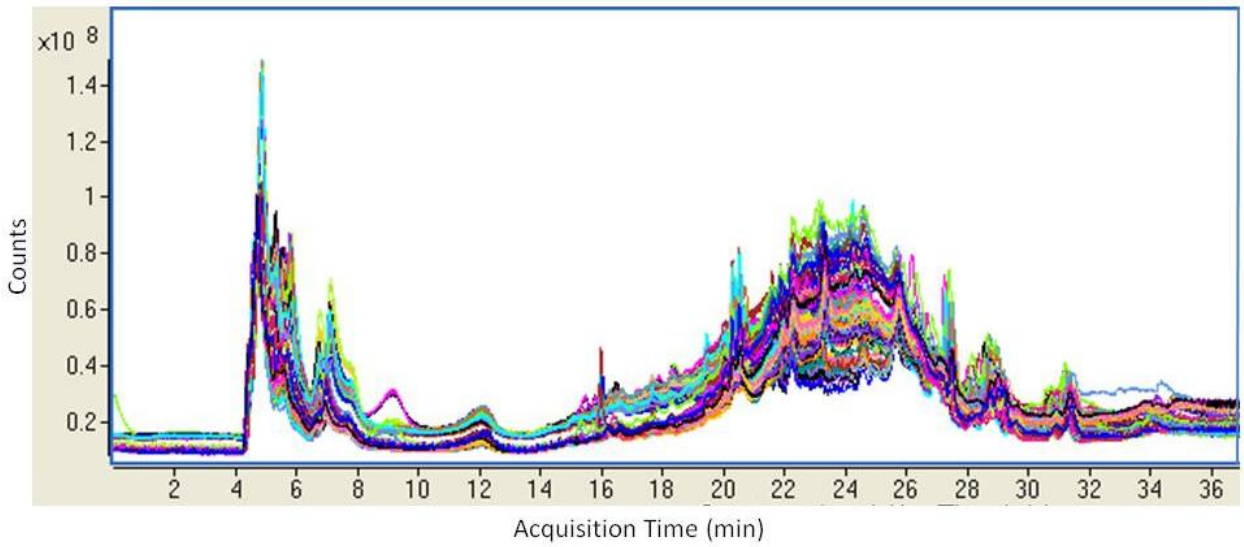


Figure 2-1. Overlay of 84 piglet plasma chromatograms.

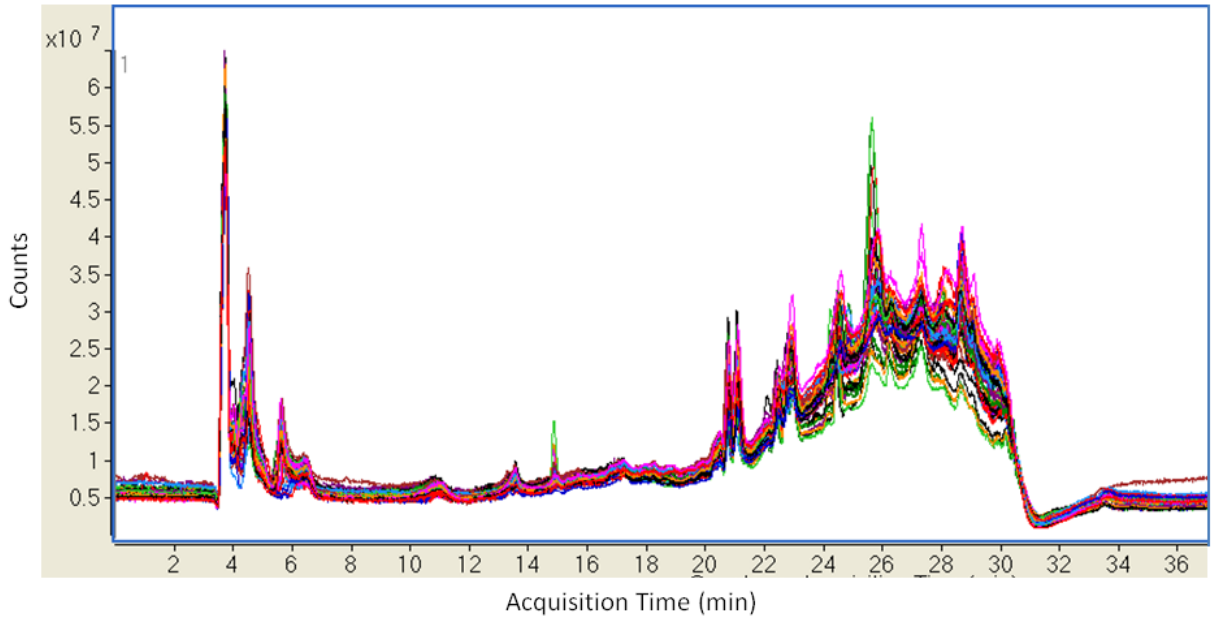


Figure 2-2. Overlay of 42 piglet RBC chromatograms.

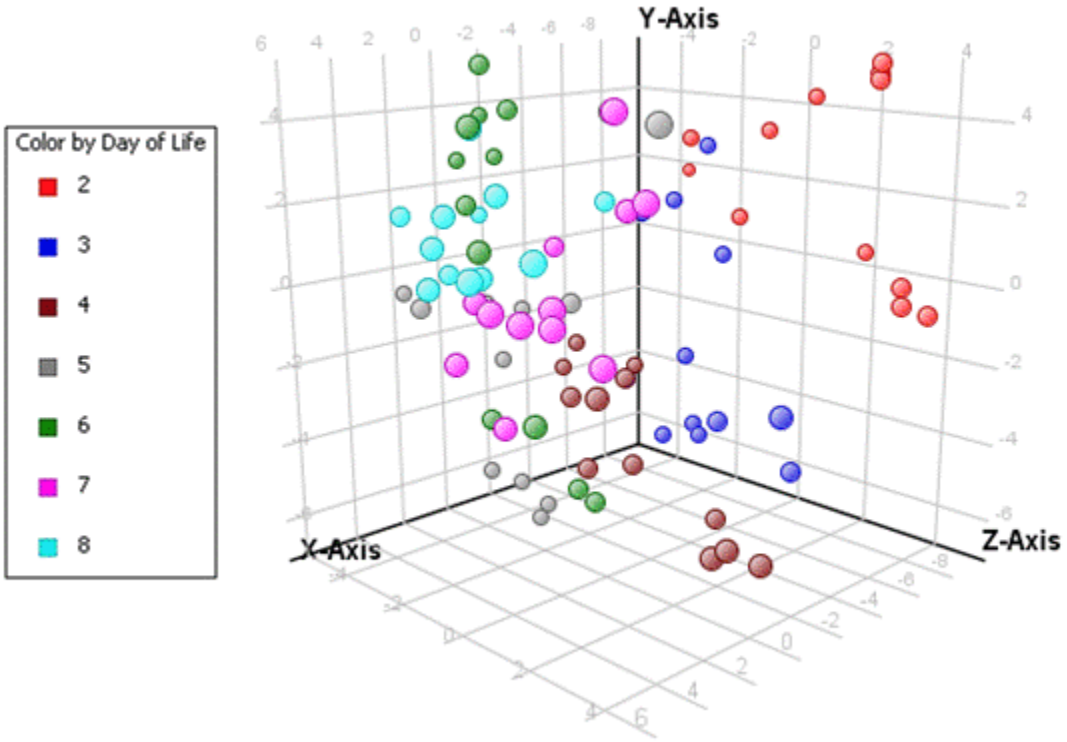


Figure 2-3. 3D-PCA plot of ANOVA masses in piglet plasma, filtered for 80% frequency (82 masses). Each color represents a different day of life. Each point of the PCA plot depicts one replicate.

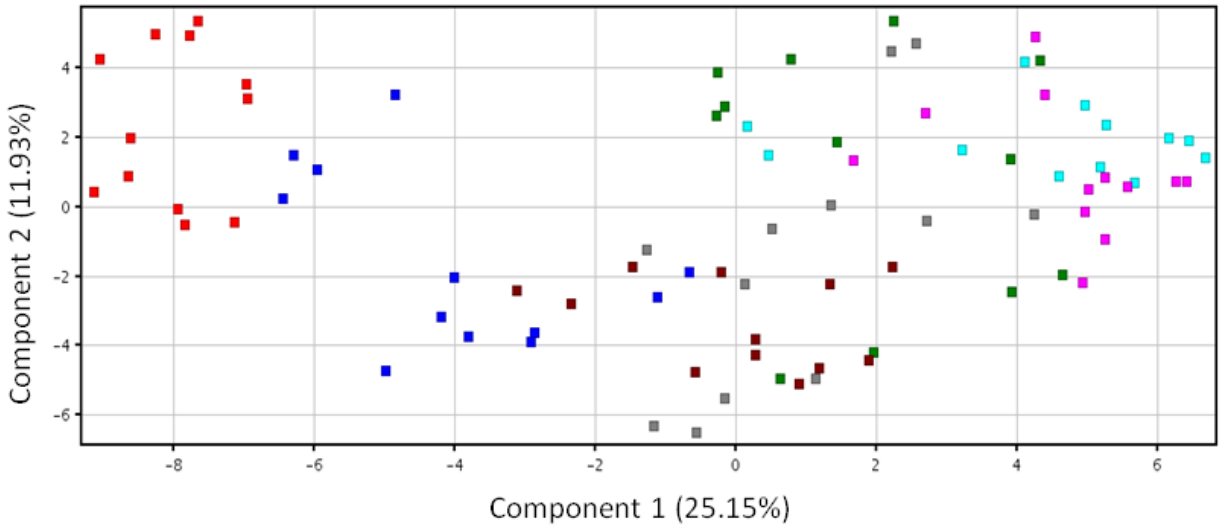


Figure 2-4. 2D-PCA plot of ANOVA masses in piglet plasma, filtered for 80% frequency (82 masses). Each color represents a different day of life and each point is one replicate. The color code is the same as Figure 2-3.

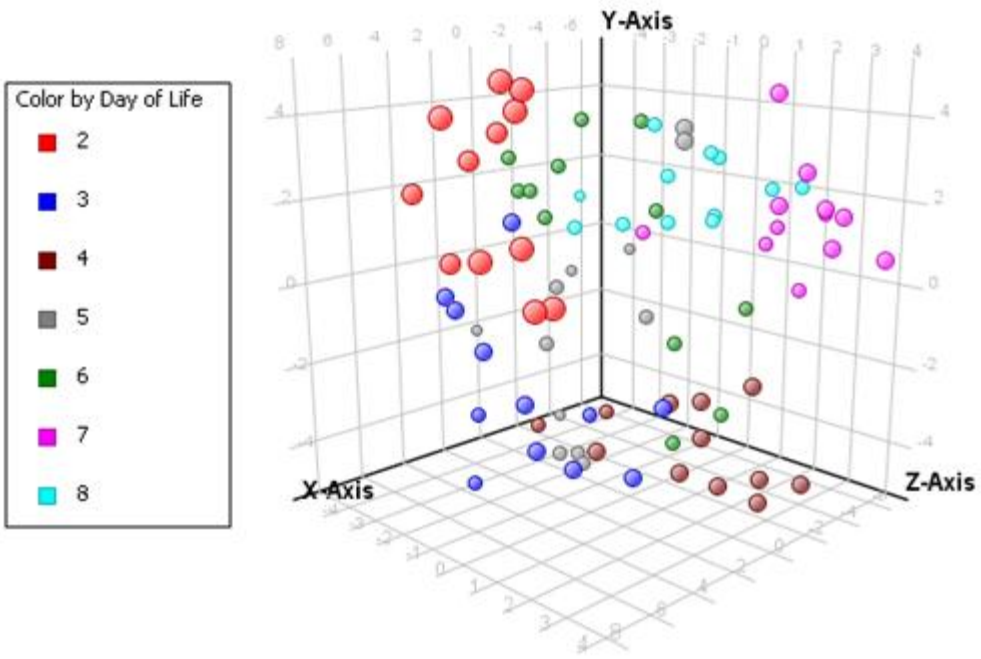


Figure 2-5. 3D-PLS-DA plot of ANOVA masses in piglet plasma, filtered for 80% frequency (82 masses). Each color represents a different day of life. Each point on the plot depicts one replicate.

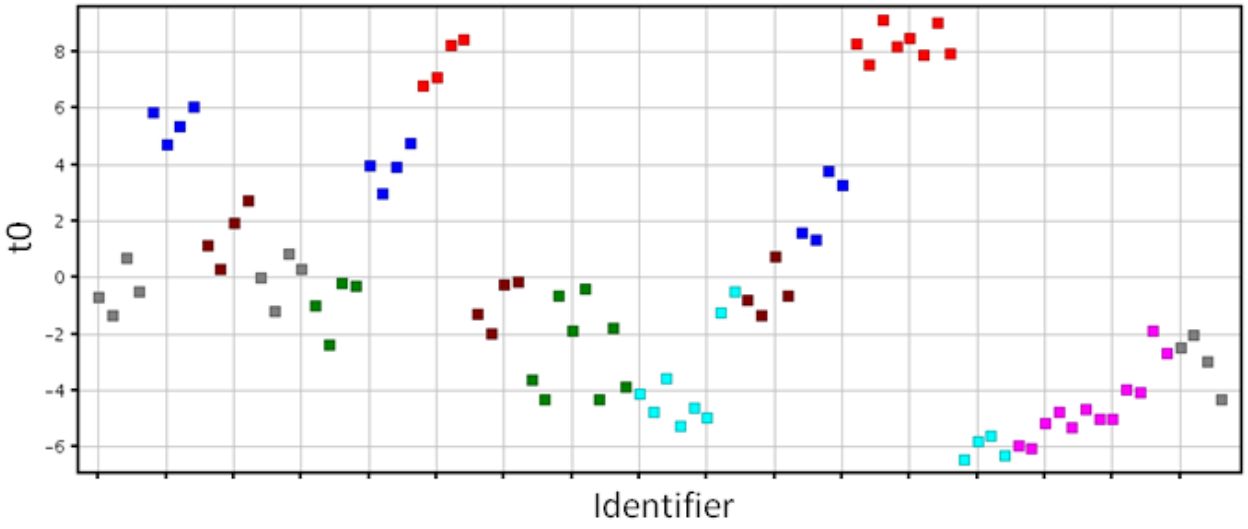


Figure 2-6. 2D-PLS-DA plot of ANOVA masses in piglet plasma, filtered for 80% frequency (82 masses). Each color represents a different day of life. Each point on the plot depicts one replicate. The x-axis (Identifier) shows individual replicates. The color code is the same as Figure 2-5.

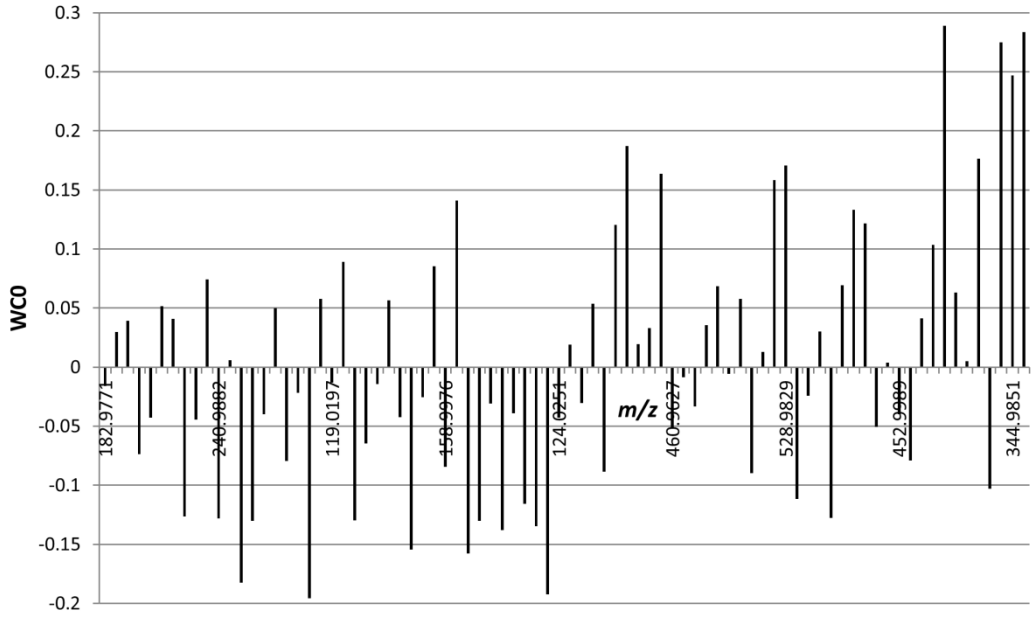


Figure 2-7. PLS-DA loadings plot of ANOVA masses in piglet plasma, filtered for 80% frequency (82 masses). WC0 indicates how influential a *m/z* value is (largest and smallest WC0 values are more influential in PLS-DA separation).

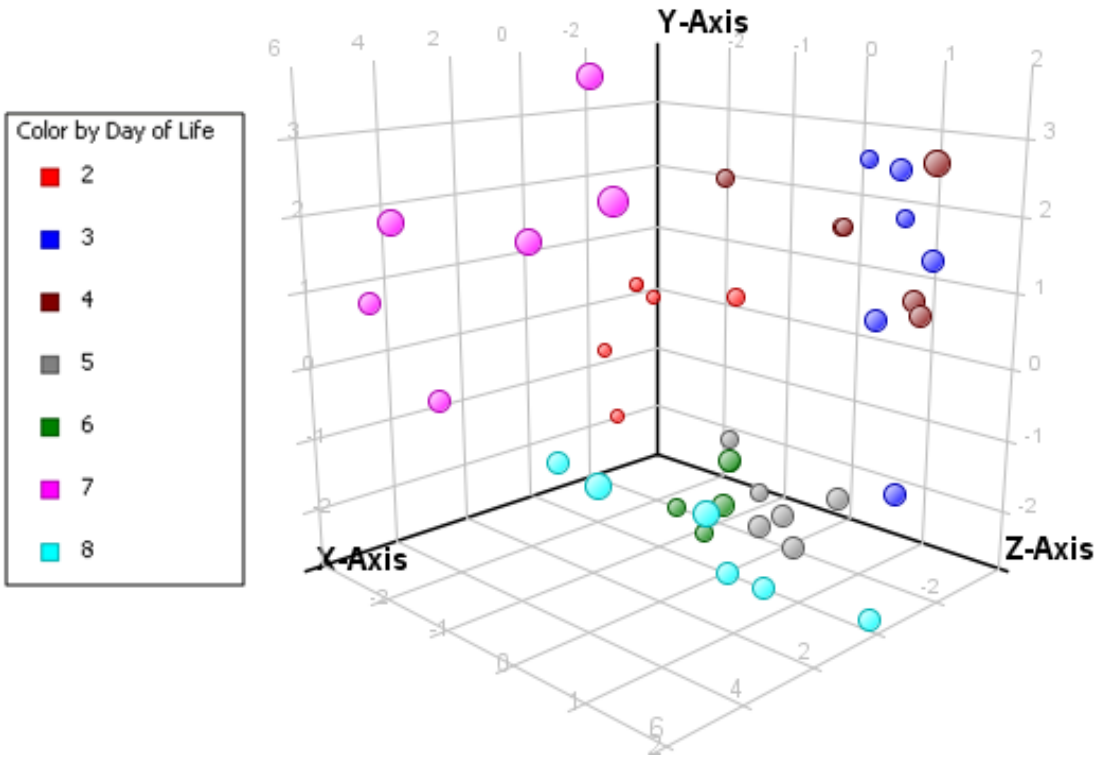


Figure 2-8. 3D-PCA plot of ANOVA masses in piglet RBCs, filtered for 80% frequency (23 masses). Each color represents a different day of life and each point is one replicate.

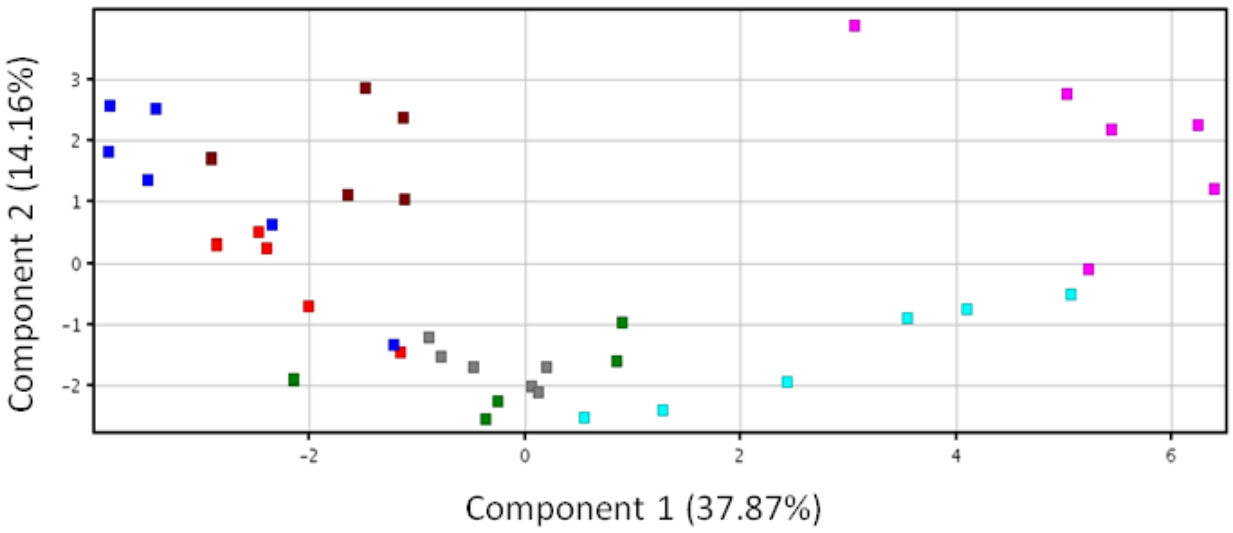


Figure 2-9. 2D-PCA plot of ANOVA masses in piglet RBCs, filtered for 80% frequency (23 masses). Each color represents a different day of life and each point is one replicate. The color code is the same as Figure 2-8.

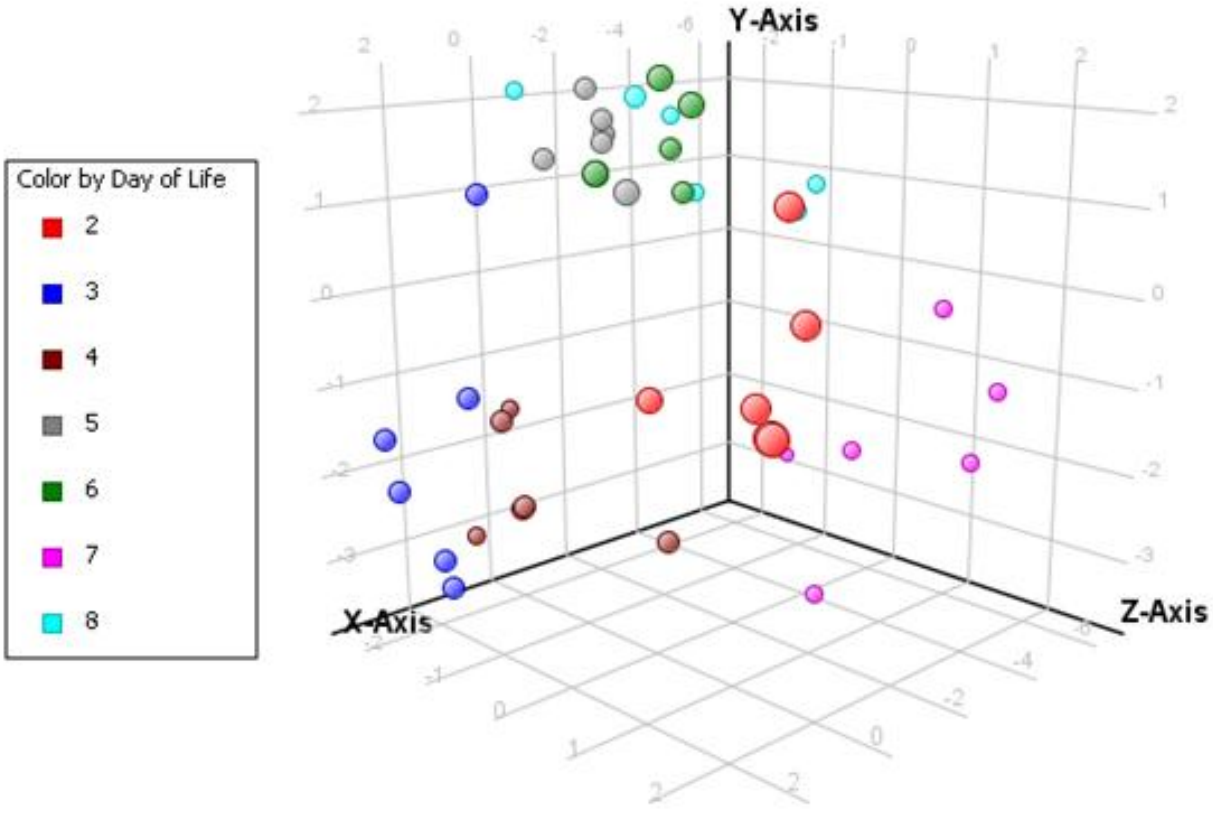


Figure 2-10. 3D-PLS-DA plot of ANOVA masses in piglet RBCs, filtered for 80% frequency (23 masses). Each color represents a different day of life. Each point on the plot depicts one replicate.

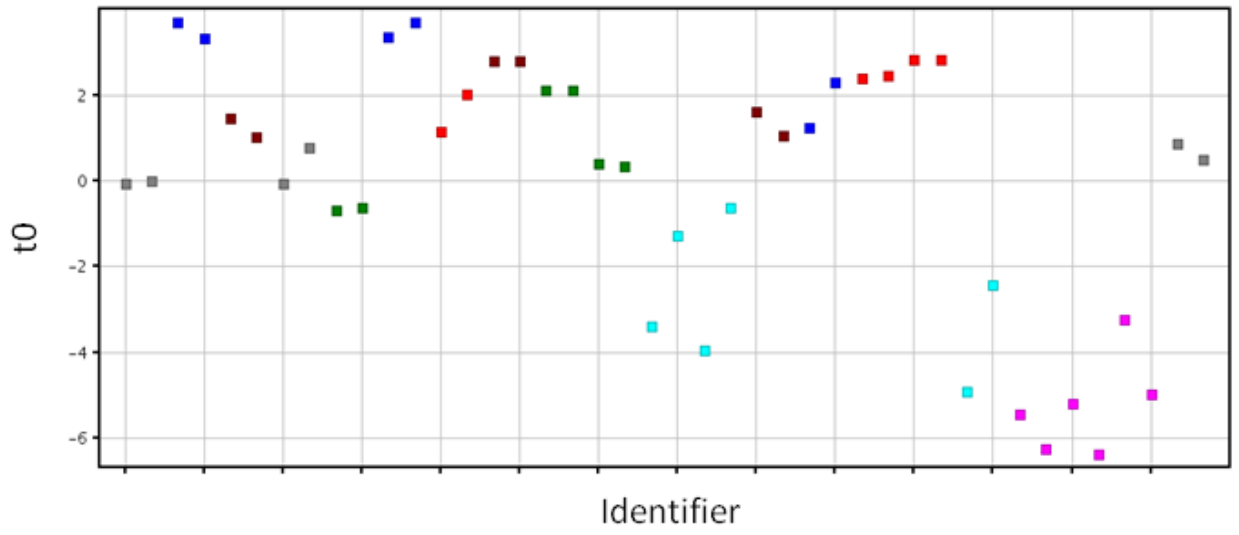


Figure 2-11. 2D-PLS-DA plot of ANOVA masses in piglet RBCs, filtered for 80% frequency (23 masses). Each color represents a different day of life. Each point on the plot depicts one replicate. The x-axis (Identifier) shows individual replicates. The color code is the same as Figure 2-10.

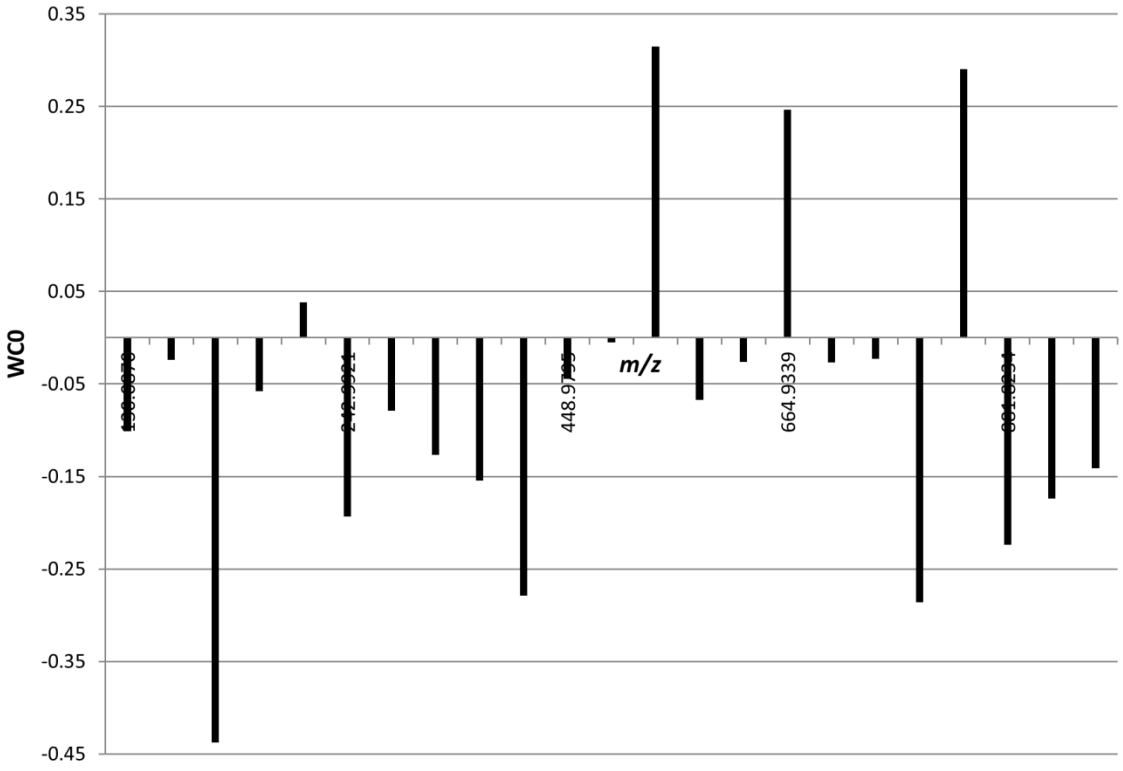


Figure 2-12. PLS-DA loadings plot of ANOVA masses in piglet RBCs, filtered for 80% frequency (23 masses). WC0 indicates how influential a *m/z* value is (largest and smallest WC0 values are more influential in PLS-DA separation).

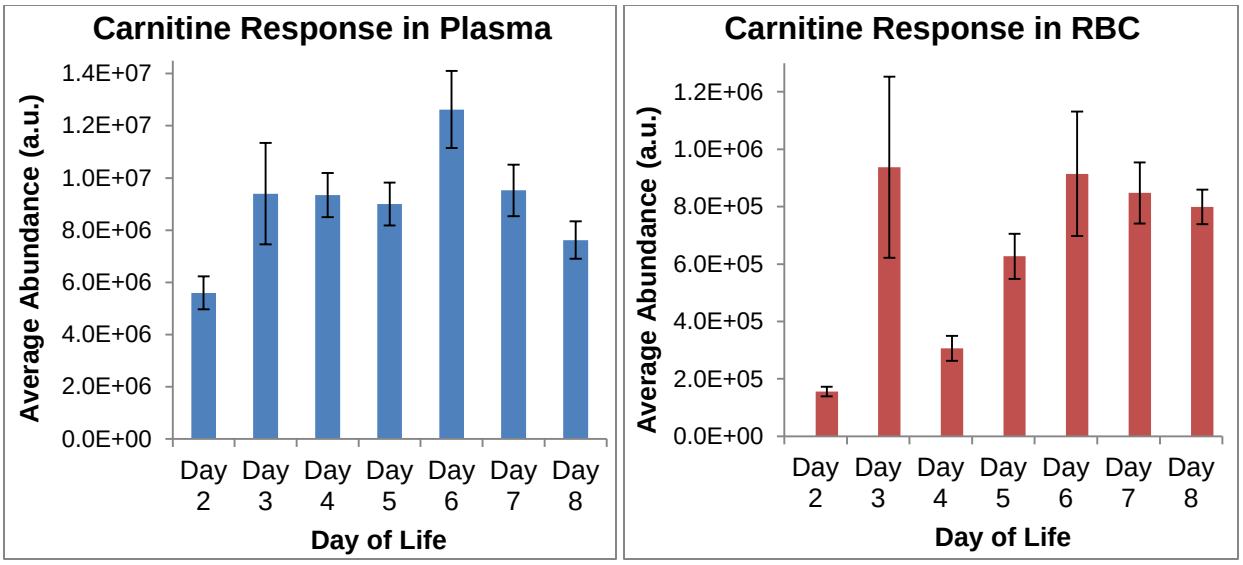


Figure 2-13. Bar graphs of carnitine response in piglet plasma and RBC. Error bars indicate the standard deviation of the mean.

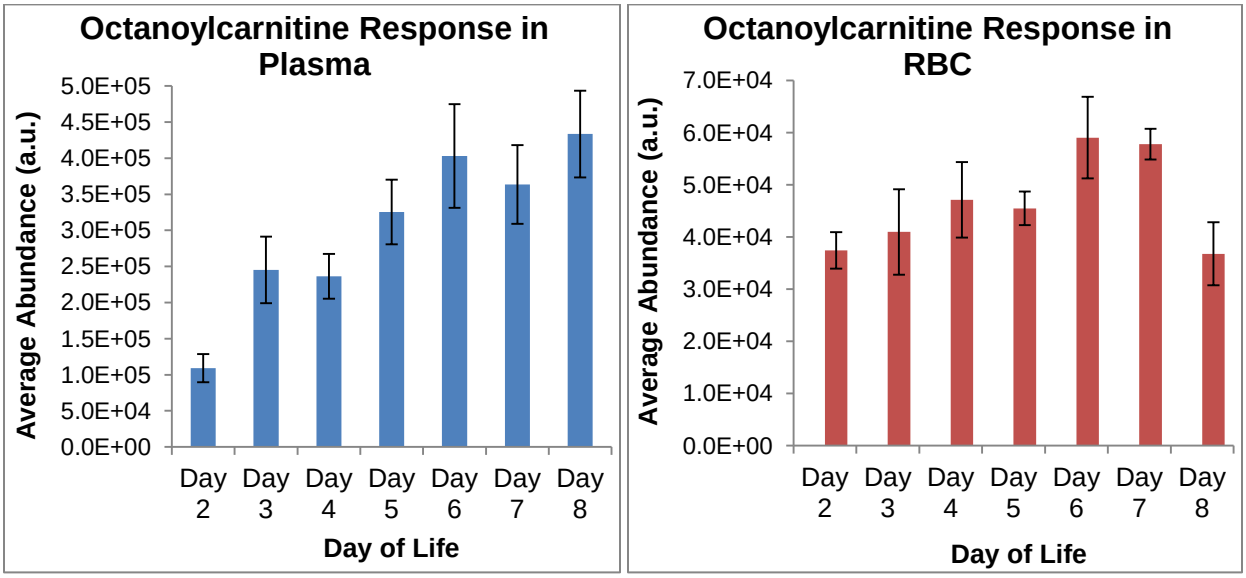


Figure 2-14. Bar graphs of octanoylcarnitine response in piglet plasma and RBC. Error bars indicate the standard deviation of the mean.

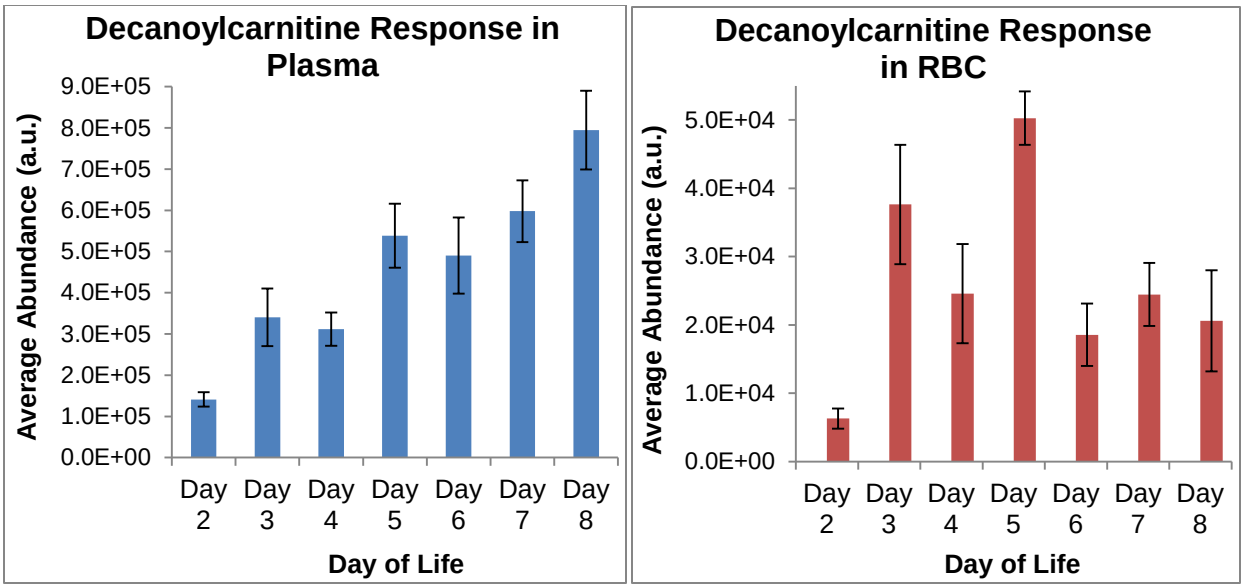


Figure 2-15. Bar graphs of decanoylcarnitine response in piglet plasma and RBC. Error bars indicate the standard deviation of the mean.

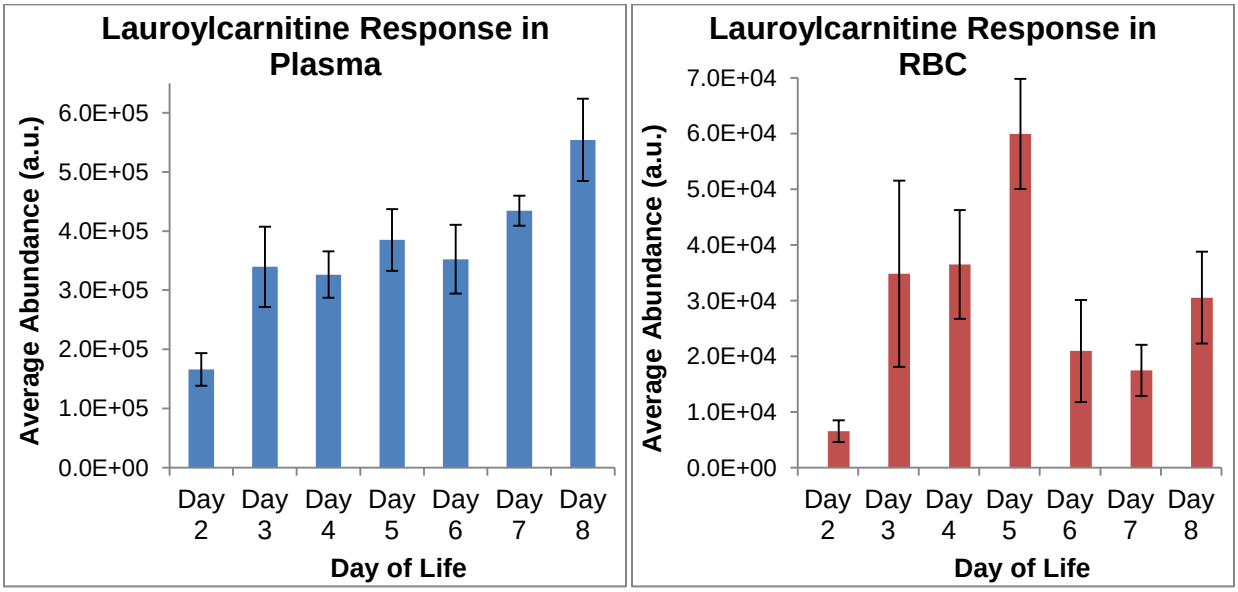


Figure 2-16. Bar graphs of lauroylcarnitine response in piglet plasma and RBC. Error bars indicate the standard deviation of the mean.

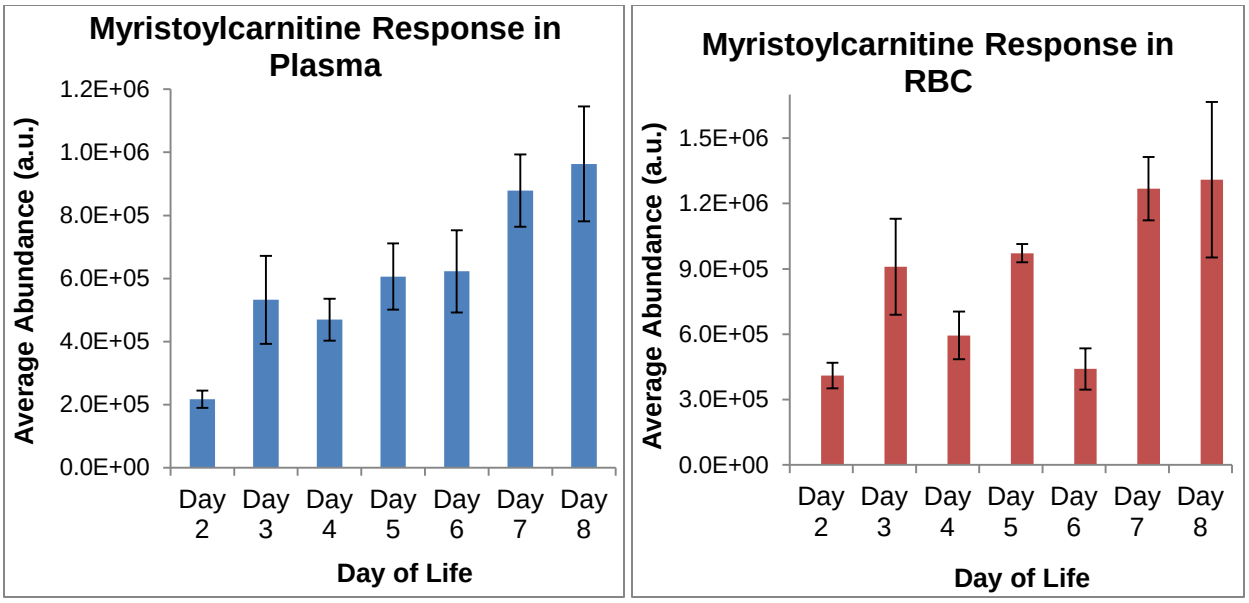


Figure 2-17. Bar graphs of myristoylcarnitine response in piglet plasma and RBC. Error bars indicate the standard deviation of the mean.

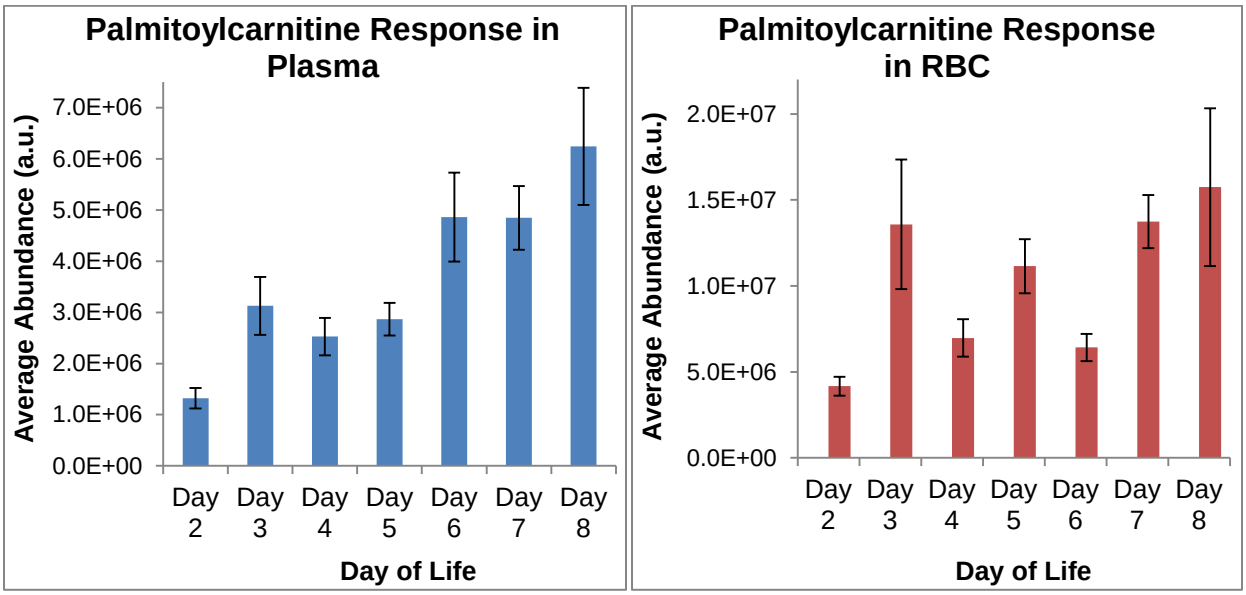


Figure 2-18. Bar graphs of palmitoylcarnitine response in piglet plasma and RBC. Error bars indicate the standard deviation of the mean.

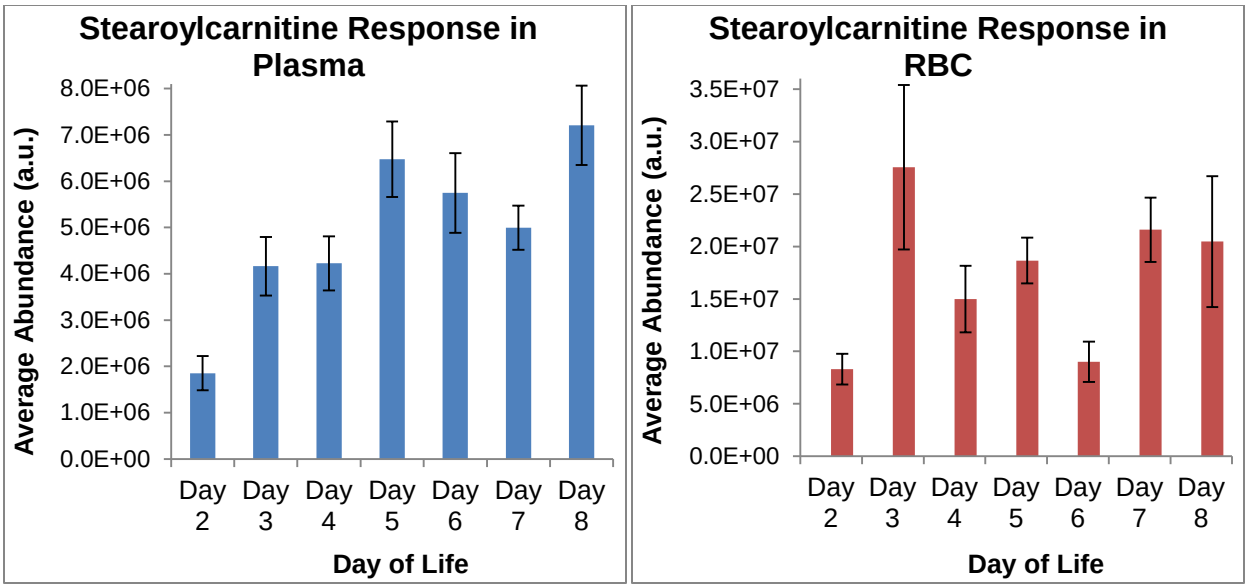


Figure 2-19. Bar graphs of stearoylcarnitine response in piglet plasma and RBC. Error bars indicate the standard deviation of the mean.

CHAPTER 3
METABOLOMICS OF COLON CANCER BY ANALYSIS OF HUMAN PLASMA

Overview

Colorectal cancer (CRC) affects the colon and/or rectum and is the second highest cancer-related cause of death when considering both males and females.[59] When men and women are considered individually, CRC is the third highest cancer-related cause of death, after lung cancer and breast and prostate cancers in women and men, respectively.[59] For 2012, the number of new colon and rectal cancer cases is estimated to be 103,170 and 40,290, respectively.[59] Although colon cancer and rectal cancer are frequently grouped together, 72% of the estimated new CRC cases are colon cancer, proving how important it is to better understand the disease. Colon cancer results from changes in polyps in the lining of the bowel. It takes approximately 10-15 years for an adenoma polyp to change to a carcinoma polyp; however, the chance of a benign one centimeter polyp to change to a malignant polyp is only 10%.[60] There are some things that can be done to try to protect oneself against CRC, including exercising and maintaining a diet that avoids a high amount of fat, as well as eating plenty of fruits and vegetables.

Because of the high mortality rate with CRC, as well as the readily available screening tools, early detection is the key to overcoming both colon and rectal cancers. It is recommended that anyone at average or slightly above average risk (98% of the population) undergo a colonoscopy at the age of 50 to screen for CRC.[61] While the colonoscopy can remove 90% of polyps, it is an uncomfortable and invasive procedure that requires preparation in order to remove the contents of the colon and rectum.[60] A less invasive screening tool is the fecal occult blood test. The fecal occult blood test

detects blood in stool samples, but it is not a very sensitive test. It is recommended that after more than 4/6 positive fecal tests the patient receives a colonoscopy. The fecal occult blood test can detect 72% of cancers with screening every two years.[60] The lack of a non-invasive, sensitive screening tool is one reason why the metabolomics of colon cancer in plasma should be examined. If a plasma biomarker can be determined, it could be possible for colon cancer to be screened by a simple blood drawing.

In order to better understand the samples investigated in this chapter, the method of staging colon cancer should be mentioned. The staging system used in this work is that given by the American Joint Committee on Cancer (AJCC).[62] By the AJCC standards there are five stages of cancer (0-IV). The AJCC staging is determined by considering how far the primary tumor has progressed through the colon wall (T), how many nearby lymph nodes have been afflicted (N), and how far the disease has metastasized (M). Because of this, the AJCC staging system is also sometimes called the TNM system.[63] As the staging of colon cancer is discussed below many areas in and around the colon will be mentioned; a cross section of the colon is shown in Figure 3-1 to better understand the complexity of the stages. Essentially, the inner most area of the colon is called the mucosa and is affected first with colon cancer cells. The second inner most region of the colon is called the submucosa, followed by muscle layers. The outer layer of the colon is called the serosa. In stage 0, abnormal cells are in the mucosa of the colon wall. The cancer then spreads from the mucosa to the muscle layer in stage I of colon cancer. With this early stage of cancer the five year survival rate for stage I colon cancer is 93.2%.[64] Within stage II there are three sub-stages: stage IIA, stage IIB and stage IIC. If the cancer has spread through the muscle

to the serosa then it is considered stage IIA. Stage IIB indicates that the colon cancer spread through the serosa but not to any lymph nodes/organs. The cancer is considered stage IIC if it has spread through the serosa, and to nearby organs. Stage II five year survival rates for IIA and IIB are 84.7% and 72.2% respectively. No survival rate for stage IIC was reported but is expected to be lower than stage IIB.[64] Colon cancer at stage III is again divided into three sub-stages: stage IIIA, stage IIIB and stage IIIC. Stage IIIA is characterized by two different conditions: 1) the cancer has spread through the mucosa to the submucosa and a muscle layer, and has spread to 1-3 nearby lymph nodes or tissue OR 2) the cancer has spread through the mucosa to the submucosa and 4-6 nearby lymph nodes. Three different scenarios can be seen in stage IIIB of colon cancer: 1) cancer spread through the muscle and the serosa, as well as to 1-3 nearby lymph nodes or tissue OR 2) cancer spread through the muscle to the serosa and 4-6 nearby lymph nodes OR 3) cancer spread through the mucosa to the submucosa/muscle layer and 7 or more nearby lymph nodes. The final sub-stage of stage III is stage IIIC, where there are the following three possibilities: 1) cancer has spread through the serosa to 4-6 nearby lymph nodes but to no nearby organs OR 2) cancer has spread to or through the serosa and 7 or more nearby lymph nodes but to no nearby organs OR 3) cancer has spread through the serosa to nearby organs and to one or more nearby lymph nodes/tissues. Five year survival rates are much lower for stage III colon cancer with the exception of stage IIIA. Stage IIIA, IIIB, and IIIC have five year survival rates of 83.4%, 64.1% and 44.3%, respectively.[64] When colon cancer has spread through the blood and lymph nodes to other parts of the body it is considered to be in the last stage, stage IV. At this point, the five year survival rate is

only 8.1%.[64] All the descriptions of the stages of colon cancer were taken from the National Cancer Institute at the National Institutes of Health's website.[65]

Most previous research has focused on transcriptional regulation of genes associated with CRC; therefore, the effect of CRC on metabolites and protein expression is not well understood. [60, 66] Previous research has shown that different cancers could be distinguished (colon and rectal), but successful discrimination between stages of the cancer could not be observed.[66-67] One study that compared CRC tissue to normal control tissue found higher up-take of glucose in CRC tissues, as well increased glycolysis in those same tissues.[68] Interestingly, Warburg proposed the idea of cancer cells using glycolysis to create energy in 1956.[66, 69] This phenomenon was shown in another study of CRC and normal control tissues as well.[70]

The majority of metabolomic CRC studies have been on tissues,[67, 71-72] but some studies have been performed on serum from persons with CRC.[73] Advantages of plasma analyses, compared to tissue analyses include: simple sample collection, the potential for rapid screening and no extraction method is needed (after protein precipitation). For these reasons, we have focused on studying the metabolomics of human plasma from 18 healthy controls and 18 colon cancer patients. Colon and rectal cancers are different from each other and therefore may have different metabolic profiles. Our study focuses on colon cancer to avoid complicating the experiment with two cancers.

Due to the nature of this disease, both the large number of individuals that are affected by it and the large number of deaths caused by it each year, new research on colon cancer is very important. This is the first global metabolomics work to focus only

on colon cancer patient plasma compared to healthy controls. Comparing these results to previous CRC studies may provide more insight about the metabolomics of colon cancer. This may allow for a better understanding of such a fatal disease so that better treatments can eventually be implemented.

Experimental

The experimental conditions are the same as those described in Chapter 2, unless otherwise indicated below.

Sample Collection and Preparation

Whole blood was collected between May 2005 and May 2009 from 36 individuals; 18 healthy controls and 18 colon cancer patients with stage II or stage III cancer (each stage is composed of level A, B, and C). Within each group (controls and cases) there were equal numbers of males and females (9 each). The mean patient age was 65.6 years old and the median age was 67.5 years old. The control group had a mean age of 65.5 years old and a median age of 67.5 years old. The sample demographics are shown in Table 3-1.

The samples were prepared as discussed in Chapter 2. Plasma was protein precipitated and evaporated to dryness. The dried products were dissolved in 100 µL of 40% acetonitrile, 40% methanol, 10% water and 10% isopropanol.

Liquid Chromatography Conditions

The LC conditions are described in Chapter 2 and Table 2-2. An injection volume of 5 µL was used. All samples were analyzed in triplicate for a total of 108 LC/MS files.

Electrospray Source

The source conditions used are indicated in Chapter 2. The only change made was a gas temperature of 325˚C was used instead of 300˚C, as recommended by the manufacturer.

Mass Spectrometer

The TOF mass spectrometer settings are listed in Chapter 2.

MS/MS Instrumentation

Analyses were also performed in positive ESI mode on a Velos dual pressure linear ion trap (LIT) mass spectrometer (Thermo Scientific, Waltham, MA) capable of MS/MS analyses. Chromatography conditions used for the LC/TOF analysis were copied on an Accela 600 LC system (Thermo Scientific, Waltham, MA). The MS/MS method included a parent mass list of 64 masses determined to be influential to the PLS-DA plot, as well as those masses with the highest fold change. In addition to the parent mass list a data dependent acquisition was implemented to perform MS/MS on the most intense ion if no parent mass was found. Other settings included: source voltage of 3,000 V, source temperature of 350˚C, CID activation, collision energy of 35%, isolation width of two, and a signal threshold of 10,000 counts. By using both accurate mass and MS/MS fragmentation compound identification was possible for some ions of interest.

Data Analysis

As described in Chapter 2, data analysis was carried out with Agilent's Qualitative Analysis (for feature extraction and file conversion) and Agilent's MPP for statistical analysis (Agilent Technologies, Inc., Santa Clara, California). The abundance threshold was set to 2,500 a.u., as determined by examining the intensity of the background.

Files did not require recalibration since a reference solution was sprayed along with the LC flow for constant real-time calibration.

No targeted approach was taken in this study, but some statistically significant *m/z*'s were extracted using Quantitative Analysis (Agilent Technologies, Inc., Santa Clara, California). This allowed for the calculation and comparison of average abundances of particular compounds in healthy controls compared to cancer patients and vice versa.

MS/MS data analysis was performed with Thermo Scientific's Xcalibur Qual Browser (Thermo Scientific, Waltham, MA).

Results and Discussion

A total of 3,858 aligned compounds were determined to be present in the 108 data files. After performing an asymptotic *t*-test ($p = 0.01$), with a Benjamini-Hochberg FDR multiple testing correction, 165 masses were determined to be significant. A 3D-PCA plot of the 165 significant masses can be seen in Figure 3-2. In the 3D-PCA the x-axis accounts for 39.17% variation, while the y- and z-axes represent 20.67% and 6.11% of the variations, respectively. The 2D-PCA is also shown in Figure 3-3. By examining the PCA plots it can be noted that all the controls group together, as well as a few cases, but most of the cases seem to spread out away from the controls. There are six case samples that make up the small cluster near the control samples in the PCA plots. This suggests that although the colon cancer is present in these six patients, the samples may not contain as many of the significant features found in the rest of the case samples. This is why the six samples group closer to the controls than the remainder of the case samples. When the significant masses were examined more closely, this was

exactly the cause. Many of the significant masses determined to be more intense in the cancer patients than in the controls are missing from these six individuals.

The significant masses determined by the *t*-test ($p = 0.01$) were also used to create a 3D- and 2D-PLS-DA plot (Figure 3-4 and 3-5). The PLS-DA yielded a prediction accuracy of 93.52% (Table 3-2). In the PLS-DA model the case samples had a prediction accuracy of 100%, while the control samples were only predicted correct for 47 of the 54 replicates. The loadings plot (Figure 3-6) was used to examine the masses that were farthest from WC0 = 0 (WC0 < -0.1 or WCO > 0.1). These masses (22) will be discussed below.

Influential PLS-DA Masses

From the PLS-DA plots, 22 masses were determined to be the most influential in the separation observed. These masses were further investigated with Agilent's Quantitative Analysis software (Agilent Technologies, Inc., Santa Clara, California). More specifically, the extracted ion chromatographs (EICs) for each of the 22 *m/z*'s were plotted for all 108 data files and the peak areas were extracted. The extraction was performed using Quantitative Analysis and was manually inspected. Any ions that had unusual repeatability between replicates, or weren't found using Quantitative Analysis were then manually extracted and integrated in Agilent's Qualitative Analysis. For each ion, two average peak areas were calculated: one for the cases and one for the controls. The standard deviation of the mean was also calculated for each average. Of the 22 masses, 11 were used to create a bar graph comparing the average peak areas for each group (control and case). The reason for only 11 masses being depicted in Figure 3-7 is that a mass had to be present in at least 60% of the samples, in at least one of the groups (in at least 11 of the 18 samples for case or control), to be

considered. Additionally, a signal-to-noise ratio (S/N) of at least 3 was necessary to be considered present in any sample. By examining Figure 3-7, it is obvious that many of the significant masses influencing the PLS-DA plots are present only in the cancer patient's plasma samples. These masses were investigated further by a database search as well as with MS/MS analysis.

Significant Masses with Highest FC

Fold change is often used to determine what masses are changing most from one group to another. Additionally, it can be used to filter a large number of significant masses down to a more manageable number. This is sometimes done by concentrating on significant masses that also have a fold change of at least 2 from one group to another. The masses with the highest fold change (FC) from either case to control, or control to case, were considered in this work. The masses that passed the *t*-test and also had the highest FC (FC >1400) are compared in Figure 3-8 (14 masses). This was done in the same manner described under the heading Influential PLS-DA Masses, which is above. As can be seen in Figure 3-8, all of the masses are found in either the case or control samples, but not both. Also, a majority of the masses are present in only the cancer patient's plasma, as was observed for the influential PLS-DA masses. A database search and MS/MS analysis of these masses will be discussed later.

The fact that the FC masses were only present in either the case or control samples is both a con and a pro for data analysis. The negative aspect is that the statistical software is putting in a very small value for the abundance of a mass that is actually not present in either the case or the control samples. Thus, when the fold change is used to filter masses it results in very high changes and is essentially only

determining masses present in only one of the groups. This makes it difficult to look at masses that are actually present in both groups and changing greatly. The advantage of this is that a compound not present in control samples but present in cancer patient samples may be a very easy biomarker to screen for. This is because the concentration of the biomarker present would not be as important as just determining if it is detectable. Often times, when a biomarker is found it is either elevated or found to be decreased in the diseased samples. This means developing a diagnostic tool for colon cancer would be more difficult, since the concentration would need to be taken into account.

Database Search of Significant Masses

As mentioned, the *t*-test masses were further filtered by considering masses in the PLS-DA loadings plot that satisfied the conditions of WC0 < -0.1 or WC0 > 0.1 (Figure 3-6). This resulted in 11 masses to consider, which were then applied to an accurate mass database search. The database search was performed using METLIN.[58] The mass of the positively charged ion was used in the database search and a mass accuracy of 10 ppm was applied. Two different ion options were searched, $[M+H]^+$ and $[M+Na]^+$, as discussed in Chapter 2. The database search of $[M+H]^+$ ions for the PLS-DA masses resulted in hits for 8 of the 11 masses searched. These are listed in Table 3-3. Most of the results generated multiple potential compounds for one *m/z* with the same formula. No ID is given for those search results. As would be expected, some ions resulted in multiple potential formulas, such as *m/z* 303.2159 in Table 3-3. These possible identifications may help make the MS/MS characterization easier. The $[M+Na]^+$ ion search results are listed in Table 3-4 for the PLS-DA influential masses. Only 4 of the 11 masses received hits in the database search. This may be because the masses are more likely $[M+H]^+$ ions.

Masses that passed the *t*-test were also filtered by considering the significant masses that had the highest FC from control to case or case to control (Figure 3-8). These 14 masses were also applied to an accurate mass database search, as described above and in Chapter 2. A search of $[M+H]^+$ ions resulted in hits for 4 of the 14 masses, three of which had single formula hits (Table 3-5). The $[M+Na]^+$ database search resulted in hits for 7 of the 14 ions searched. The results for the $[M+Na]^+$ ion search are shown in Table 3-6. All the potential formulas and ID's will be considered when characterizing the significant masses by MS/MS.

MS/MS of Significant Masses

Identifying compounds based on accurate mass alone is impossible. By implementing MS/MS, fragmentation patterns were found for most of the significant ions. Identification was done for some ions by use of accurate mass and MS/MS databases as well as literature searches for specific ions or neutral losses observed in the spectra.

Before describing the resulting data the MS/MS instrumentation will be mentioned briefly. A schematic of the dual pressure LIT is shown in Figure 3-9. In general, ions formed at the source are transferred into the MS and go through the S-lens. The S-lens is a series of flat ring electrodes with radio frequency (RF) voltage applied to them. Every other electrode has an opposite phase RF and each ring is an increased distance from the last. The S-lens openings were designed to allow optimal ion transfer while minimizing the RF voltage required for focusing.[74] After ions are guided through two quadrupoles and an octopole they reach two trapping cells maintained at different pressures. The first cell is held at approximately 5×10^{-3} Torr. One benefit of higher pressure in the first cell is improved fragmentation efficiency over typical LITs. The

second cell is maintained at about 4 X10^{-4} Torr and allows improved mass resolution for a given scan rate, or this resolution can be sacrificed for a higher scan rate.[74] The dual pressure LIT is faster and more sensitive than its previous models but still lacks the mass accuracy that is desired for metabolomic applications.

When examining the PLS masses four ions were quickly identified due to correct accurate mass database results and corresponding MS/MS spectra. All four of these compounds were carnitines. More specifically, myristoleylcarnitine (C14:1), myristoylcarnitine (C14:0), palmitoleylcarnitine (C16:1), and linoleoylcarnitine (C18:2) were identified. Table 3-7 lists the experimental masses observed with the LC/TOF instrumentation, the theoretical masses, as well as the calculated retention times observed. The MS/MS spectra for each of the four identified carnitines in plasma from colon cancer patients can be seen in Figures 3-10 to 3-13. Carnitines are relatively easy to identify by MS/MS fragmentation due to a characteristic neutral loss of 59, corresponding to the trimethyl amine, as well as the neutral loss of 161 from the loss of carnitine. For carnitines, an ion is also observed at m/z 144 that is dehydrated carnitine. This ion can be seen in Figure 3-11, although it is present in all four of the MS/MS spectra. A standard was analyzed for C14:0 and is compared to the MS/MS spectrum from plasma in Figure 3-11.

MS/MS spectra were acquired for several of the fold change masses. Unfortunately, no compounds have been able to be identified. Some characteristics of the masses could be determined, such as if it was containing a hydroxy group, a phosphate group or a fatty acid chain. However, enough information to identify the compounds was not obtained.

Spectra was acquired for several of the remaining significant masses (not grouped with the FC masses or the influential PLS masses) as well. Another carnitine, which was observed to be elevated in cancer cases, was identified at *m/z* 344 as dodecanoylcarnitine (C12:0). The same neutral losses and similar fragment ions were observed in the MS/MS spectrum of *m/z* 344 from colon cancer patient plasma as other carnitines already discussed (Figure 3-14). A standard was obtained for C12:0 and the standard spectrum matched the experimental spectrum from colon cancer patient plasma (Figure 3-14).

Biological Significance

It is interesting that five significant compounds were identified as carnitines with varying fatty acid tail lengths. Previous work with CRC patient serum samples compared to healthy controls has suggested a down regulation of fatty acids such as myristic acid (C14 chain) and palmitic acid (C16 chain) in cancer samples.[73] The conclusion of this finding was a dysfunction in the β-oxidation pathway used for fatty acid metabolism. Another study of CRC tissue compared to healthy tissue resulted in up regulation of oleic acid (C18:1 chain) and stearic acid (C18 chain) in CRC samples.[67] The authors suggested that this increase occurs with increased cell death that may be occurring more frequently in the malignant tissue.[75] Another study of colon tissue and healthy tissue showed elevated palmitic and oleic acids in healthy tissue.[71] These results match those already discussed.

In is unclear if the carnitines identified in this work would be associated with the cell death mentioned by Lehtimaki et al., but it is likely that that some type of defect is occurring in the β-oxidation pathway. In particular, when medium and long chain carnitines are elevated metabolism disorders are often the culprit. These disorders can

sometimes be the result of a defect in carnitine palmitoyltrasferase (CPT2) in the mitochondria. As shown in Figure 3-15, free carnitine is used to transfer fatty acids into the mitochondria via the translocase. The acylcarnitines (carnitine with varying fatty acid tails) are then converted back to free carnitine and Acyl-CoA by CPT2.[76] If a defect occurs in CPT2 then one might suspect an increase in acylcarnitines, as was a result in this research.

Finally, Wright et. al. found an increase in the myristoylation of proteins in a rat colon cancer model.[77] If increased myristoylation is occurring in individuals with colon cancer then an increase in the transportation of C14 chains may be necessary. Thus, an elevation of myristoylcarnitine may be observed in colon cancer patients, such as was found here.

Obviously more biological understanding of the processes of the identified metabolites must be acquired. With a better understanding treatments for colon cancer can be considered and experiments can be designed for these treatments.

Challenges Faced

There were several challenges associated with the MS/MS aspect of this work. Firstly, although no data was presented here, a significant amount of time was spent attempting to utilize a quadrupole-TOF (Q-TOF) instrument located at Moffitt Cancer Center. A major flaw in the Q-TOF system was the ionization source. The Q-TOF was equipped with an “ionspray”, capable of mobile phase flow rates of approximately 2-40 µL/min. Due to the higher flow rate used in this work (800 µL/min) a split was required for compatibility. However, when the split was used there was a large decrease in the sensitivity of the instrument. This was a result of most of the analyte going to waste. In order to solve this problem a capillary version of the monolithic column was purchased

from Phenomenex. Unfortunately, the capillary column offered its own difficulties, especially with sample overloading. It is believed that the column was quickly damaged due to the fragile nature of capillary columns. Finally, a “turbo-ionspray” was received from the manufacturer and configured to the Q-TOF. The new source was capable of up to 1 mL/min flow rates. However, when the chromatography method used in this was implemented with the new source desolvation suffered. This was evident by the accumulation of solvent within the source, even with a flow rate of only 0.6 mL/min. The lack of desolvation is believed to have caused a loss of signal that was observed at the end of one day of research. The periodic loss of signal resulted in the necessity of finding a different MS/MS capable instrument to use.

In addition to instrumental problems, the process of identification of compounds by accurate mass and MS/MS fragmentation proved to be difficult. While accurate mass libraries such as HMDB and METLIN are useful they are still incomplete. This means that many significant masses that are searched result in no hits. Also, the amount of MS/MS spectra available in these databases is even more limited. Due to this, more literature searching is necessary to try to find MS/MS work that has already been done on database results from accurate masses. Additionally, specific neutral losses or ions observed in MS/MS can be searched in the literature. Another challenge that was faced when examining the MS/MS spectra of the significant masses was that many were related to each other by either a loss of water or the addition of sodium. This is supposed to be addressed in the software, or least the addition of sodium should be easily determined. Obviously this is not always the case, so careful consideration of retention times and ions present in the MS/MS spectra should be used. A recent

publication also mentioned the many challenges associated with metabolomics research, and the bottleneck of the work flow is the actual identification of features.[78] With this in mind, there are many advances to be made in software capable of processing metabolomics data, as well as in expanding accurate mass and MS/MS libraries.

Summary

LC/TOF analysis of plasma from colon cancer patients and healthy controls resulted in 165 significant masses. These masses were also analyzed by LC/LIT to obtain MS/MS fragmentation patterns. To simplify the identification process, masses that were found to be influencing the PLS-DA plot or to have the highest fold change were considered initially. This resulted in the identification of five carnitines. Considering all other significant features one more compound was identified. Some of these results suggest a disturbance in the β-oxidation pathway.

Table 3-1. Sample demographics. The median age for both groups is 67.5 years old.

	Colon Cancer Patients	Healthy Controls
# of Samples	18	18
Age (mean, range)	65.6, 40-86	65.5, 40-88
Sex (male, female)	9, 9	9, 9
Stage 2 (A, B, C)	7, 0, 0	N/A
Stage 3 (A, B, C)	2, 4, 5	N/A

Table 3-2. PLS-DA prediction results for human plasma from healthy controls and colon cancer patients.

	Predictions	
Case	Control	Accuracy (%)
54	0	100.0
7	47	87.04
		93.52

There are 54 files for each group, thus each should have a prediction of 54, those that vary cause the accuracy of the model to decrease.

Table 3-3. Potential formulas from $[M+H]^+$ database search of influential PLS-DA masses.

m/z Searched	Formula	Theoretical *m/z*	Error (ppm)	ID
289.2005	$C_{15}H_{28}O_5$	289.2010	-1	1,2-Dihexanoyl-sn-glycerol
303.2159	$C_{12}H_{26}N_6O_3$	303.2139	6	
	$C_{16}H_{30}O_5$	303.2166	-2	
	$C_{17}H_{26}N_4O$	303.2179	-6	Emedastine
372.3111	$C_{21}H_{41}NO_4$	372.3108	0	Tetradecanoylcarnitine
398.3264	$C_{23}H_{43}NO_4$	398.3265	0	Hexadecenoyl carnitine
424.3417	$C_{25}H_{45}NO_4$	424.3421	-1	Linoleyl carnitine
427.2694	$C_{23}H_{38}O_7$	427.2690	0	
	$C_{19}H_{41}NO_7P$	427.2693	0	
431.2640	$C_{23}H_{34}N_4O_4$	431.2653	-2	
467.2619	$C_{22}H_{43}O_6PS$	467.2591	6	(2S)-OMPT
	$C_{25}H_{38}O_8$	467.2639	-4	

A mass accuracy of 10 ppm was required. ID was given if no other compounds with the same formula resulted.

Table 3-4. Potential formulas from $[M+Na]^+$ database search of influential PLS-DA masses.

m/z Searched	Formula	Theoretical *m/z*	Error (ppm)	ID
431.2640	$C_{26}H_{36}N_2O_2$	431.2669	-6	O-2545 (hydrochloride)
467.2619	$C_{23}H_{40}O_8$	467.2615	0	
485.3108	$C_{28}H_{46}O_3S$	485.3060	9	
	$C_{30}H_{42}N_2O_2$	485.3138	-6	Arachidonoyl Serotonin
489.3037	$C_{27}H_{46}O_4S$	489.3009	5	Cholesterol Sulfate

A mass accuracy of 10 ppm was required. ID was given if no other compounds with the same formula resulted.

Table 3-5. Potential formulas from [M+H]+ database search of masses with the highest FC.

m/z Searched	Formula	Theoretical *m/z*	Error (ppm)	ID
277.0626	$C_{13}H_{12}N_2O_3S$	277.0641	-5	Sulfabenzamide
283.0705	$C_{13}H_9F_3N_2O_2$	283.0689	5	Niflumic Acid
445.2800	$C_{23}H_{40}O_8$	445.2796	0	
597.3603	$C_{33}H_{48}N_4O_6$	597.3647	-7	L-Urobilinogen

A mass accuracy of 10 ppm was required. ID was given if no other compounds with the same formula resulted.

Table 3-6. Potential formulas from $[M+Na]^+$ database search of masses with the highest FC.

m/z Searched	Formula	Theoretical *m/z*	Error (ppm)	ID
261.0885	$C_{16}H_{14}O_2$	261.0886	0	4'-Methoxychalcone
277.0626	$C_{11}H_{14}N_2O_3S$	277.0617	3	
485.3108	$C_{28}H_{46}O_3S$	485.3060	9	
	$C_{30}H_{42}N_2O_2$	485.3138	-6	Arachidonoyl Serotonin
485.3107	$C_{28}H_{46}O_3S$	485.3060	9	
	$C_{30}H_{42}N_2O_2$	485.3138	-6	Arachidonoyl Serotonin
513.3416	$C_{30}H_{50}O_3S$	513.3373	8	
537.2824	$C_{30}H_{42}O_7$	537.2823	0	
	$C_{31}H_{46}O_2S_2$	537.2831	-1	Probucol spiroquinone
583.3451	$C_{25}H_{48}N_6O_8$	583.3426	4	Deferoxamine

A mass accuracy of 10 ppm was required. ID was given if no other compounds with the same formula resulted.

Table 3-7. List of metabolites identified by accurate mass and MS/MS fragmentation from the PLS-DA influential mass list.

Metabolite	Experimental *m/z*	Theoretical *m/z*	Error (ppm)	Retention (min)
Dodecanoylcarnitine* $C_{19}H_{37}NO_4$	344.2799	344.2795	1.16	22.9
Myristoleylcarnitine $C_{21}H_{39}NO_4$	370.2961	370.2952	2.43	23.7
Myristoylcarnitine* $C_{21}H_{41}NO_4$	372.3111	372.3108	0.81	24.7
Palmitoleylcarnitine $C_{23}H_{43}NO_4$	398.3264	398.3265	-0.25	25.2
Linoleoylcarnitine $C_{25}H_{45}NO_4$	424.3417	424.3421	-0.94	25.6

*Indicates verification by standard

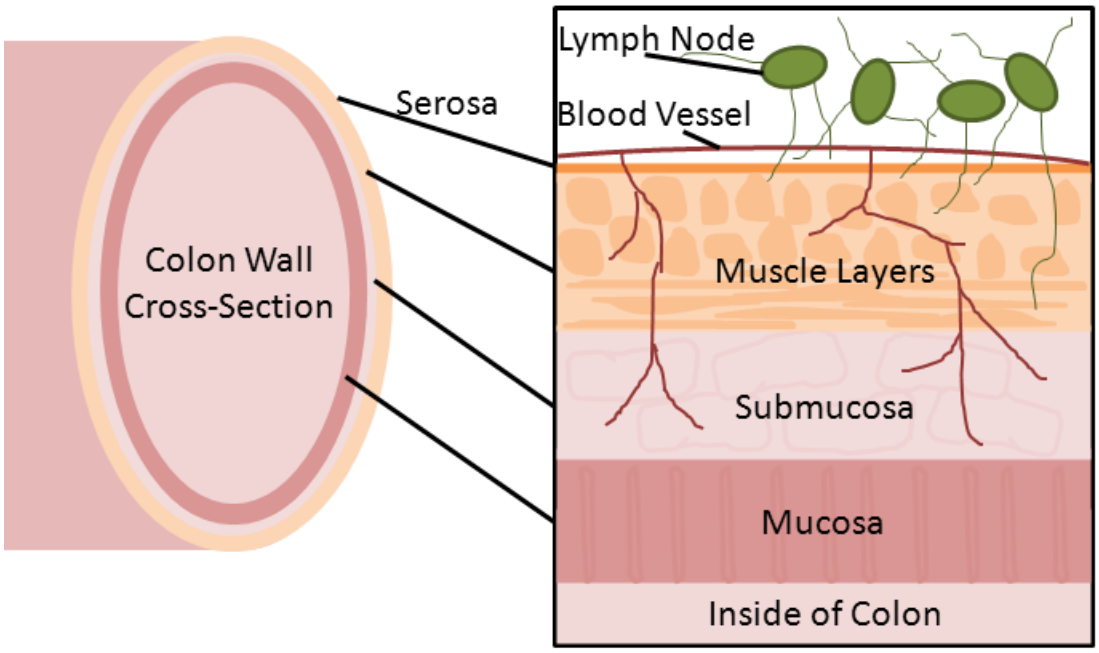


Figure 3-1. Areas affected by and used in staging colon cancer.[65]

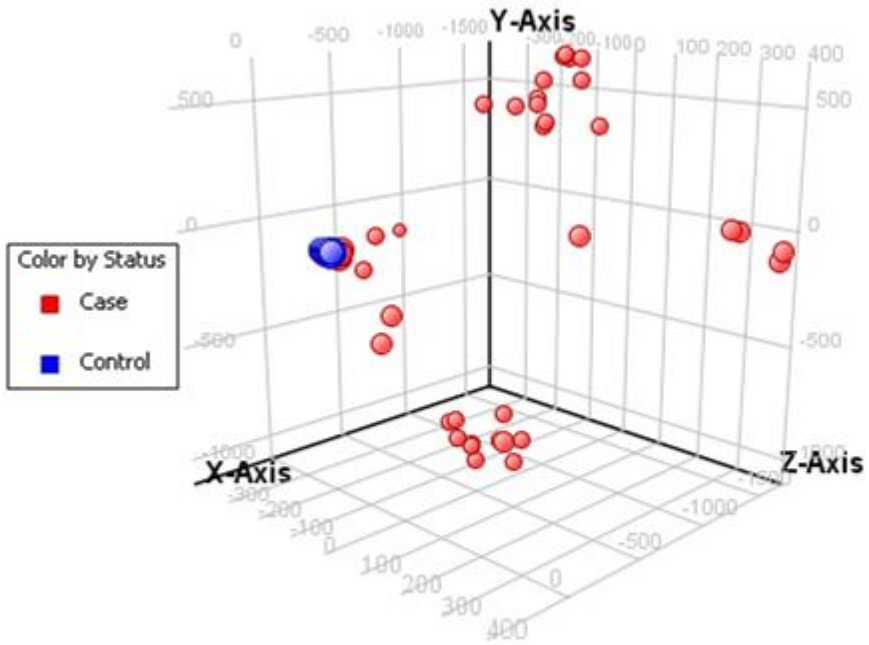


Figure 3-2. 3D-PCA plot of the significant masses in plasma from colon cancer patients and healthy controls (165 masses, p = 0.01). Each point represents one replicate.

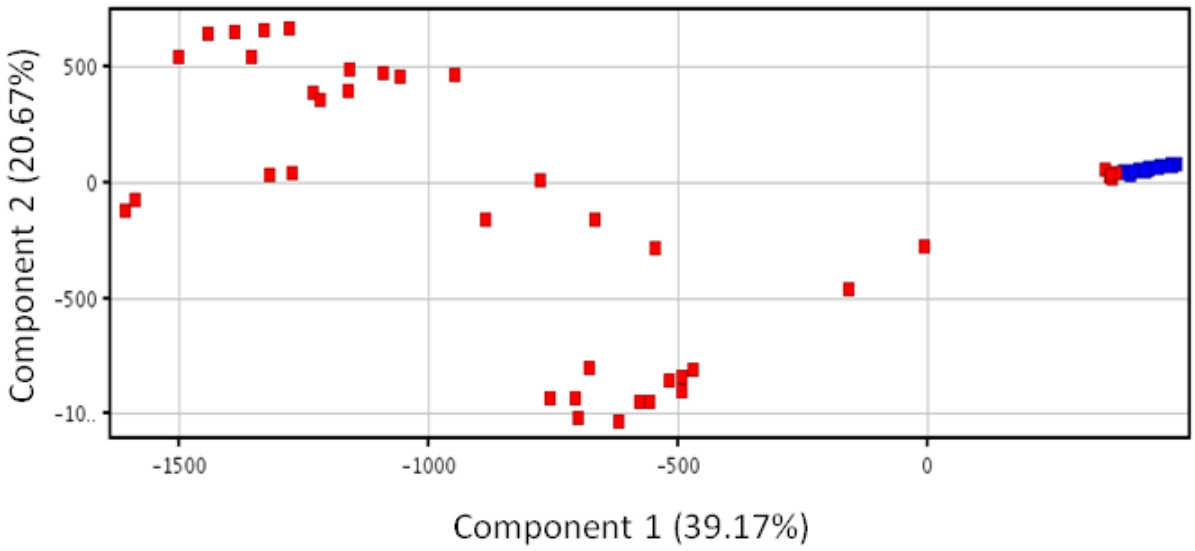


Figure 3-3. 2D-PCA plot of the significant masses in plasma from colon cancer patients and healthy controls (165 masses, $p = 0.01$). Each point represents one replicate. The color code is the same as Figure 3-3.

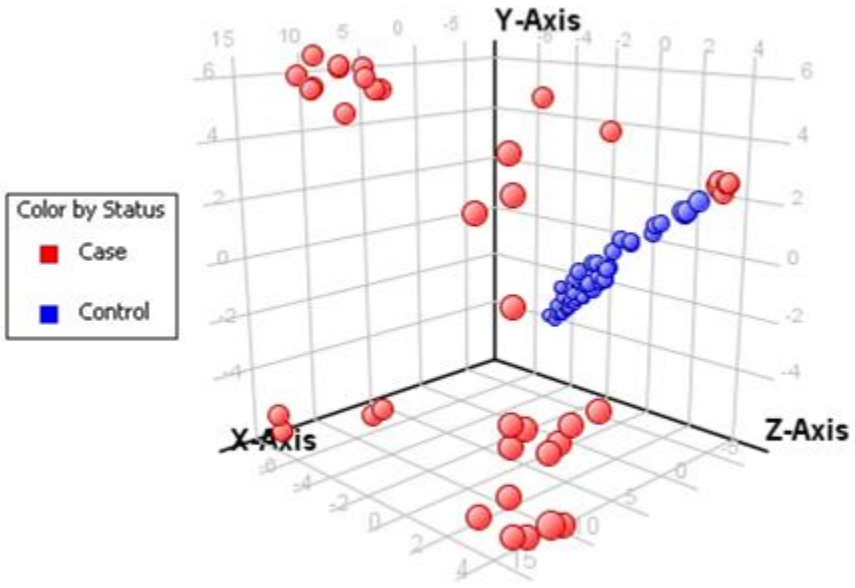


Figure 3-4. 3D-PLS-DA plot of *t*-test ($p = 0.01$) masses in human plasma from healthy controls and colon cancer patients (165 masses). Each point on the plot depicts one replicate.

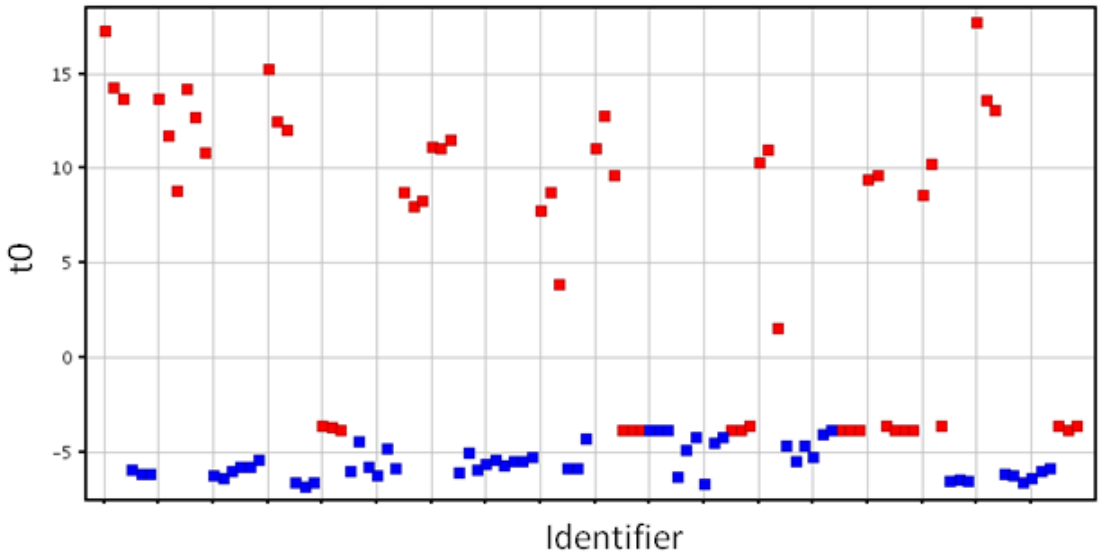


Figure 3-5. 2-PLS-DA plot of *t*-test (p = 0.01) masses in human plasma from healthy controls and colon cancer patients (165 masses). Each point on the plot depicts one replicate. The x-axis (Identifier) values are the sample replicates. The color code is the same as Figure 3-5.

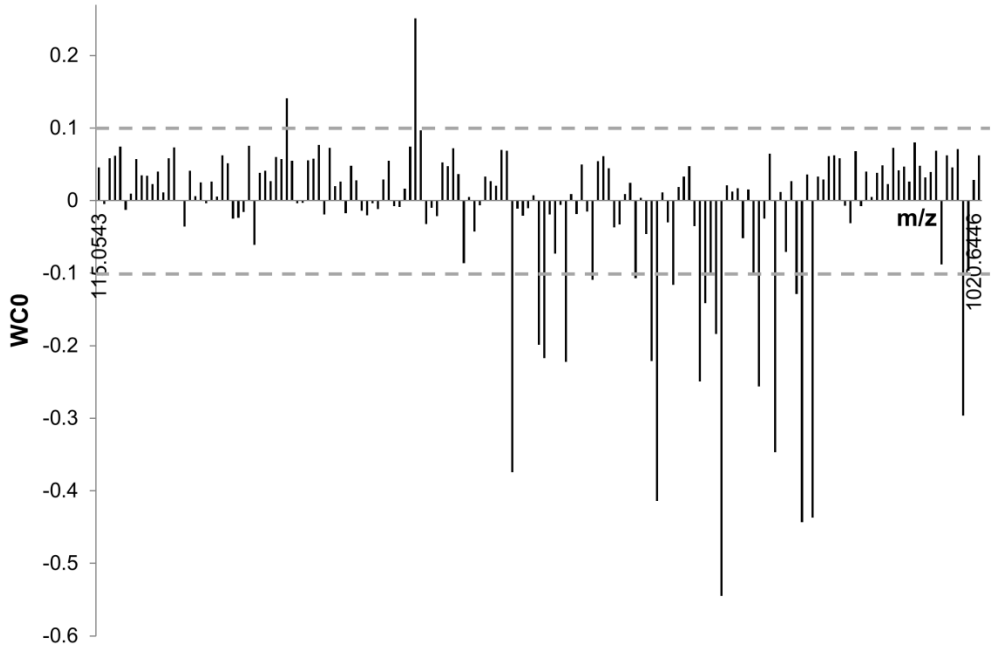


Figure 3-6. PLS-DA loadings plot of the significant masses determined by a *t*-test (165 masses, p = 0.01) in human plasma from healthy controls and colon cancer patients. WC0 indicates how influential a *m/z* value is (largest and smallest WC0 values are more influential in PLS-DA separation).

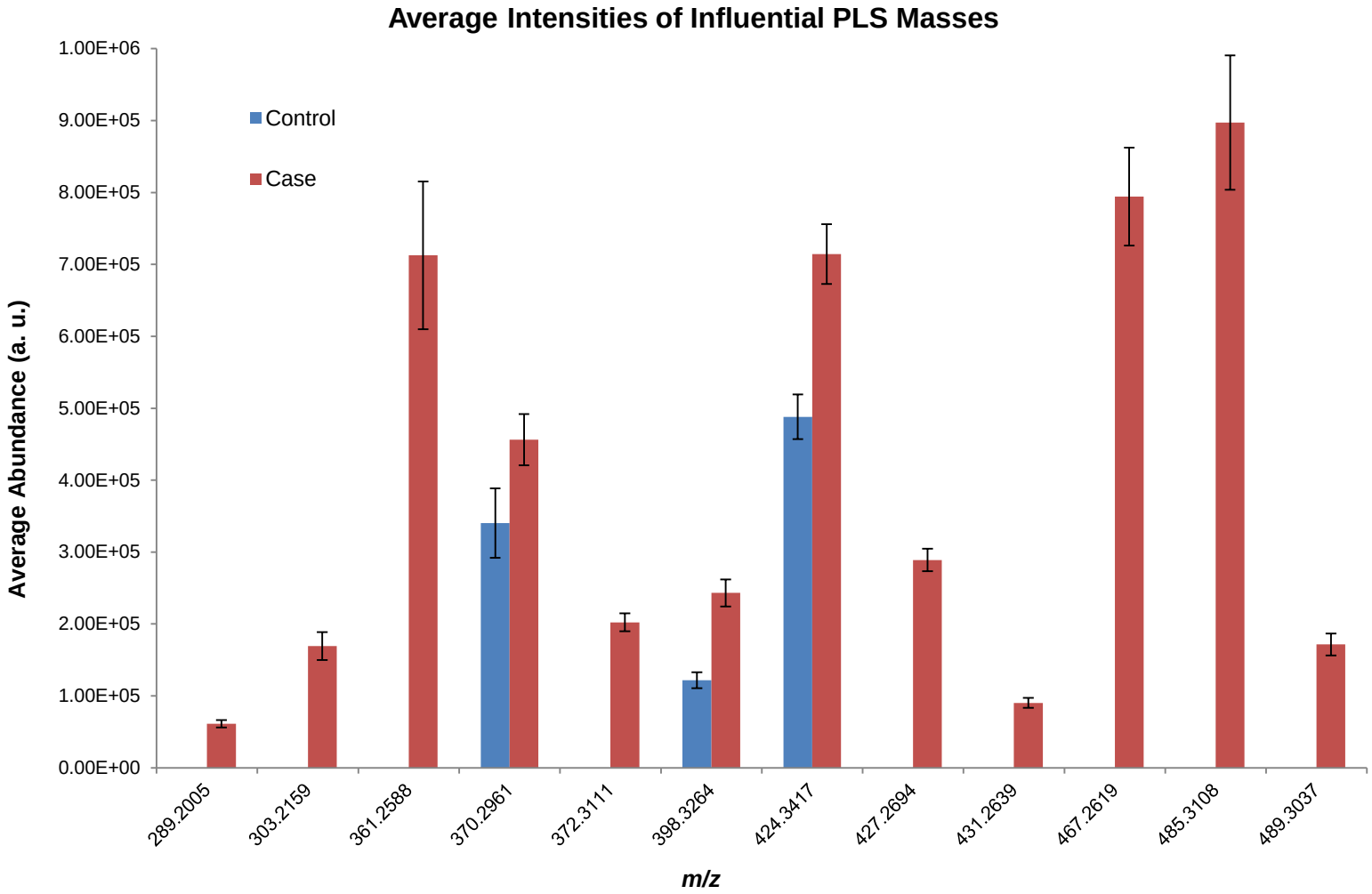


Figure 3-7. Bar graph of the average abundances of masses furthest from WC0 = 0 in the PLS loadings plot. A mass had to be present in 60% of the samples in one condition (case or control) to be considered. Error bars indicate the standard deviation of the mean. Metabolites identified were *m/z* 370.2961, 372.3111, 398.3264, and 424.3417 (all carnitines).

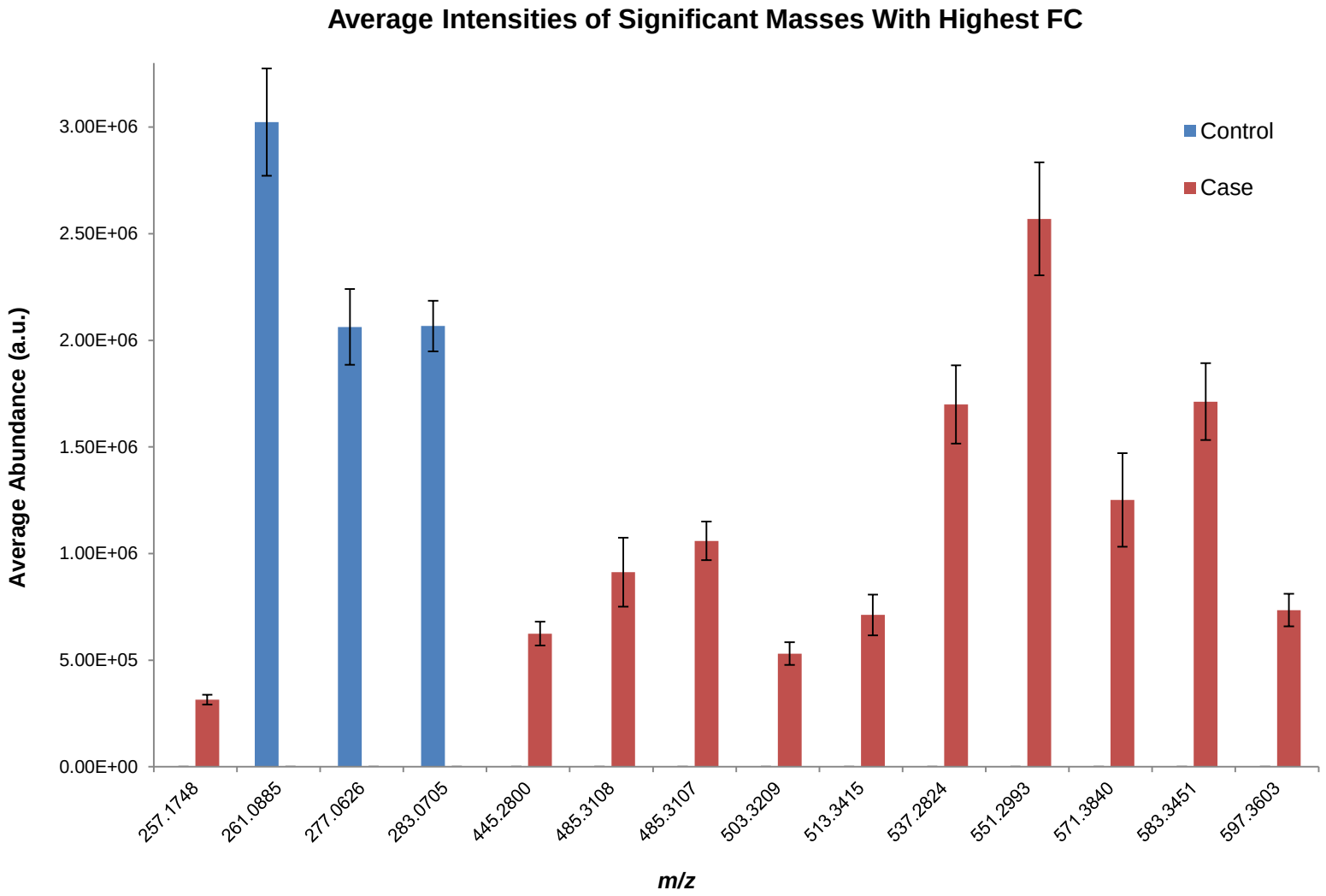


Figure 3-8. Bar graph of the average abundances of the significant masses with the highest fold change. A mass had to be present in 60% of the samples in one condition (case or control) to be considered. Error bars indicate the standard deviation of the mean.

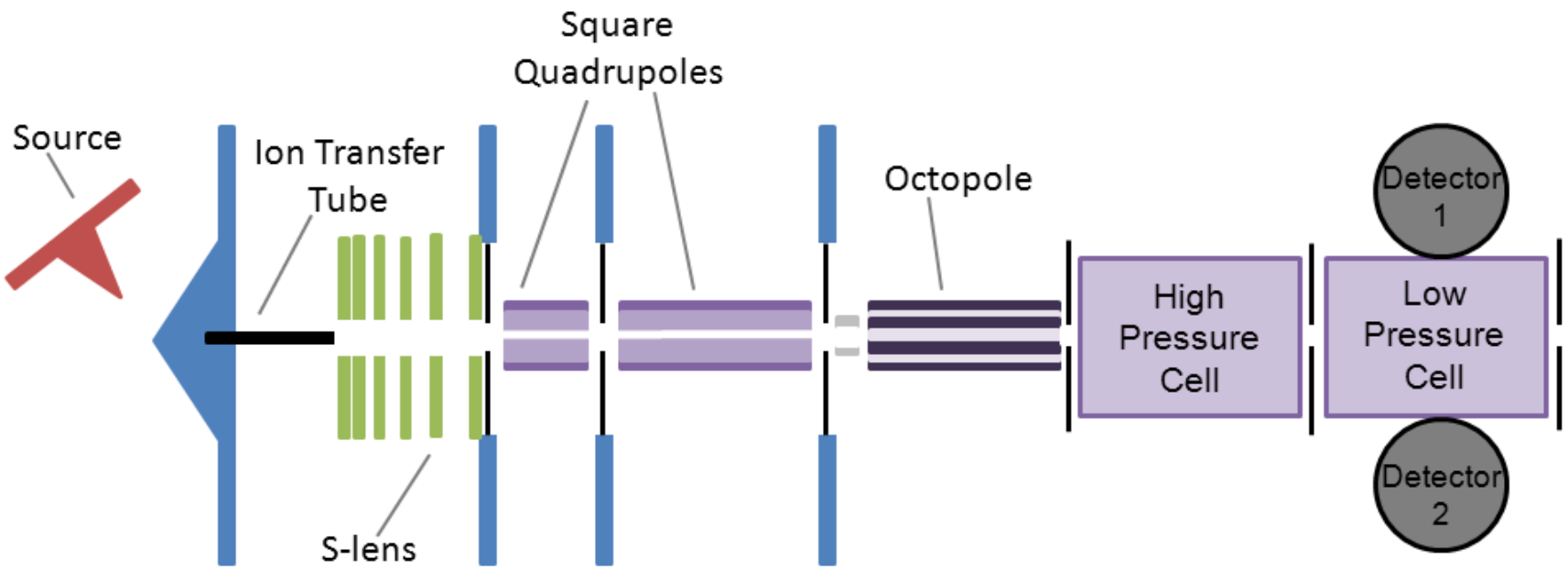


Figure 3-9. Schematic of the Velos dual pressure LIT.[74]

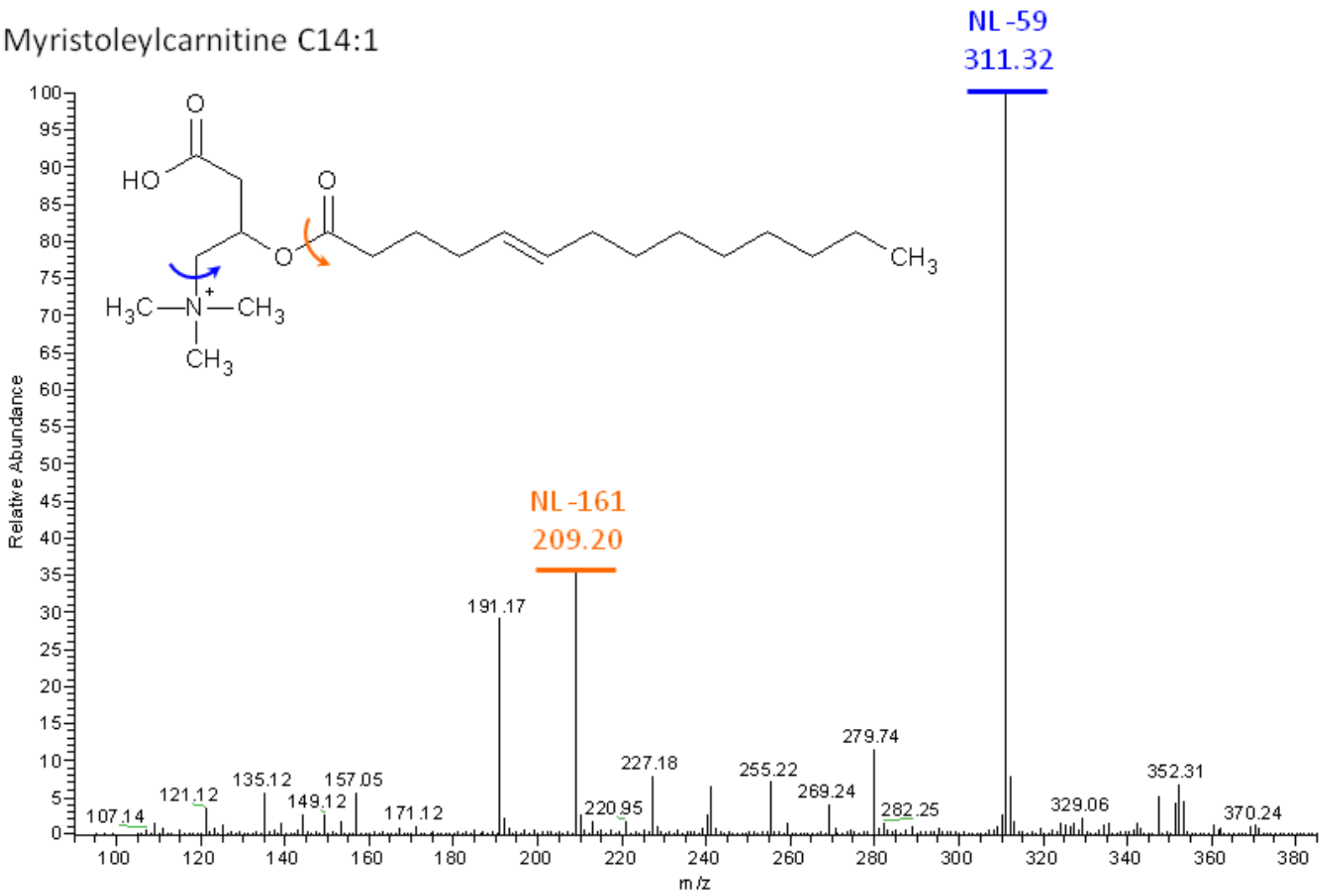


Figure 3-10. MS/MS spectrum of *m/z* 370 in plasma from a colon cancer patient. The neutral losses (NL) of 59 and 161 are due to the loss of the trimethyl amine (below blue arrow) and the loss of carnitine (left of orange arrow), respectively.

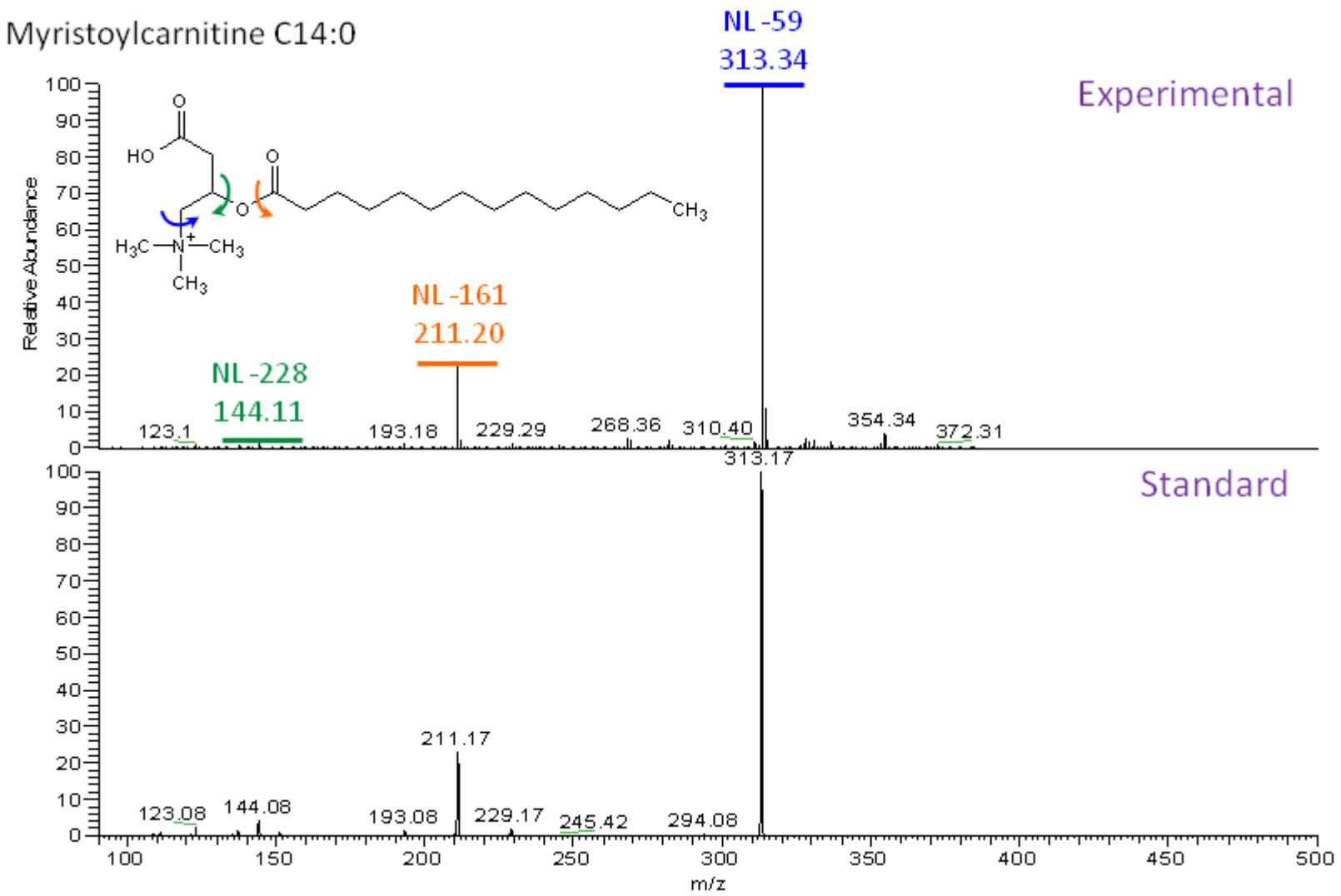


Figure 3-11. MS/MS spectra of *m/z* 372 in plasma from a colon cancer patient (top) and a standard (bottom). The neutral losses (NL) of 59 and 161 are due to the loss of the trimethyl amine (below blue arrow) and the loss of carnitine (left of orange arrow), respectively. The ion at *m/z* 144 is indicative of dehydrated carnitine (left of green arrow).

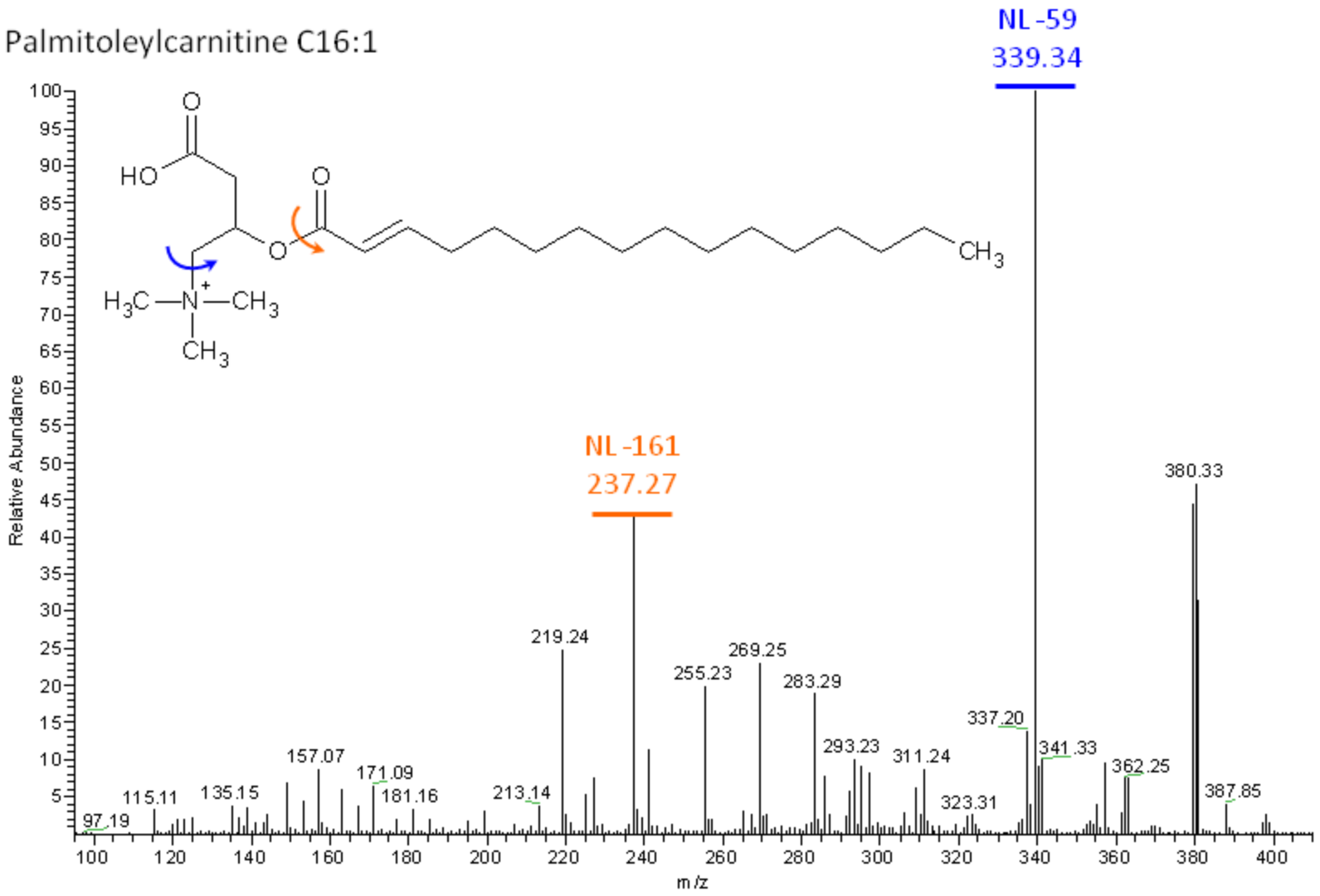


Figure 3-12. MS/MS spectrum of *m/z* 398 in plasma from a colon cancer patient. The neutral losses (NL) of 59 and 161 are due to the loss of the trimethyl amine (below blue arrow) and the loss of carnitine (left of orange arrow), respectively.

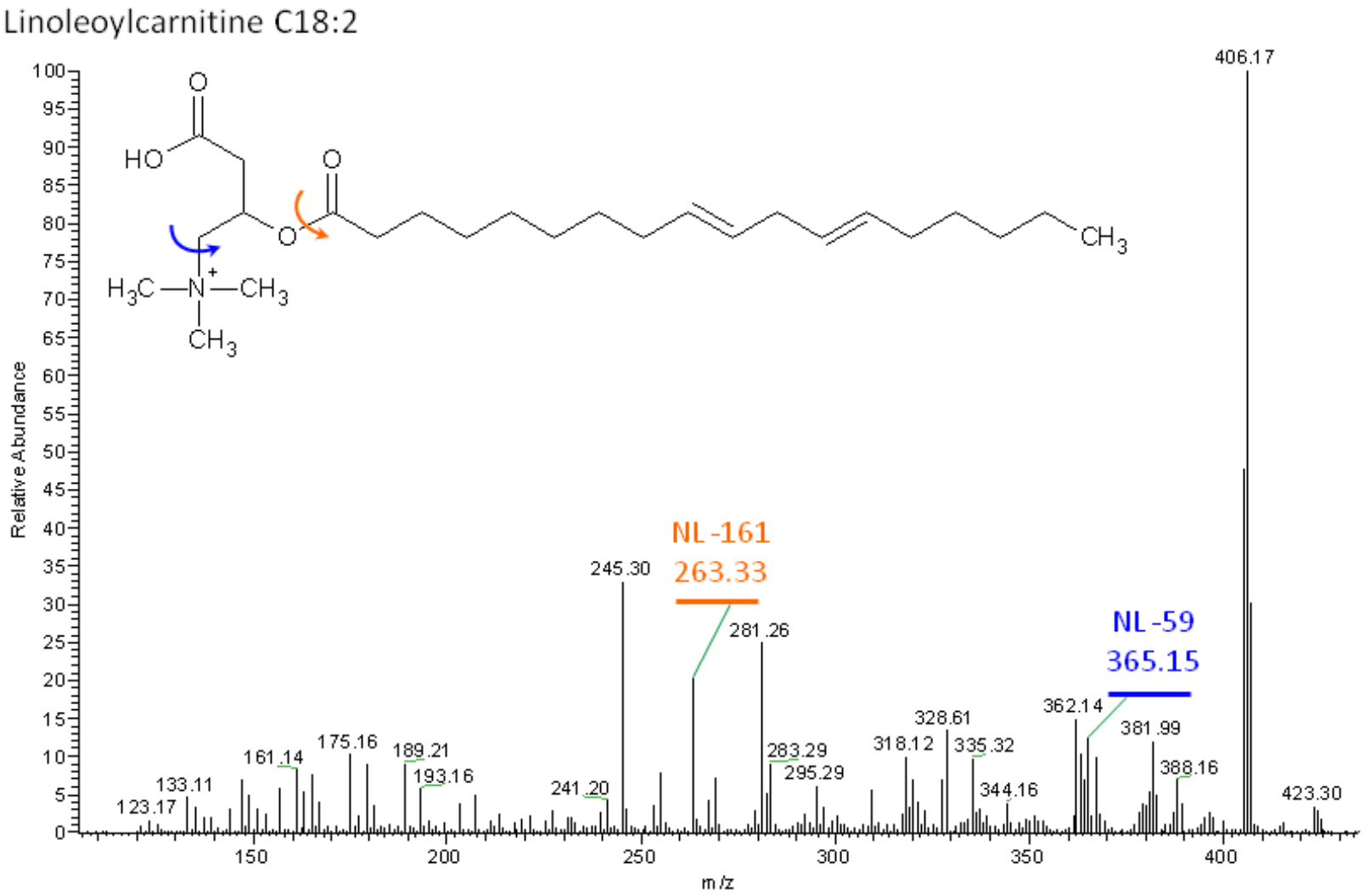


Figure 3-13. MS/MS spectrum of *m/z* 424 in plasma from a colon cancer patient. The neutral losses (NL) of 59 and 161 are due to the loss of the trimethyl amine (below blue arrow) and the loss of carnitine (left of orange arrow), respectively.

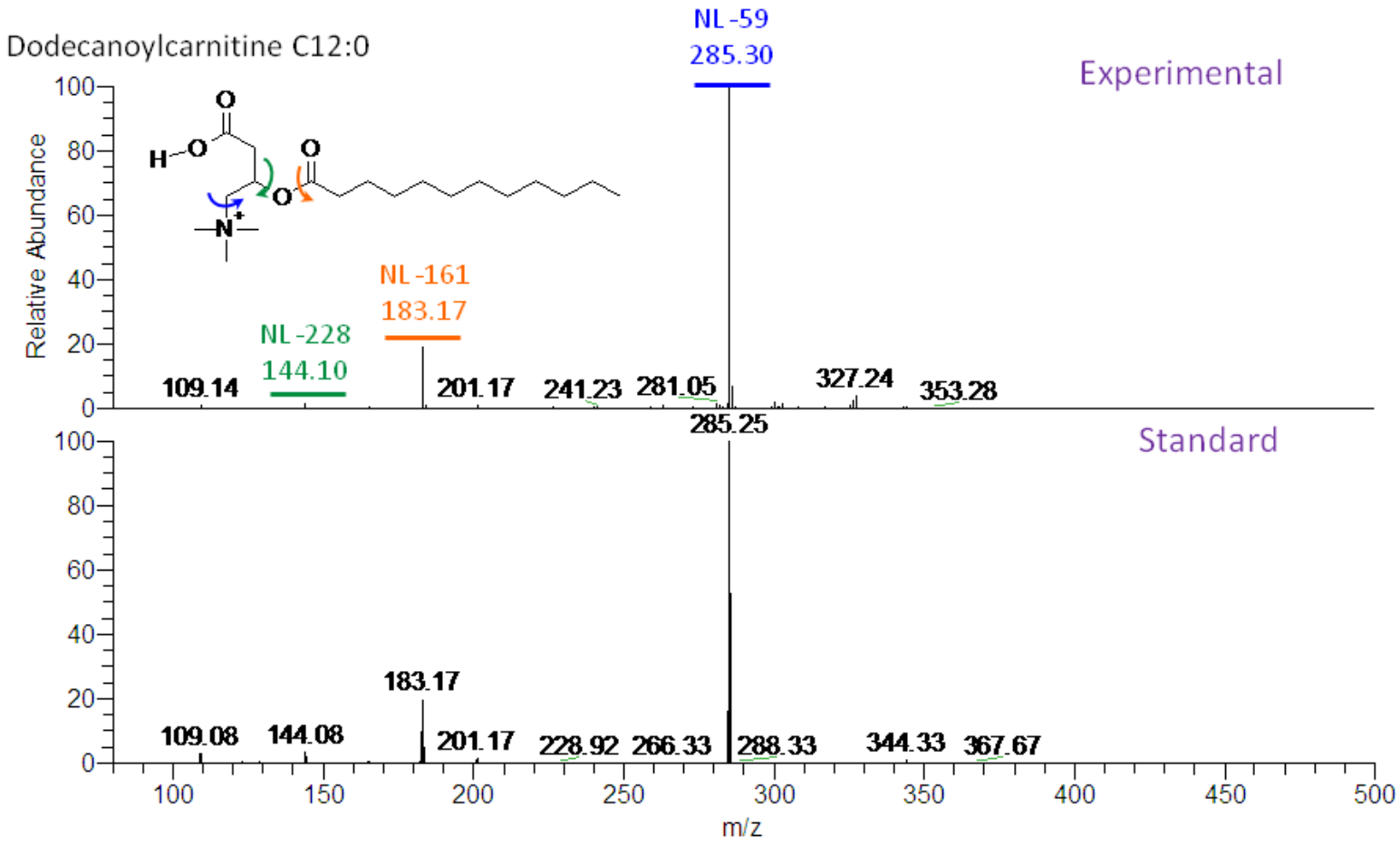


Figure 3-14. MS/MS spectra of *m/z* 344 in plasma from a colon cancer patient (top) and a standard (bottom). The neutral losses (NL) of 59 and 161 are due to the loss of the trimethyl amine (below blue arrow) and the loss of carnitine (left of orange arrow), respectively. The ion at *m/z* 144 is indicative of dehydrated carnitine (left of green arrow).

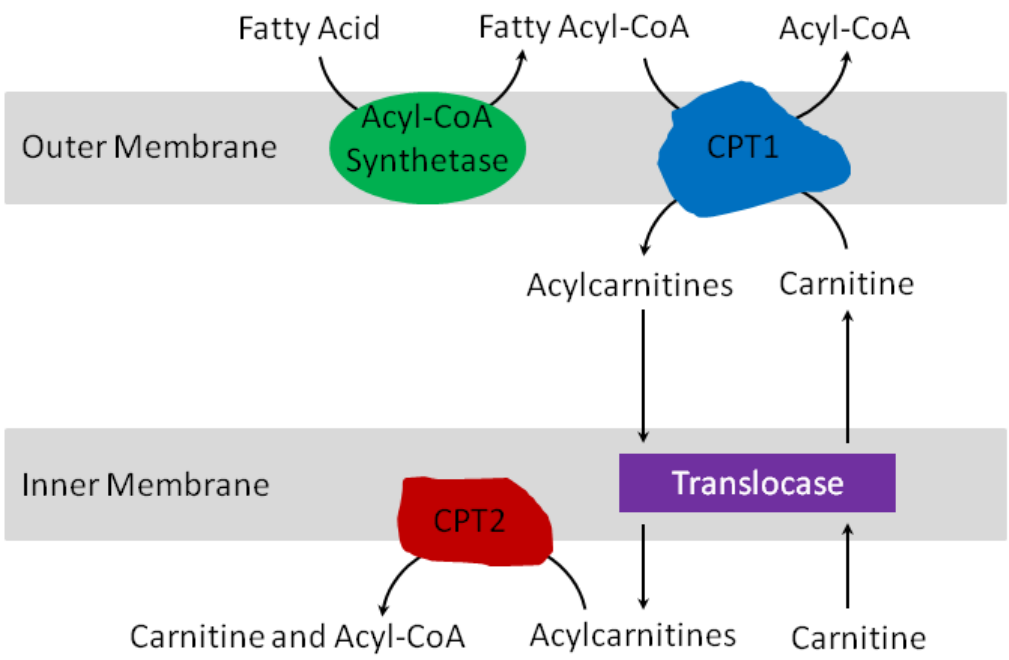


Figure 3-15. Carnitine palmitoyltransferase (CPT) enzyme activity in the mitochondria.[76]

CHAPTER 4
METHOD OPTIMIZATION: TOWARDS MORE EFFICIENT IONIZATION AND SHORTER ANALYSIS TIME

Overview

In metabolomic experiments it is important to be able to ionize and detect as many compounds as possible. Compounds that are not detected could still be potential biomarkers. Frequently, researchers will run experiments in both positive and negative modes as well as with ESI and atmospheric pressure chemical ionization (APCI) sources to broaden the types of compounds detected.[9, 36, 79] Each of these ionization techniques should be optimized separately. Efficient ionization is often achieved by good chromatographic separation prior to ionization; this prevents or reduces ion suppression. Although chromatography is necessary, it should be timely to allow all samples to be analyzed together, instead of in multiple batches, where instrument variability is increased. In particular, shared instruments have limited free time so shorter chromatographic run times are helpful. It is also more economical since solvent consumption is reduced. This chapter concentrates on the optimization of positive ESI as well as different LC time gradient lengths and LC columns to be coupled to the source.

One technique that is used in this work, in order to compare the various optimization conditions and LC methods, is the extraction and comparison of features detected with LC/MS analyses. This technique is also used in the lab of one of the leaders in metabolomic research, Gary Siuzdak, at The Scripps Research Institute.[9, 80-81] Metabolomics software packages typically begin data analysis by extracting all features (or *m/z* and retention time pairs). These features can therefore be used to represent the number of different components that were able to be detected by the

mass spectrometer. Thus, one can compare the number of features from one method to another to determine the conditions that allow the highest number of compounds to be detected. As previously mentioned, in global metabolomics it is desirable to detect the largest number of compounds. Because of this, conditions that achieve the highest number of features were implemented in this work.

Finally, a hydrophilic interaction liquid chromatography (HILIC) column was optimized. As can be seen by the chromatograms in Figures 5-1 and 5-2, urine contains many polar analytes that are not well retained with RP chromatography. Due to the nature of HILIC, the elution order of analytes is essentially reversed, allowing polar compounds to be better retained, and in return non-polar compounds are less well retained.[82] Sometimes both HILIC and RP chromatography are implemented in order to achieve complimentary results.[83-84] The optimized HILIC method was later implemented for the separation of manatee urine prior to mass spectrometric detection (Chapter 5).

Experimental

Chromatography Methods

Three chromatography methods were used for ionization optimization and feature extraction of urine and plasma samples. Feature extractions were done to determine the parameters that led to the highest number of features being detected. The three methods are shown in Table 4-1 and described below.The monolithic method uses two Phenomenex (Phenomenex, Torrance, CA) monolithic columns and a 50 minute gradient elution analysis. The short monolithic method uses similar conditions and the same columns but the analysis time is only 27 minutes. The final chromatography method, the Zorbax method, uses an Agilent Zorbax column (Agilent Technologies, Santa Clara, CA). The Zorbax column is a small particle column, as opposed to a

monolithic column, and is more frequently used with fast chromatography. The Zorbax method is a 25 minute analysis.

Four different methods were also used to compare the Zorbax and monolithic columns to determine if one was more chromatographically efficient than the other. As described in Table 4-2, each column was implemented under its normal conditions and then a second method was created to mimic the other columns conditions. Feature extractions were compared for piglet plasma analyzed by each of the methods.

Finally, work was done to optimize the solvents used with a HILIC column. Four different solvent systems were implemented with the HILIC column and amino acids and carnitines were examined for chromatographic efficiency and retention. The four methods are shown in Table 4-3. The column implemented was a Halo Penta-HILIC (Advanced Materials Technology, Inc., Wilmington, DE). Similar to the Zorbax method, the HILIC column used a 25 minute time analysis, allowing for a more time efficient analysis of large sample sets.

Electrospray Source

The conditions used are the same as those listed in Chapter 3 or are otherwise stated (Capillary Voltage and Fragmentation Voltage sections).

Mass Spectrometer

The mass spectrometer conditions are identical to those of Chapter 2 and Chapter 3.

Capillary Voltage

The capillary voltage is applied to the front end of the capillary (Vcap) at the entrance of the MS, as shown in Figure 1-3. For the Agilent TOF MS, common capillary setting are between 3000-4000 V.

The fragmentor and all other settings were held constant while the capillary was varied between 3000 and 4200 V. Carnitine standards were used to determine the optimal capillary voltage. The standards included: carnitine (C0), hexanoylcarnitine (C6), octanoylcarnitine (C8), lauroylcarnitine (C12), myristoylcarnitine (C14), and stearoylcarnitine (C18). The short monolithic method was used for this experiment (Table 4-1). The average peak area (n = 3) of the extracted ion chromatograms (EIC) was determined and plotted as a function of the capillary voltage.

Fragmentation Voltage

The fragmentor voltage is applied at the back end of the capillary (fragmentor), as shown in Figure 1-3. For the Agilent TOF MS, common fragmentor voltages are 120-180 V.

For this study the capillary and all other settings were held constant while the fragmentor was varied between 70 and 240 V. Carnitine standards of carnitine (C0), hexanoylcarnitine (C6), octanoylcarnitine (C8), decanoylcarnitine (C10), lauroylcarnitine (C12), and myristoylcarnitine (C14) were injected (n = 3) and the Zorbax method was implemented (Table 4-1). The average peak area from the EIC was plotted as a function of the fragmentor voltage.

Feature Extraction of Urine Samples

Manatee urine was analyzed in triplicate to find the optimal conditions to produce the largest number of features in urine. Conditions that were considered were different capillary/fragmentor voltages as well as a longer and shorter LC time gradient. This was done by using both the monolithic and short monolithic methods (Table 4-1) and capillary/fragmentor settings of 3000 V/120 V, 3000 V/180 V, 4000 V/120 V, 4000 V/180 V, and 3600 V/150 V for each of the LC methods. The features were extracted using

Mass Profiler Professional (Agilent Technologies) from each data file and averaged for each group (n = 3).

ANOVA was implemented to determine if there was a significant difference between the average number of features detected with the short monolithic method compared to the monolithic method.

Feature Extraction of Plasma Samples

Piglet plasma was analyzed in triplicate to find the optimal conditions to produce the largest number of features in plasma. Conditions that were considered were different capillary/fragmentor voltages as well as a longer and shorter LC time gradient. This was done by using both the monolithic and short monolithic methods (Table 4-1) and capillary/fragmentor settings of 3000 V/120 V, 4000 V/120 V, 4000 V/180 V, and 3600 V/150 V for each of the LC methods. The features were extracted using MPP from each data file and averaged for each group (n = 3).

ANOVA was also used to determine if a significant difference existed between the average number of features detected with the short monolithic method compared to the monolithic method.

In order to compare the methods further, five different compounds were also evaluated in detail. The EIC for each *m/z* value were used to determine the retention times (t_r) and base peak widths (w). These values were then used to calculate the column efficiency (N). This was performed for the capillary/fragmentor settings of 4000 V/120 V only.

Comparison of Monolithic and Small Particle Columns

In order to determine if a different column would be more efficient, the monolithic column was compared to a small particle Zorbax column. Table 4-2 shows the four

different LC methods that were used with the two different columns.[79] In addition to each individual method (MC and ZC methods) another method was made that attempts to use similar conditions for a fair comparison (MC with ZC conditions and ZC with MC condtions methods). Piglet plasma was injected in triplicate for each of these methods. Capillary and fragmentor voltages were maintained at 4000 V and 120 V, respectively. It is important to note that since the Zorbax column contains small particles it was always heated to 60 ˚C to maintain a pressure within the LC's limits.

The number of features for each analysis was determined with MPP and the same five compounds were explored as in the Feature Extraction of Plasma Samples section. The EICs for each *m/z* value were used to determine the average retention times (t_r) and base peak widths (w). These values were then used to calculate the column efficiency (N). In addition to the five compounds chosen at random, standard mixtures of carnitines were analyzed using each LC method and the same calculations were performed. Carnitines that were in the mixture include: carnitine (C0), hexanoylcarnitine (C6), octanoylcarnitine (C8), lauroylcarnitine (C12), myristoylcarnitine (C14), and stearoylcarnitine (C18).

Evaluation of Different Mobile Phases for use with the Zorbax Column

Different mobile phase compositions were explored for uses with the Zorbax column method. The LC method used was the ZC method shown in Table 4-2. The solvents were the only condition changed for this study. Three different pairs of mobile phases were used. The first set of mobile phases analyzed were the same as those listed on Table 4-2. The second set of mobile phases used water (H_2O) as mobile phase A and acetonitrile (ACN) as mobile phase B, the amount of AA added to each (A and B) was 0.1% (v/v) instead of 1% (v/v) used in the first set. The final set of mobile

phases used water as mobile phase A and methanol (MeOH) as mobile phase B, the amount of AA added to each (A and B) was 0.2% (v/v).[79] These three different LC solvent sets were used with the Zorbax column to analyze piglet plasma in triplicate. The number of features were determined with the MPP software and averaged for the three replicates in each solvent set. The same five compounds that were explored as in the Feature Extraction of Plasma Samples section were examined. The EICs for each *m/z* value were used to determine the average retention times (t_r) and base peak widths (w). These values were then used to calculate the column efficiency (N).

Optimization of Solvent Systems for use with HILIC

Standards of carnitine (C0), hexanoylcarnitine (C6), octanoylcarnitine (C8), decanoylcarnitine (C10), lauroylcarnitine (C12), myristoylcarnitine (C14), L-anserine, L-arginine, L-carnosine, creatine, L-histidine, L-homocystine, gamma-DL-hydroxylysine, L-lysine, 1-methyl-L-histidine, 3-methyl-L-histidine, L-ornithine and L-tryptophan were examined by HILIC-ESI/TOF-MS with four different solvent systems. The same HILIC column, flow rate, column temperature, and gradient were used, only the solvents differed. The four solvent systems are described in Table 4-3. The first two solvent systems had 5 mM ammonium formate (AmF) in each mobile phase A and B and were at pH 3.75. Both AmF methods used the same mobile phase B, 50% water and 50% acetonitrile with 5 mM AmF. The first method, named AmF 95:5 had mobile phase A of 95% acetonitrile and 5% water with 5 mM AmF. The second method (AmF 90:10) only differed from the first by mobile phase A, which was 90% acetonitrile and 10% water with 5 mM AmF. The second two solvent systems used with HILIC had 5 mM ammonium acetate (AmA) in each mobile phase A and B and were at pH 5.75. Both AmA methods used the same mobile phase B, 50% water and 50% acetonitrile with 5

mM AmA. The first AmA method, named AmA 95:5, had mobile phase A of 95% acetonitrile and 5% water with 5 mM AmA. The second AmA method, called AmA 90:10, differed from the previous AmA method by having a mobile phase A of 90% acetonitrile and 10% water with 5 mM AmA. The gradient used for all methods held 100% mobile phase A for three minutes and then changed to 90% mobile phase B over 17 minutes and then immediately headed back to 100% mobile phase A over three minutes and held 100% mobile phase A for 2 minutes for a total analysis time of 25 minutes.

The 18 standards were mixed together and injected in triplicate with each solvent system. The peak areas for each analyte were determined from the EIC of each standard's $[M+H]^+$ ion. The average retention time and peak widths were also noted so that the efficiency (N) could be calculated for each analyte.

Results and Discussion

Capillary and Fragmentor Voltages

In general, the carnitine standards appeared to have a maximum peak area at a capillary voltage of 3000 V and a fragmentor voltage of 120 V (Figure 4-1 and Figure 4-2). It is clear that a lower fragmentor voltage would promote more efficient ionization, as it would be gentler on the ions and less likely to break them. However, it is less apparent why a lower capillary voltage would lead to more ionization. It seems more likely that a higher fragmentation voltage would lead to more ionization.

While this information is useful, carnitines are not the only compounds of interest in metabolomic experiments. Different types of compounds may not ionize efficiently under these types of conditions. Thus, this study was extended to different biological fluids in order to determine what conditions produce the largest number of features.

Feature Extraction of Biofluids

Table 4-4 depicts the average number of features (three replicates) detected in manatee urine for each chromatography method and capillary/fragmentor setting. The short monolithic method and capillary/fragmentor voltages of 4000 V/120 V yield the largest average number of features in manatee urine. The piglet plasma results can be seen in Table 4-5, shown as the average number of features (three replicates) detected for each chromatography method and capillary/fragmentor setting. The short monolithic method and capillary/fragmentor voltages of 4000 V/120 V yield the largest average number of features in piglet plasma. This verifies that a shorter time gradient can be implemented without losing features when working with plasma or urine. ANOVA verified that no significant differences existed between the monolithic method and short monolithic method in both piglet plasma and manatee urine.

The column efficiencies in the monolithic and short monolithic methods were calculated and compared for the piglet plasma analysis. This was done to verify the chromatographic efficiency did not suffer too extensively with the shorter time gradient. Table 4-6 lists the column efficiencies for five *m/z*'s chosen at random from the piglet plasma analyses with the monolithic method. The same calculations were performed for the piglet plasma analyses with the short monolithic method and listed in Table 4-7. These results show little change in the column efficiency when the time gradient is reduced. This is likely because the peak widths are also decreasing with the shorter gradient, maintaining the efficiency. The short monolithic method can thus be used for metabolomic experiments with capillary and fragmentor voltages of 4000 V and 120 V, respectively.

Implementing Different LC Columns

To decrease the time length of a gradient and also maintain chromatographic efficiency it is common to either increase the pressure or temperature or to use different types of particles.[23] For these reasons we compared the monolithic column to a column that utilizes those properties (Zorbax column). The results of the feature extraction for this study can be seen in Table 4-8. The numbers of features detected for each of the three replicates of piglet plasma, analyzed with four LC methods, were averaged. As noted in Table 4-8 the largest number of features was observed while the ZC method was implemented.

The carnitine standard mixture produced resolved chromatographic peaks for all four LC methods, shown in Figure 4-3. Addtionally, Tables 4-9 to 4-12 list retention times, peak widths and the calculated plate heights of the carnitines for each of the four methods. The Zorbax methods allowed for smaller plate heights (except in the first two minutes of the analysis) than the methods that equipped the monolithic column. The two Zorbax column methods also allowed for sharper peaks than the two monolithic column methods. This can be seen by comparing the peak widths shown in Tables 4-9 to 4-12.

The chromatograms for individual piglet plasma injections, under each of the four methods, are shown in Figure 4-4. Much sharper peaks are observed for the two Zorbax column methods than for the monolithic column methods. Tables 4-13 to 4-16 list the retention times, peak widths, and calculated plate heights for five randomly chosen compounds. Similar trends were observed for the carnitine standards as were seen with the five random masses. Those trends include smaller peak widths and plate heights for the methods that use the Zorbax column.

When considering the number of features found for each of the four LC methods, as well as the chromatography results for both carnitine standards and piglet plasma, it is obvious that the ZC methods allows for efficient chromatography and the detection of the largest number of features.

Use of Different Solvents with Zorbax Column

The results of the feature extraction from the three solvent sets tested are shown in Table 4-17. The largest average number of features was observed for the third solvent set that implemented water (A) and methanol (B) with 0.2% AA (v/v) in each. For the two solvent sets that used acetonitrile instead of methanol, the method that used 0.1% AA instead of 1% AA (v/v) resulted in more extracted features. The chromatographic results are shown in Tables 4-18 to 4-23.

Tables 4-18 to 4-20 are the chromatographic results of carnitine standard mixtures for each of the three solvent sets. The most column efficiency is observed for the method using water and methanol with 0.2 % AA (Table 4-20). This method (0.2% AA) also has longer retention times than the two methods using ACN, adding to the efficiency of the method.

The chromatographic results of five randomly chosen compounds are shown in Tables 4-21 to 4-23 for each of the solvent systems. Other than the first two *m/z*'s listed in Table 4-23, a much higher column efficiency was observed for the method using methanol and 0.2% AA. The two methods using ACN yielded similar column efficiencies but in general, much lower than the method implementing methanol and 0.2% AA.

Figures 4-5 and 4-6 show the chromatograms for carnitine mixtures and piglet plasma, respectively. Similar chromatographic profiles are observed with ACN with 1%

AA and ACN with 0.1% AA for both the carnitine standard mixtures and the piglet plasma. However, the use of methanol and 0.2% AA shows a shift to longer retention times in the carnitine standard mixture and the piglet plasma sample.

Increased column efficiency for the carnitine standard mixture as well as the piglet plasma was observed with the solvent system utilizing methanol and 0.2% AA. Therefore, the solvent system using methanol and 0.2% AA will be implanted in metabolomic studies. This will allow for efficient chromatography and more features to be determined based on the work in this section.

Optimization of HILIC

Analysis of the standard mixture resulted in higher efficiencies for the two solvent systems that used ammonium acetate. The efficiencies observed for carnitine standards, which experienced much lower efficiencies than the amino acids, are shown in Figure 4-7. C0 was the only carnitine to be retained with solvents containing ammonium formate, explaining the low efficiencies obtained. The carnitines were retained better with the ammonium acetate solvents, but still had lower efficiencies than most of the amino acids since the retention was not as long for most of the carnitines as it was for the amino acids. The efficiencies calculated for amino acid standards are depicted in the bar graph in Figure 4-8. Most of the efficiencies for the amino acids with ammonium acetate solvents were three times that of the carnitines despite what method was used. Figure 4-9 shows the retention of all the standards with each of the four solvent systems explored. For each analyte, an ammonium acetate solvent system led to the most retention and the highest efficiency. More frequently, the method labeled AmA 95:5 had more efficiency than AmA 90:10 and will be used in future studies of urine.

Summary

The work in this chapter has shown that a capillary/fragmentor voltage of 4000 V/120 V is optimal when studying the metabolomics of urine or plasma. It was also shown that a shorter time gradient can be implemented with the monolithic column, for urine and plasma, while maintaining efficient chromatography and not losing any molecular features. Additionally, more efficient chromatography (smaller plate heights and peak widths) is seen when studying plasma with a Zorbax column, as compared to a monolithic column. Also, a solvent system utilizing methanol and 0.2% AA allowed for more efficient chromatography in addition to the detection of more molecular features than with acetonitrile and 1% AA or 0.1% AA. Finally, experiments with HILIC using a Halo column resulted in optimal retention and efficiency with solvents containing ammonium acetate at pH 5.75. Particularly, any work with polar analytes would benefit from analysis with the AmA 95:5 HILIC method, since more analytes had the highest efficiency with this method than the other three methods.

In conclusion, future metabolomics work will be performed with capillary/fragmentor voltages of 4000 and 120 V, respectively. A 25 minute LC method will be used that implements the Zorbax column. The LC method will use solvents containing 0.2% AA in water (A) and methanol (B). The metabolomic study of manatee urine, discussed in Chapter 5, used this method and the Halo column for HILIC.

Table 4-1. Chromatography methods used to optimize ESI settings for standards and biofluids.

Method	Column	Column Dimensions (mm)	Mobile Phase A	Mobile Phase B	Flow Rate (mL/min)	Gradient [A:B(min)]
Monolithic	Phenomenex Onyx C18 Monolithic (2)	4.6 x 100	1% acetic acid (AA) in water	1% AA in acetonitrile	0.8	95:5(0.0-6.5) 0:100(25.5-35.0) 95:5(45.0-50.0)
Short Monolithic	Phenomenex Onyx C18 Monolithic (2)	4.6 x 100	1% AA in water	1% AA in acetonitrile	0.8	95:5(0.0-4.0) 0:100(13.0-17.0) 95:5(24.0-27.0)
Zorbax	Agilent Zorbax SB-Aq and SB-C8 guard	2.1 x 50 (1.8 µm particles) and 2.1 x 30 (3.5 µm particles)	0.2% AA in water	0.2% AA in methanol	0.35	95:5(0.0-13.0) 0:100(13.0-19.2) 95:5(24.0-25.0)

Table 4-2. Chromatography methods used to compare the monolithic and small particle column.

Method	Column	Column Dimensions (mm)	Mobile Phase A	Mobile Phase B	Flow Rate (mL/min)	Gradient [A:B(min)]
MC	Phenomenex Onyx C18 Monolithic (2)	4.6 x 100	1% acetic acid (AA) in water	1% AA in acetonitrile	0.8	95:5(0.0-4.0) 0:100(13.0-17.0) 95:5(24.0-27.0)
MC with ZC Conditions	Phenomenex Onyx C18 Monolithic (2)	4.6 x 100	1% AA in water	1% AA in acetonitrile	0.8	98:2(0.0) 2:98(13.0-19.0) 98:2(24.0-25.0)
ZC	Agilent Zorbax SB-Aq and SB-C8 guard	2.1 x 50 (1.8 µm), 2.1 x 30 mm (3.5 µm)	1% AA in water	1% AA in acetonitrile	0.6	98:2(0.0) 2:98(13.0-19.0) 98:2(24.0-25.0)
ZC with MC Conditions	Agilent Zorbax SB-Aq and SB-C8 guard	2.1 x 50 (1.8 µm), 2.1 x 30 mm (3.5 µm)	1% AA in water	1% AA in acetonitrile	0.6	95:5(0.0-4.0) 0:100(13.0-17.0) 95:5(24.0-27.0)

The column is heated to 60 ˚C in the ZC and ZC with MC Conditions methods. MC and ZC are abbreviated for monolithic column and Zorbax column, respectively.

Table 4-3. Chromatography methods used to optimize a HILIC method.

Method	Column	Column Dimensions (mm)	Mobile Phase A	Mobile Phase B	Flow Rate (mL/min)	Gradient [A:B(min)]
AmF 95:5	Halo Penta HILIC	2.1 x 100 (2.7 µm)	95% acetonitrile and 5% water with 5 mM AmF	50% acetonitrile and 50% water with 5 mM AmF	0.5	100:0(0.0-3.0) 10:90(20.0) 100:0(23.0-25.0)
AmF 90:10	Halo Penta HILIC	2.1 x 100 (2.7 µm)	90% acetonitrile and 10% water with 5 mM AmF	50% acetonitrile and 50% water with 5 mM AmF	0.5	100:0(0.0-3.0) 10:90(20.0) 100:0(23.0-25.0)
AmA 95:5	Halo Penta HILIC	2.1 x 100 (2.7 µm)	95% acetonitrile and 5% water with 5 mM AmA	50% acetonitrile and 50% water with 5 mM AmA	0.5	100:0(0.0-3.0) 10:90(20.0) 100:0(23.0-25.0)
AmA 90:10	Halo Penta HILIC	2.1 x 100 (2.7 µm)	90% acetonitrile and 10% water with 5 mM AmA	50% acetonitrile and 50% water with 5 mM AmA	0.5	100:0(0.0-3.0) 10:90(20.0) 100:0(23.0-25.0)

The column is heated to 35 ˚C in all methods. AmF and AmA are abbreviated for ammonium formate and ammonium acetate, respectively. All AmF methods were at pH of 3.75 and AmA methods were at pH 5.75.

Table 4-4. Average number of features (n =3) extracted from LC analysis with manatee urine by the monolithic and short monolithic methods at varying capillary/fragmentor voltages.

Capillary V/Fragmentor V	Average # of Features Monolithic Method	Average # of Features Short Monolithic Method
3000/120	121	129
3000/180	97	99
4000/120	120	138
4000/180	104	105
3600/150	115	129

Table 4-5. Average number of features (n =3) extracted from LC analysis with piglet plasma by the monolithic and short monolithic methods at varying capillary/fragmentor voltages.

Capillary V/Fragmentor V	Average # of Features Monolithic Method	Average # of Features Short Monolithic Method
3000/120	133	144
4000/120	137	154
4000/180	109	125
3600/150	131	141

Table 4-6. List of retention times, peak widths and column efficiencies for the piglet plasma analysis with the monolithic method (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	N
132.0639	4.072	0.007	0.54	0.02	917.87
205.0910	14.51	0.03	0.44	0.01	17741.34
304.3033	23.220	0.004	0.48	0.04	37028.60
332.3243	24.719	0.004	0.522	0.007	35879.04
360.3658	26.281	0.004	0.54	0.02	37664.18

Table 4-7. List of retention times, peak widths and column efficiencies for the piglet plasma analysis with the short monolithic method (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	N
132.0639	4.075	0.004	0.55	0.03	880.30
205.0910	11.122	0.003	0.267	0.007	27762.80
304.3033	15.011	0.004	0.29	0.03	43869.79
332.3243	15.796	0.003	0.34	0.02	35081.44
360.3658	16.635	0.004	0.33	0.02	40741.12

Table 4-8. Average number of features (n =3) extracted from LC analysis with piglet plasma by four LC methods.

Average # of Features MC	Average # of Features MC with ZC Conditions	Average # of Features ZC	Average # of Features ZC with MC Conditions
135	214	301	294

Table 4-9. List of retention times, peak widths and plate heights for carnitine standard mixture using the MC method.

Carnitine	t_r (min)	w (min)	H (mm)
C0	3.701	0.447	0.182
C6	11.659	0.318	0.009
C8	12.589	0.330	0.009
C10	13.378	0.341	0.008
C12	14.202	0.400	0.010
C14	15.132	0.730	0.029

Table 4-10. List of retention times, peak widths and plate heights for carnitine standard mixture using the MC with ZC Conditions method.

Carnitine	t_r (min)	w (min)	H (mm)
C0	3.831	0.459	0.179
C6	8.870	0.294	0.014
C8	10.259	0.365	0.016
C10	11.413	0.413	0.016
C12	12.543	0.483	0.019
C14	13.779	0.765	0.039

Table 4-11. List of retention times, peak widths and plate heights for carnitine standard mixture using the ZC method.

Carnitine	t_r (min)	w (min)	H (mm)
C0	0.396	0.153	0.466
C6	1.785	0.836	0.685
C8	5.199	0.318	0.012
C10	6.317	0.235	0.004
C12	7.224	0.271	0.004
C14	8.001	0.271	0.004

Table 4-12. List of retention times, peak widths and plate heights for carnitine standard mixture using the ZC with MC Conditions method.

Carnitine	t_r (min)	w (min)	H (mm)
C0	0.411	0.212	0.831
C6	3.849	0.789	0.131
C8	8.322	0.188	0.002
C10	9.099	0.235	0.002
C12	9.723	0.235	0.002
C14	10.253	0.294	0.003

Table 4-13. List of retention times, peak widths and plate heights for five randomly chosen *m/z*'s using the MC method (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	H (mm)
132.0639	4.483	0.003	0.95	0.03	0.57
205.0910	10.424	0.005	0.27	0.01	0.01
304.3033	14.835	0.008	0.279	0.007	0.004
332.3243	15.659	0.008	0.37	0.03	0.01
360.3658	16.56	0.01	0.35	0.02	0.01

Table 4-14. List of retention times, peak widths and plate heights for five randomly chosen *m/z*'s using the MC with ZC Conditions method (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	H (mm)
132.0639	5.96	0.03	1.36	0.08	0.65
205.0910	7.484	0.004	0.220	0.007	0.01
304.3033	13.461	0.002	0.28	0.02	0.01
332.3243	14.571	0.008	0.37	0.02	0.01
360.3658	15.756	0.003	0.349	0.007	0.01

Table 4-15. List of retention times, peak widths and plate heights for five randomly chosen *m/z*'s using the ZC method (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	H (mm)
132.0639	0.485	0.007	0.251	0.007	0.84
205.0910	0.606	0.002	0.216	0.007	0.40
304.3033	8.20	0.03	0.287	0.006	0.004
332.3243	8.94	0.03	0.30	0.02	0.004
360.3658	9.64	0.04	0.34	0.01	0.004

Table 4-16. List of retention times, peak widths and plate heights for five randomly chosen *m/z*'s using the ZC with MC Conditions method (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	H (mm)
132.0639	0.514	0.008	0.35	0.02	1.41
205.0910	1.134	0.003	0.463	0.007	0.52
304.3033	10.41	0.01	0.24	0.01	0.002
332.3243	10.924	0.009	0.26	0.02	0.002
360.3658	11.419	0.009	0.239	0.007	0.001

Table 4-17. Average number of features (n =3) extracted from piglet plasma analyzed with the Zorbax column and three different solvent sets.

Average # of Features H_2O & ACN, 1% AA	Average # of Features H_2O and ACN, 0.1% AA	Average # of Features H_2O and MeOH, 0.2% AA
278	338	455

Table 4-18. List of retention times, peak widths and theoretical plates for carnitine standard mixture using 1% AA in H2O and ACN.

Carnitine	t_r	w	N
C0	0.424	0.141	144.682
C6	3.108	0.906	188.289
C8	5.451	0.247	7792.529
C10	6.499	0.247	11076.923
C12	7.346	0.224	17207.818
C14	8.100	0.271	14293.923

Table 4-19. List of retention times, peak widths and theoretical plates for carnitine standard mixture using 0.1% AA in H2O and ACN.

Carnitine	t_r	w	N
C0	0.441	0.153	132.927
C6	3.666	0.789	345.423
C8	5.562	0.259	7378.744
C10	6.668	0.235	12881.767
C12	7.598	0.235	16725.644
C14	8.387	0.33	10334.879

Table 4-20. List of retention times, peak widths and theoretical plates for carnitine standard mixture using 0.2% AA in H2O and MeOH.

Carnitine	t_r	w	N
C0	0.458	0.153	143.373
C6	4.731	0.424	1992.022
C8	7.250	0.318	8316.522
C10	9.005	0.306	13856.213
C12	10.252	0.259	25069.037
C14	11.171	0.318	19744.668

Table 4-21. List of retention times, peak widths and theoretical plates for five randomly chosen *m/z*'s using 1% AA in H_2O and ACN (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	N
132.0639	0.55	0.02	0.33	0.03	43.05
205.0910	0.68	0.03	0.25	0.06	118.16
304.3033	8.31	0.02	0.28	0.01	14224.84
332.3243	9.03	0.02	0.306	0.000	13924.00
360.3658	9.73	0.02	0.353	0.000	12165.32

Table 4-22. List of retention times, peak widths and theoretical plates for five randomly chosen *m/z*'s using 0.1% AA in H_2O and ACN (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	N
132.0639	0.588	0.006	0.26	0.01	82.37
205.0910	0.80	0.01	0.267	0.007	144.96
304.3033	8.25	0.02	0.33	0.01	10283.12
332.3243	8.99	0.02	0.48	0.02	5642.98
360.3658	9.76	0.02	0.47	0.01	6991.40

Table 4-23. List of retention times, peak widths and theoretical plates for five randomly chosen *m/z*'s using 0.2% AA in H_2O and MeOH (standard deviation = σ). The number of significant figures is limited by the standard deviation.

m/z	Average t_r (min)	σ of t_r (min)	Average w (min)	σ of w (min)	N
132.0639	0.64	0.01	0.381	0.03	45.60
205.0910	1.56	0.06	0.683	0.02	83.30
304.3033	10.34	0.02	0.294	0.02	19758.87
332.3243	11.09	0.02	0.334	0.006	17658.98
360.3658	11.74	0.02	0.298	0.007	24817.17

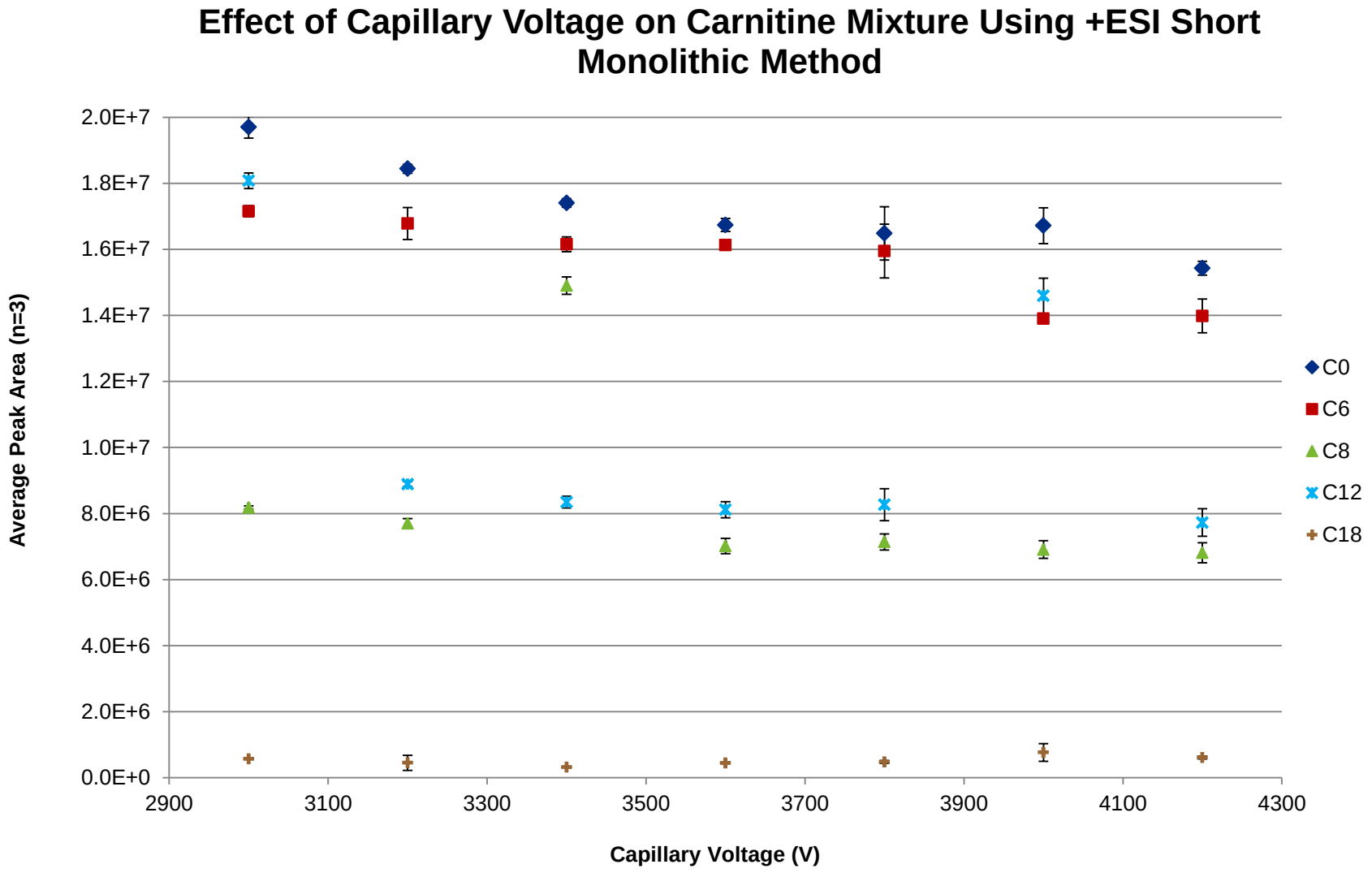


Figure 4-1. Average peak areas of different carnitines as a function of capillary voltage. Error bars indicate the standard deviation.

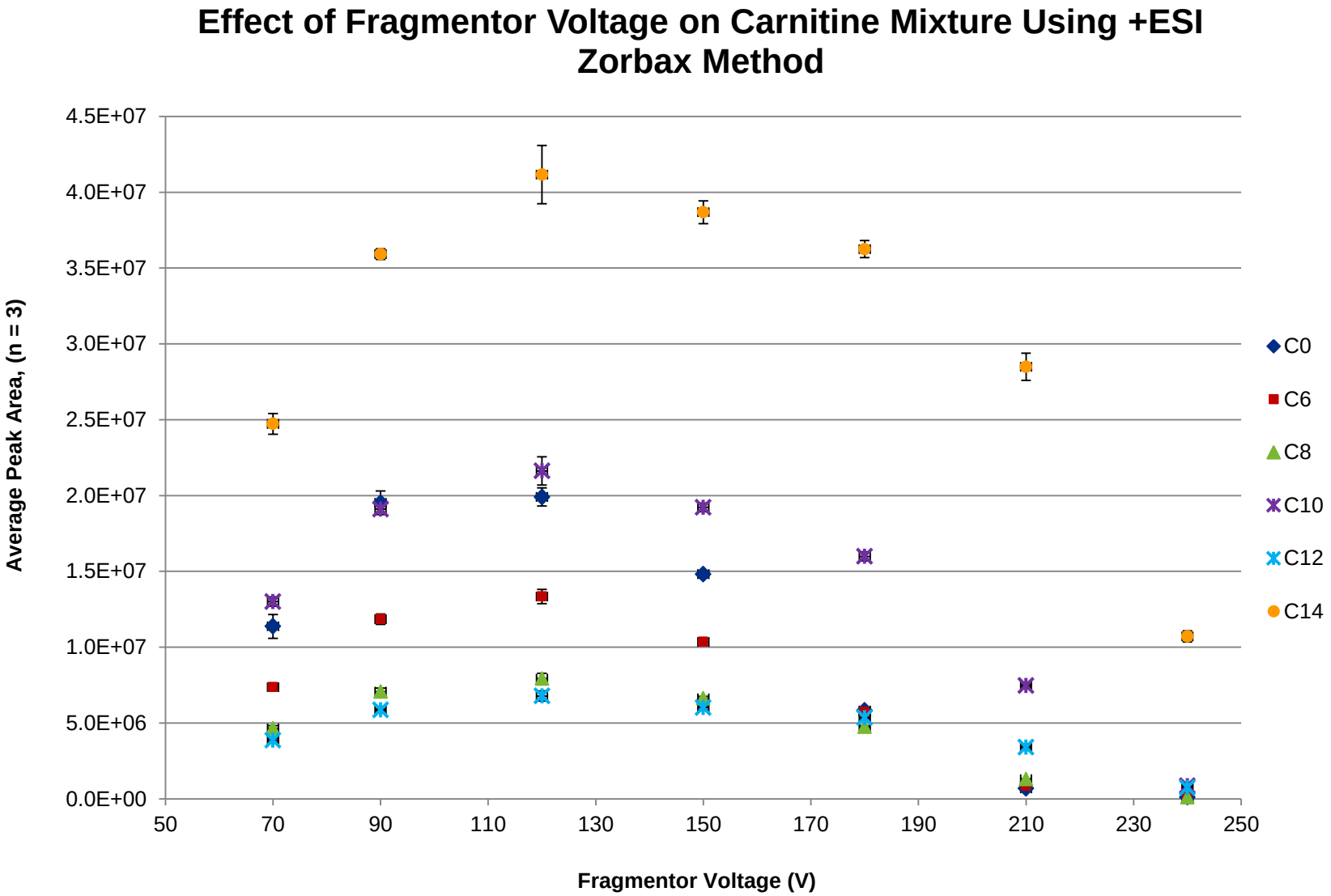


Figure 4-2. Average peak areas of different carnitines as a function of fragmentor voltage. Error bars indicate the standard deviation.

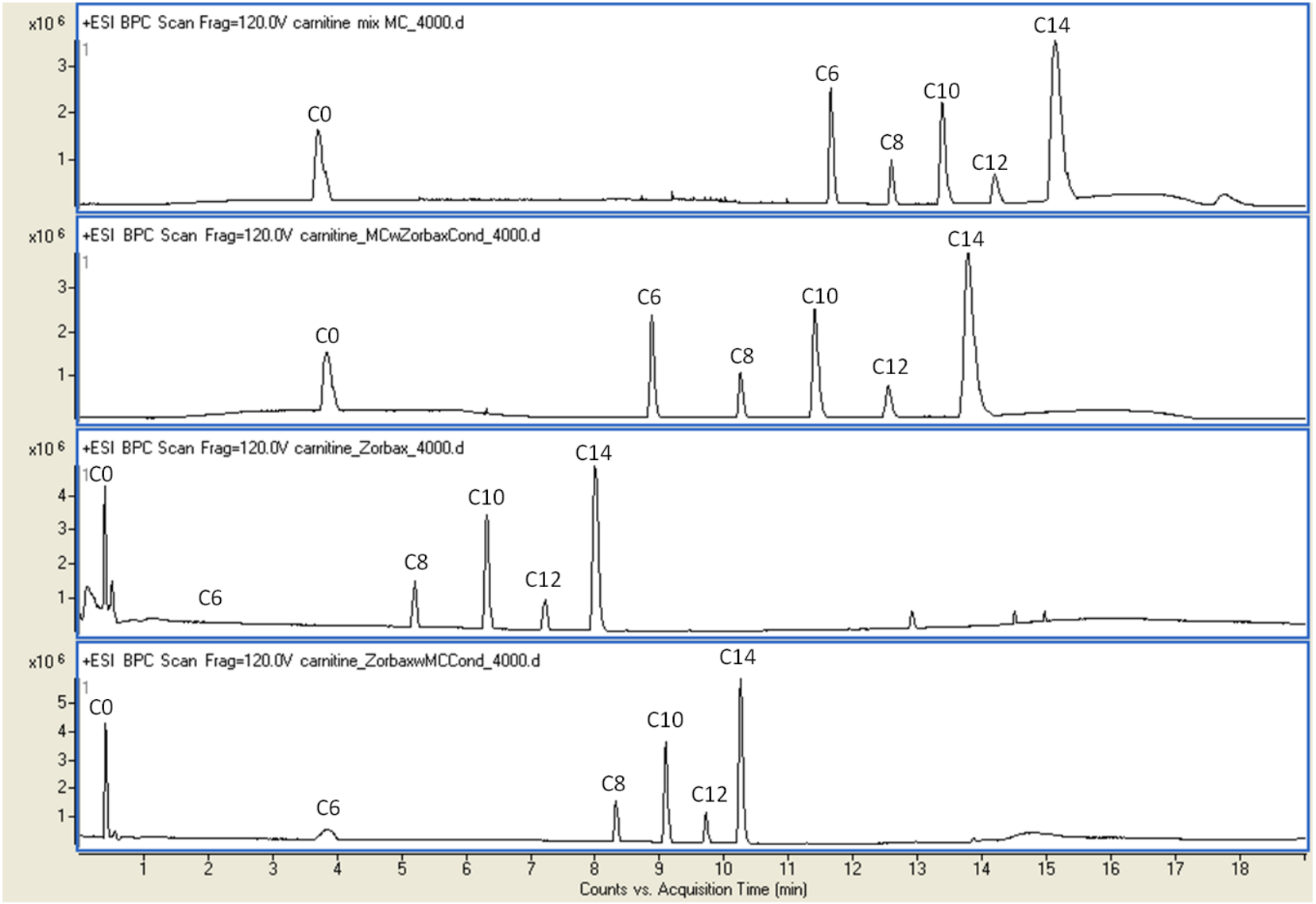


Figure 4-3. Base peak chromatograms of standard mixtures of carnitines using four different LC methods. From top to bottom: MC method, MC with ZC conditions method, ZC method, ZC with MC conditions method.

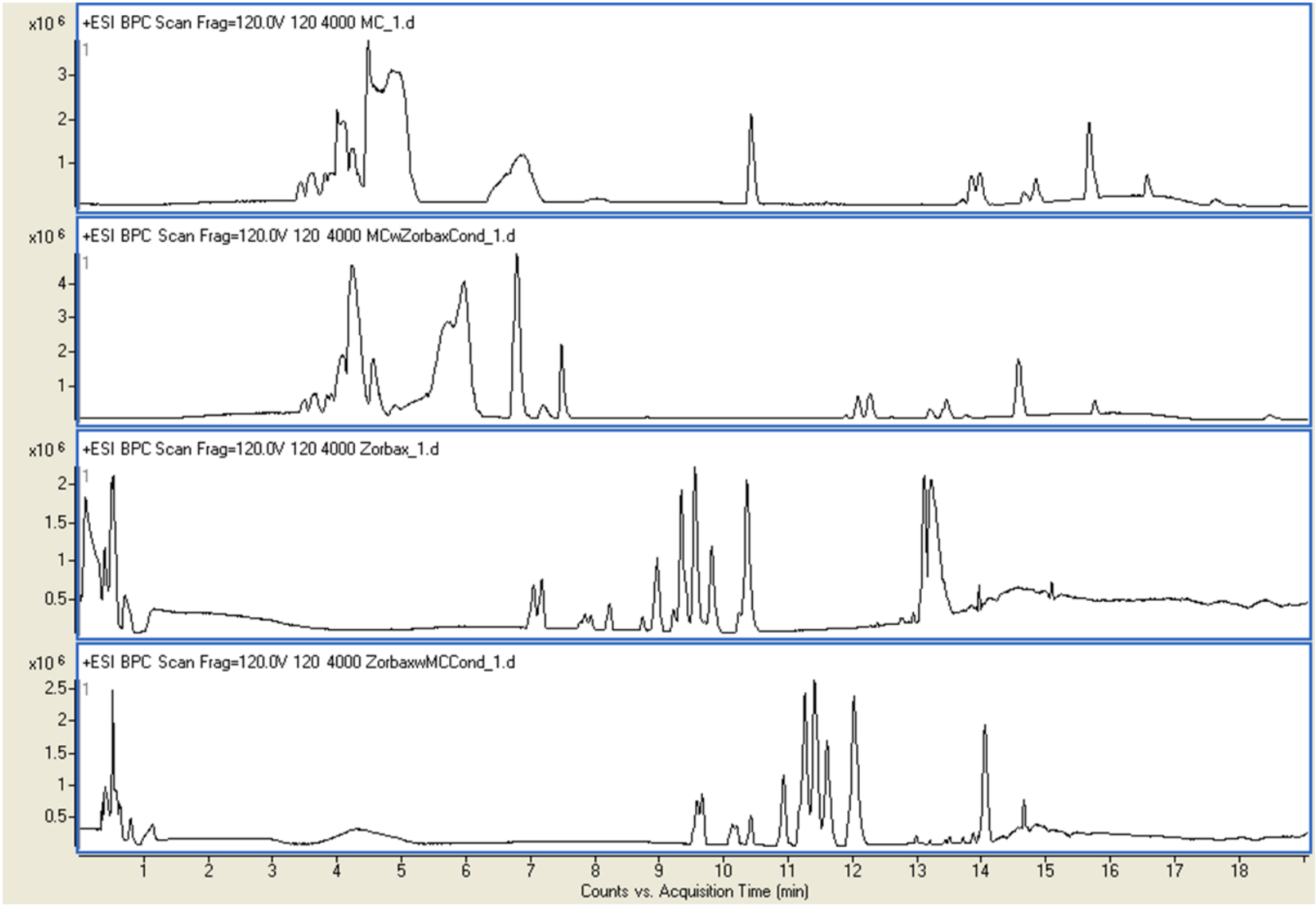


Figure 4-4. Base peak chromatograms of piglet plasma using four different LC methods. From top to bottom: MC method, MC with ZC conditions method, ZC method, ZC with MC conditions method.

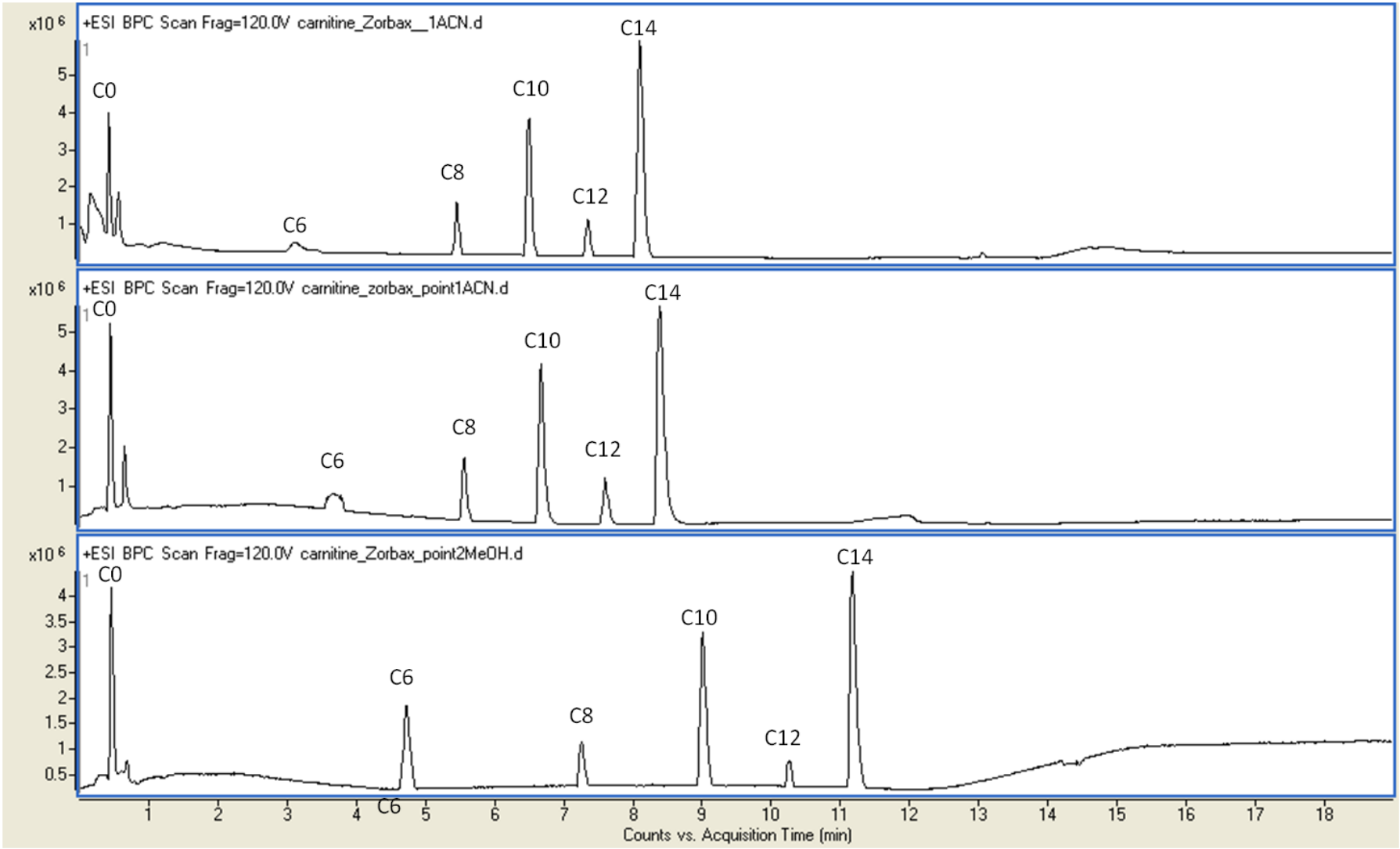


Figure 4-5. Base peak chromatograms of standard mixtures of carnitines using three different solvent sets and Zorbax column. From top to bottom: H_20 and ACN with 1% AA, H_2O and ACN with 0.1% AA, H_2O and MeOH with 0.2% AA.

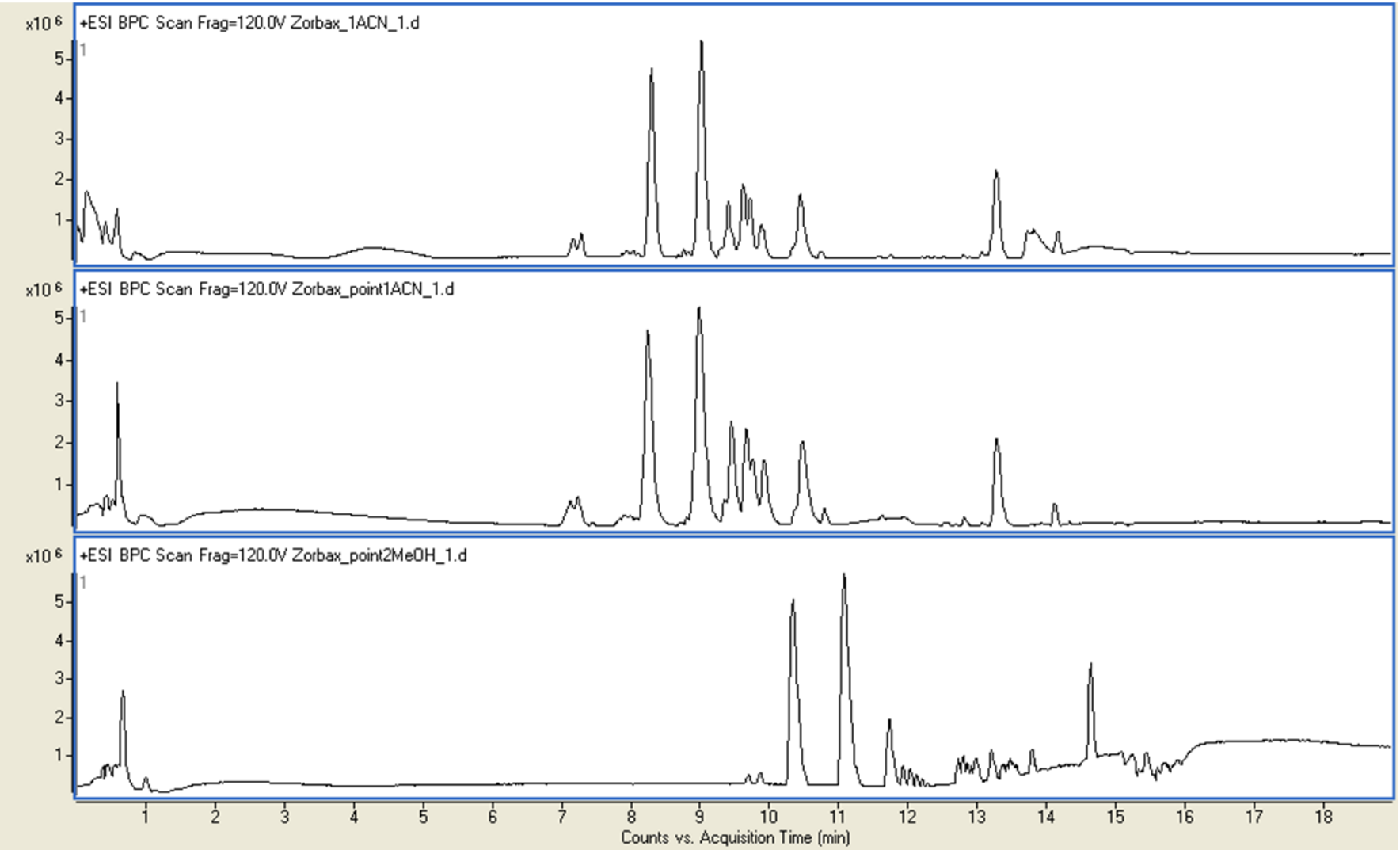


Figure 4-6. Base peak chromatograms of piglet plasma using three different solvent sets and Zorbax column. From top to bottom: H_20 and ACN with 1% AA, H_2O and ACN with 0.1% AA, H_2O and MeOH with 0.2% AA.

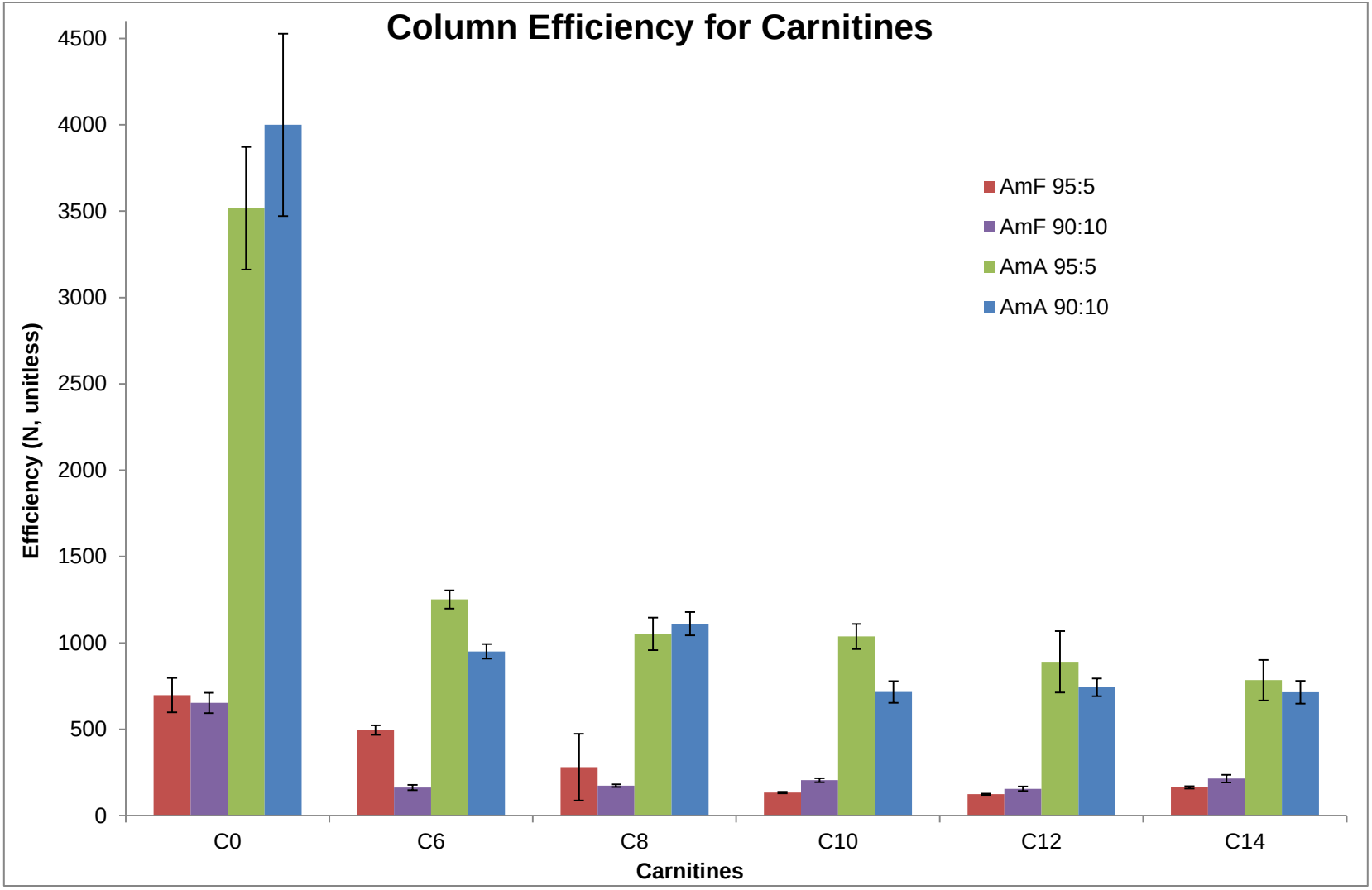


Figure 4-7. Average efficiencies (N) of carnitine standards for HILIC with four different solvent systems. AmF and AmA are abbreviations of ammonium formate and ammonium acetate, respectively. Error bars indicate the standard deviation of the mean (n = 3).

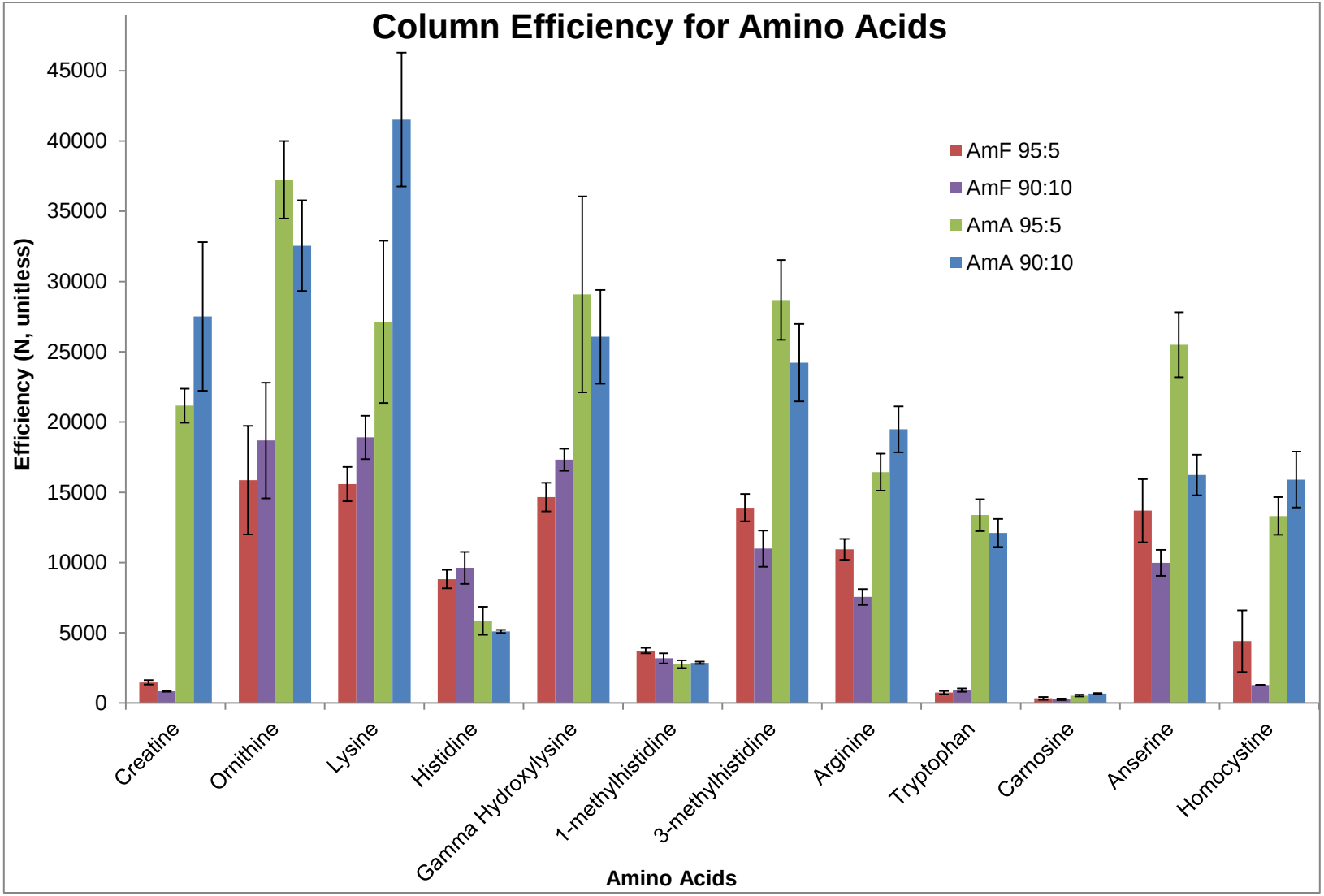


Figure 4-8. Average efficiencies (N) of amino acid standards for HILIC with four different solvent systems. AmF and AmA are abbreviations of ammonium formate and ammonium acetate, respectively. Error bars indicate the standard deviation of the mean (n = 3).

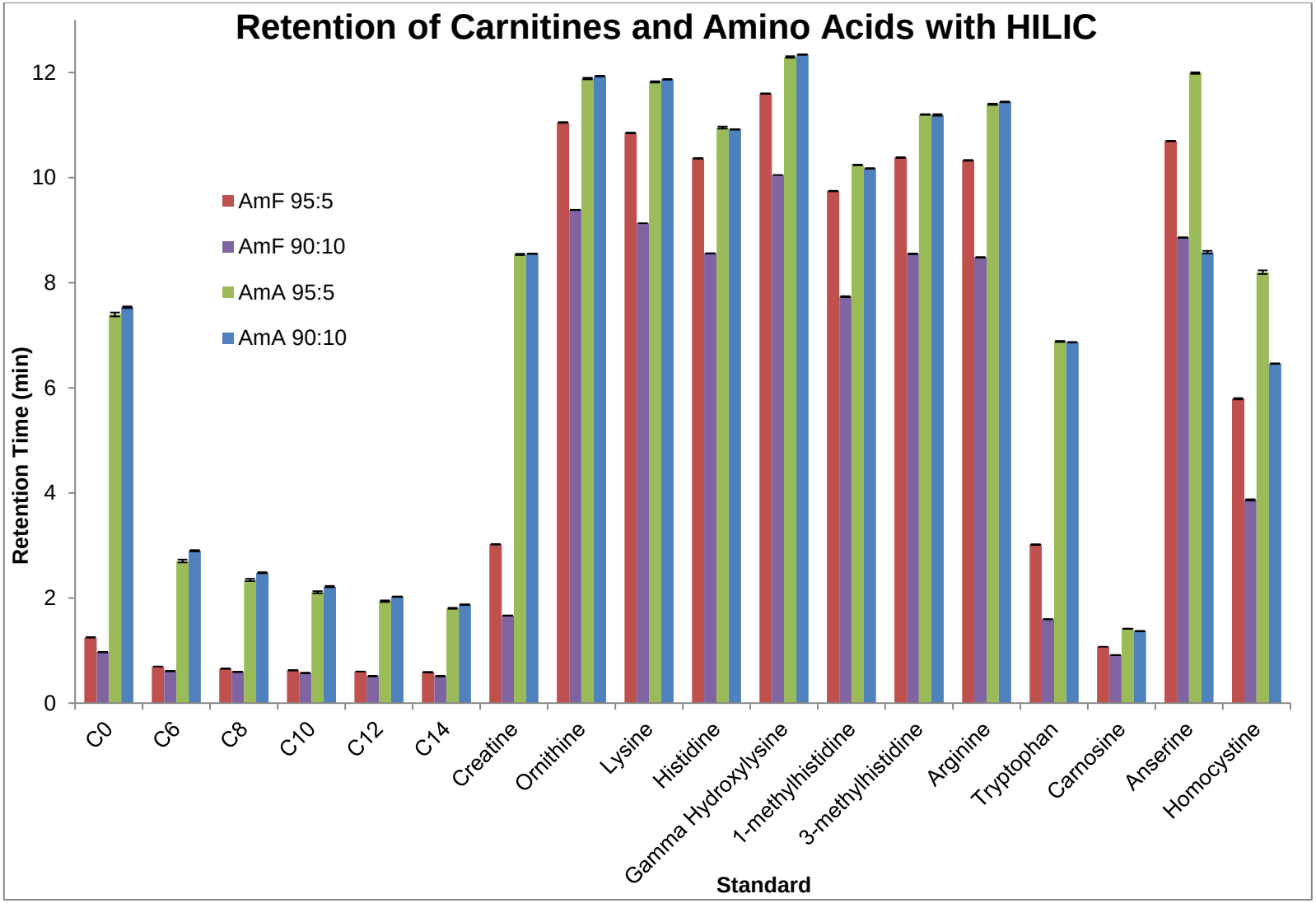


Figure 4-9. Average retention times of standards for HILIC with four different solvent systems. AmF and AmA are abbreviations of ammonium formate and ammonium acetate, respectively. Error bars indicate the standard deviation of the mean (n = 3).

CHAPTER 5
METABOLOMICS OF ESTROUS AND ANESTROUS URINE FROM MANATEES

Background

The West Indian manatees, or *Trichechus manatus*, reside around Florida and the Caribbean. Manatees are marine mammals that typically weigh 800-1,200 pounds and reach a length of about 10 feet. It is common for these large herbivores to eat more than 32 pounds of plants a day, a task that takes 5-8 hours.[85-87] Manatees are often called “gentle giants”, as they are very docile and curious animals.

Manatees are semi-social, in that they interact with one another but do not form any permanent bonds. The only exception to the bonding behavior is when a female manatee births a calf. After birth it is typical for the calf to remain with the mother 1-2 years.[88] Since manatees do not form permanent bonds it is common for male manatees to compete with one another in order to mate with a single female. When a female manatee is in estrus male manatees will pursue for as long as 2-4 weeks, creating a mating herd.[89] It is assumed that a signal, or pheromone, is given off by the female that allows male manatees to know when she is receptive to mating. Manatees have limited vision and lack communication, other than between mother and calf; because of this, the pheromone is thought to be chemical.[89-90]

Pheromones are substances utilized for intra-specific communication, or communication between the same species. They are emitted outside the body of an individual and received by other individuals of that species. The pheromone can contain information about gender, dominance, or reproductive status, as well as other characteristics. The reactions to the pheromone should be species-wide and not independent of the individual sending or receiving it.[91] It is hypothesized that the

pheromones allowing male manatees to know when a female manatee is receptive to mating are transmitted through urine. This is a common way for pheromones to be secreted, along with feces and through glands.[91] However, manatees do not have any known glands, thus urine seems the most obvious pathway.

Two compounds known to be associated with the manatee reproductive status, progesterone and β-estradiol, are commonly monitored to determine where in the reproductive cycle a female manatee is.[92-93] The urine samples obtained for this work were determined to be from estrus or anestrus periods by radioimmunoassays that measured the above urinary hormones. Additionally, some of the estrous urine samples were shown to invoke a behavioral response in some male manatees, suggesting that the pheromone is present in urine.

Elephants are mammals that are closely related to manatees and progesterone has also been monitored to determine where in the reproductive cycle female elephants are.[94] Additionally, it has been observed that mating behavior does not occur at the maximum progesterone level, indicating that progesterone is not the pheromone responsible for showing the readiness of a female to mate.[94] In fact, the sex pheromone that indicates a female elephant is ready for mating has been identified for elephants. The compound is (Z)-7-dodecen-1-yl acetate.[95] The same compound is the sex pheromone for about 140 different species of moths.[96] The pheromone in manatees may be structurally similar, in which case GC/MS analysis may be necessary. However, little is known about what type of compound the sex pheromone may be, so initial experiments will use LC/MS analyses. In addition potentially identifying a sex

pheromone we will be able to learn more about the profiling of different metabolites during the manatees estrus cycle.

The metabolomics of urine from estrus and anestrus female manatees was performed using two different LC columns. LC with MS/MS was performed for the compounds found to be significant between the estrous and anestrous groups. In this work, particular attention was paid to those compounds found to be elevated in estrous urine. The goal of this project was to potentially identify the sex pheromone in female manatee urine, but to also understand the reproductive cycle of manatees. This understanding would allow the possibility of better breeding techniques for captive manatees and potentially preserve this endangered species.[92]

Experimental

The experimental conditions are the same as those described in Chapter 2 and Chapter 3, unless otherwise indicated below.

Sample Collection and Preparation

Manatee urine was collected directly from four different females at varying times between May 2003 and August 2010. There were a total of ten estrous samples and ten anestrous samples. The urine was stored frozen until the time of analysis and was directly injected (no sample preparation).

Liquid Chromatography Conditions

The Zorbax chromatography method (ZC) introduced in Chapter 3 was implemented in this work (Table 4-2). The solvents used were methanol (A) and water (B), both with 0.2% acetic acid. The column was heated to 60 ˚C. An injection volume of 5 µL of urine was used and the samples were analyzed in triplicate for a total of 60 data files.

HILIC chromatography was also performed after noting that all 9 significant compounds found by statistical analysis of the Zorbax chromatography method were at a retention time of 0.3-0.4 minutes. The HILIC chromatography method used was AmA 95:5, explained in Chapter 4 and implementing the Halo column. The injection volume for the HILIC method was 5 µL.

Electrospray Source

The conditions used are the same as those listed in Chapter 3.

Mass Spectrometer

The mass spectrometer conditions are identical to those of Chapter 2 and Chapter 3.

MS/MS Instrumentation

The MS/MS instrumentation is described in Chapter 3. MS/MS events were created to scan for the significant ions.

Data Analysis

The LC/MS data analysis process is described in both Chapter 2 and Chapter 3. MS/MS data analysis was performed with Thermo Scientific's Xcalibur Qual Browser (Thermo Scientific, Waltham, MA).

Results and Discussion

RP/MS

The chromatograms of estrous and anestrous urine samples obtained with the Zorbax column are shown in Figures 5-1 and 5-2. In general, the chromatograms for the estrous samples appear to be more complex in the first couple of minutes than the anestrous chromatograms.

The manatee urine analyzed by LC/MS with the Zorbax method was determined to have 1,472 masses present. These masses were subjected to an asymptotic *t*-test (p = 0.05) with a Benjamini-Hochberg FDR multiple testing correction. The t-test resulted in only 9 significant masses between the two groups of estrus and anestrus. This small number is very easy to work with and MS/MS can be performed on all 9 of the masses. The reason for this low of a number of masses may be due to only subtle changes between estrous and anestrous urine. Additionally, if an anestrous sample is collected close to the beginning of estrus then fewer changes would be expected between the two groups. It is also important to note that all the 9 masses have early retention times, indicating a high solubility in water.

The 3D-PCA plot, created with the statistically significant masses, is shown in Figure 5-3. The x-axis (component 1) accounts for 60.50% of the variation, while the y- and z-axes (components 2 and 3) account for 15.73% and 7.90%, respectively. Distinct separation is not observed between the two groups but the estrous samples do appear to group towards the top right area of the plot. A two-dimensional plot is also shown in Figure 5-4. The 2D view allows better visualization of the grouping. The estrous samples group tightly on the left side of the 2D plot.

A PLS-DA was also performed and the plot can be seen in Figure 5-5 (3D view) and Figure 5-6 (2D view). The prediction model had an accuracy of 83.33%, most of the inaccuracies resulted from 9 anestrous replicates being predicted as estrus (Table 5-1). Only one estrous replicate was predicted to be anestrus. The loadings plot from the PLS-DA was not considered in this experiment, as the low number of significant features allowed further examination of all the masses.

The 9 masses determined to be significant were extracted from each of the three replicates for the estrous and anestrous urine samples. The extracted ion chromatograms were then manually integrated and the peak areas were averaged for all the estrous replicates as well as for the anestrous replicates. A bar graph was then created (Figure 5-7) comparing the average abundance of the 9 masses in estrous and anestrous urine. As can be seen, three masses show an increase from anestrous to estrous urine. However, no mass is only present in the estrous urine, as we would expect if it was a pheromone released only during estrus.

The three masses found to be elevated in estrous urine were the focus of MS/MS experiments. However, the two ions with *m/z* 143 could not be differentiated with MS/MS due to the lack of high resolution data. The significant ion at *m/z* 175 resulted in a loss of 32 mass units that could be a loss of a methoxy group (Figure 5-8). There were two accurate mass library results, neither of which would likely ionize in positive mode or have a loss of 32 Da. Formulas were calculated based on accurate mass and one resulted in a structure that would likely lose a methoxy group. That formula was $C_7H_{10}O_3S$. The MS/MS spectrum for *m/z* 143 is shown in Figure 5-9, a loss of 42 mass units is observed at *m/z* 101. It is not known if this is mixture of the two 143 ions found by RPLC/MS or one of them. An accurate mass library search resulted in two compound for *m/z* 143.0242 but none for *m/z* 143.0197. However the two compounds in the library did not have matching fragment ions and one would only ionize in negative mode. The formula for the compound that would ionize in positive mode is $C_5H_6N_2OS$. All data base searching was performed for $[M+H]^+$ ions with a mass tolerance of 30 ppm.

HILIC/MS

The results of the HILIC chromatography can be seen in Figure 5-10 for estrous urine samples and in Figure 5-9 for anestrous samples. Again, the estrous samples seem to generate a more complicated chromatogram than the anestrous samples. It is obvious that more compounds are being retained on the HILIC column than were with the Zorbax column if Figures 5-1 and 5-2 are compared to Figures 5-10 and 5-11.

When the HILIC/MS results for manatee urine were examined with MPP, more than two times the number of aligned compounds were found for HILIC than were for RP/MS with the Zorbax column. The manatee urine analyzed by HILIC/MS had 3,209 masses present, versus the 1,472 masses found aligned with the Zorbax method. This is not surprising considering polar nature of urine; better separation should be achieved with HILIC than RP chromatography and thus allow for better statistical analyses. The 3,209 aligned compounds were subjected to an asymptotic *t*-test ($p = 0.05$) with a Benjamini-Hochberg FDR multiple testing correction. The t-test resulted in 30 significant masses between the two groups of estrus and anestrus. Again, this is much higher, greater than 3 times the number of significant masses found using RPLC. Although there are more significant masses found with HILIC than with RPLC, 30 masses still allows for relatively easy MS/MS analyses.

The 3D-PCA plot, created with the 30 statistically significant masses, is shown in Figure 5-12. The x-axis (component 1) accounts for 44.70% of the variation, while the y- and z-axes (components 2 and 3) account for 11.93% and 9.59%, respectively. Fair separation is observed between the two groups, where the estrous samples appear to group towards the back right area of the plot. A two-dimensional plot is also shown in

Figure 5-13. The 2D view allows better visualization of the grouping. The estrous samples group on the left side of the 2D plot.

A PLS-DA was also performed and the plot can be seen in Figure 5-14 (3D view) and Figure 5-15 (2D view). The prediction model had an accuracy of 98.33%, the inaccuracy resulted from one anestrous replicate being predicted as estrus (Table 5-2). No estrous replicates were predicted to be anestrus. The loadings plot from the PLS-DA was not considered in this experiment, as the low number of significant features allowed further examination of all the masses.

The masses determined to be significant and also elevated (10 masses) in the estrous samples were extracted from each of the three replicates for the estrous and anestrous urine samples. The extracted ion chromatograms were then manually integrated and the peak areas were averaged for all the estrous replicates as well as for the anestrous replicates. A bar graph was then created (Figure 5-16) comparing the average abundance of the 10 masses in estrous and anestrous urine. As can be seen, when manually extracted one *m/z* is actually not elevated in the estrous urine despite being identified as up regulated in the estrous urine with MPP. Similar to the RPLC/MS results, no mass is only present in the estrous urine, as we would expect if it was a pheromone released only during estrus.

The 10 masses found to be elevated in estrous urine were the focus of MS/MS experiments. One of the accurate masses matched that of creatine and the fragmentation pattern observed in the urine samples matched the fragmentation pattern of a standard of creatine (Figure 5-17). In the MS/MS spectrum the ion at *m/z* 90 is due to the loss of the CN_2H_2 group. The retention time of creatine with HILIC/MS also

matched the ion of interest at *m/z* 132. Thus, the significant mass at *m/z* 132 was determined to be creatine. Another significant mass, at *m/z* 148, had a result based on accurate mass as glutamic acid and also had fragmentation that matched that observed for a standard of glutamic acid. In Figure 5-18, the transition of *m/z* 148 to *m/z* 130 and 102 is due to the loss of water and the loss of a carboxylic acid, respectively. The MS/MS spectrum for the significant mass at *m/z* 176 is shown in Figure 5-19. The loss of 44 mass units is observed at *m/z* 132 and indicates the loss of a CN_2H_4 group. One accurate mass library result was argininic acid, and although no MS/MS data was available the structure of argininic acid support the MS/MS fragmentation observed. Finally, the significant mass at *m/z* 244 resulted in a large fragment ion at *m/z* 112 (Figure 5-20). One of the results of the accurate mass library search was cytidine. The fragmentation observed in the library matched that observed for the ion at *m/z* 112 as well as a synthetic standard, indicating a loss of the oxygen containing ring.

A summary of the identified metabolites found by both RPLC/MS and HILIC/MS can be seen in Table 5-3. Compounds not included on this table resulted in: no hits from metabolite libraries, no conclusive calculated formulas based on accurate mass, MS/MS fragments that did not match for any hits for potential metabolites, and/or were not detected by the LIT.

Summary

Other than the possible identification of $C_7H_{10}O_3S$ as *m/z* 175 found by RP/MS, most of the other identified metabolites are responsible for energy supply or waste removal. The exception is cytidine, which is a nucleoside. It is likely that none of the identified compounds are the pheromone of interest. The fact that none of the significant ions were exclusive to estrous urine supports that claim. Additionally, *m/z*

values of known sex pheromones, including the elephant pheromone, were extracted from the data files but yielded no significant peaks. Those urine samples that did invoke a response in male manatees were also compared to those that did not, but no additional potential pheromones could be found. This work is still novel, as no one has explored the metabolites of manatee urine. It is also not common that urine is directly injected, but this work has shown differences can still be seen between estrous and anestrous urine without sample preparation. It has been shown that HILIC is a better chromatographic technique for the separation of urine but different ionization modes/sources can still be explored, such as –ESI and +/-APCI. Any potential pheromone would also need to be tested for behavior responses within male manatees. Finally, this work may help our understanding of the manatee estrus cycle, so that reproduction can be assisted to better maintain this endangered species.[92]

Table 5-1. PLS-DA prediction results for urine from manatees in estrus and anestrus by RPLC/MS.

	Predictions	
Estrus	Anestrus	Accuracy (%)
29	1	96.67
9	21	70.00
		83.33

There are 30 files for each group, thus each should have a prediction of 30, those that vary cause the accuracy of the model to decrease

Table 5-2. PLS-DA prediction results for urine from manatees in estrus and anestrus by HILIC/MS.

	Predictions	
Estrus	Anestrus	Accuracy (%)
30	0	100.0
1	29	96.67
		98.33

There are 30 files for each group, thus each should have a prediction of 30, those that vary cause the accuracy of the model to decrease

Table 5-3. Overview of identified significant masses elevated in estrous urine by RPLC/MS and HILIC/MS.

Metabolite	Formula	Experimental *m/z*	Theoretical *m/z*	Error (ppm)	Retention (min)
Unknown	$C_7H_{10}O_3S$	175.0441	175.0423	10.3	0.3, RPLC
Creatine*	$C_4H_9N_3O_2$	132.0760	132.0768	-6.1	8.1, HILIC
Glutamic acid*	$C_5H_9NO_4$	148.0596	148.0604	-5.4	10.0, HILIC
Argininic Acid	$C_6H_{13}N_3O_3$	176.1021	176.1030	-5.1	9.5, HILC
Cytidine*	$C_9H_{13}N_3O_5$	244.0917	244.0928	-4.5	2.8, HILIC

Metabolites are noted with RPLC or HILIC under retention to indicate if they were identified as significant with RPLC/MS or HILIC/MS. An asterisk indicates verification with a synthetic standard.

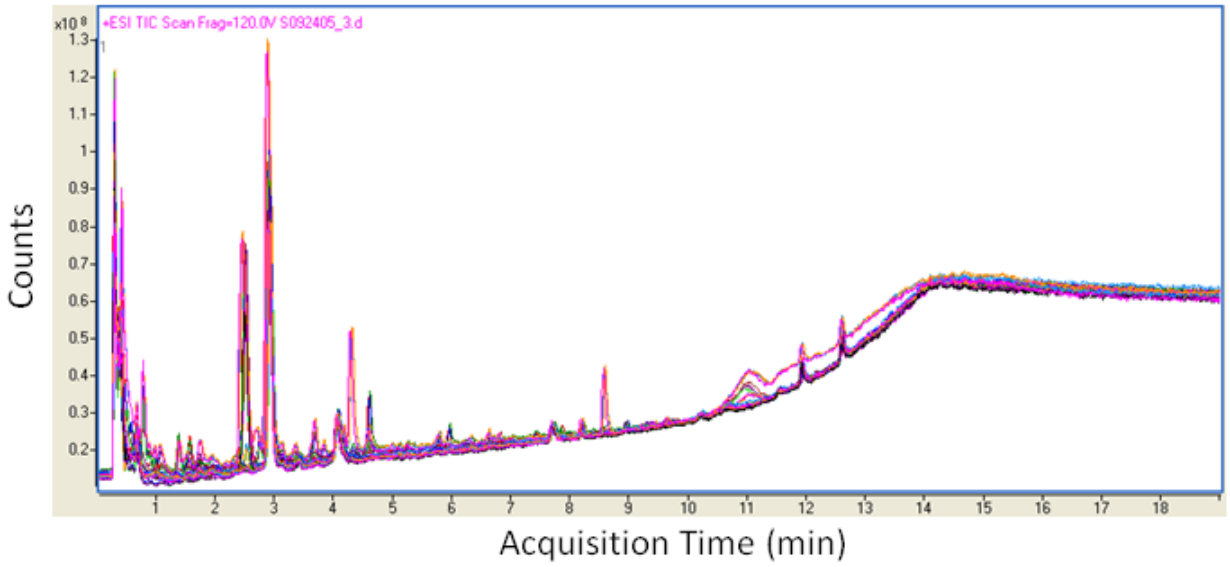


Figure 5-1. Overlay of 30 chromatograms of manatee urine from animals in estrus using RPLC.

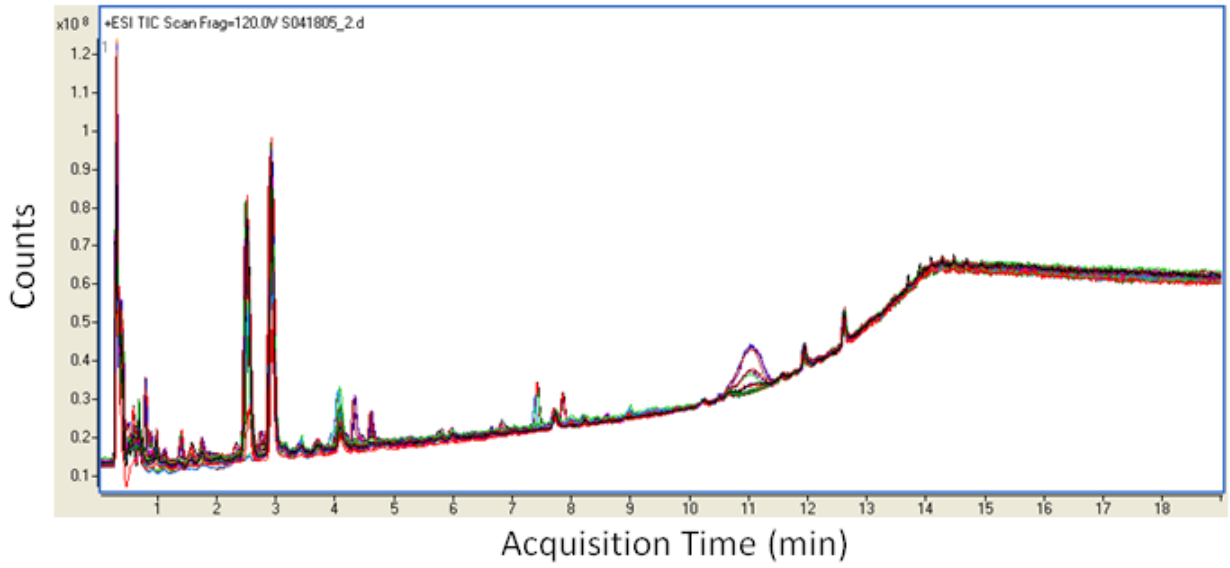


Figure 5-2. Overlay of 30 chromatograms of manatee urine from animals in anestrus using RPLC.

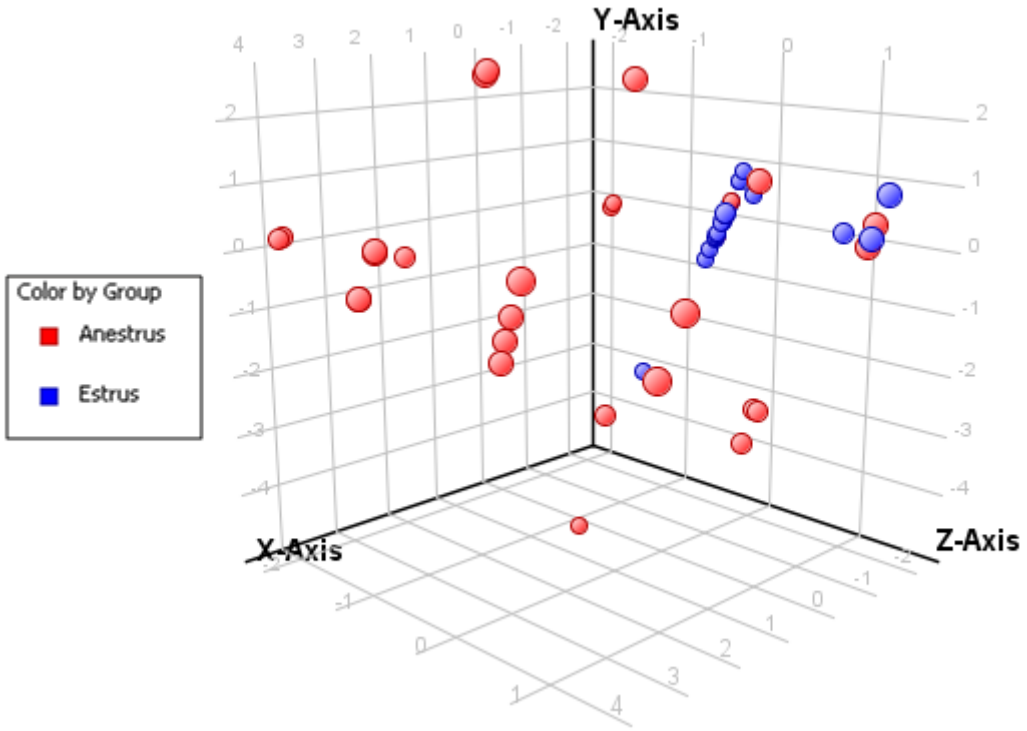


Figure 5-3. 3D-PCA plot of the significant masses in urine from manatees in estrus and anestrus (9 masses, p = 0.05) by RPLC/MS. Each point represents one replicate.

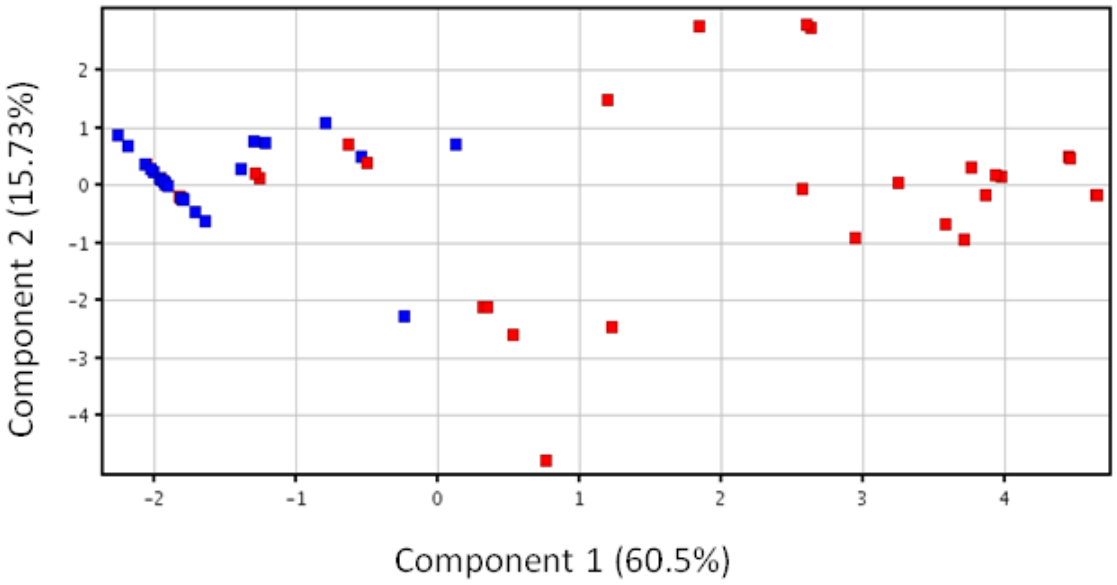


Figure 5-4. 2D-PCA plot of the significant masses in urine from manatees in estrus and anestrus (9 masses, p = 0.05) by RPLC/MS. Each point represents one replicate. The color code is the same as Figure 5-5.

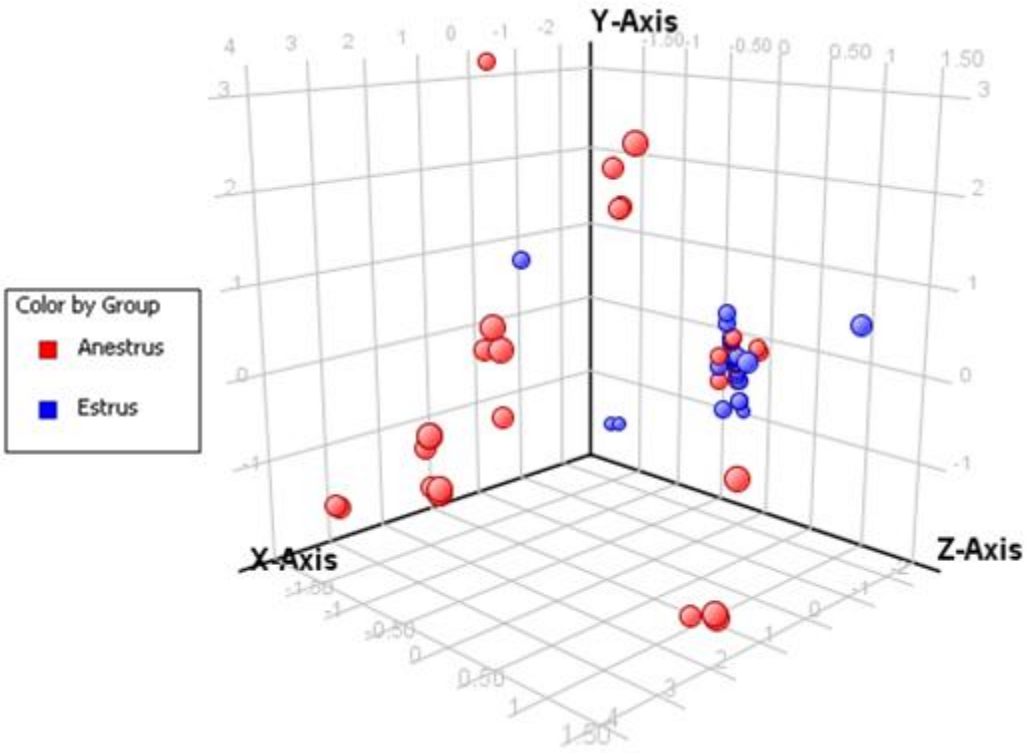


Figure 5-5. 3D-PLS-DA plot of *t*-test (p = 0.05) masses in urine from manatees in estrus and anestrus (9 masses) by RPLC/MS. Each point on the plot depicts one replicate.

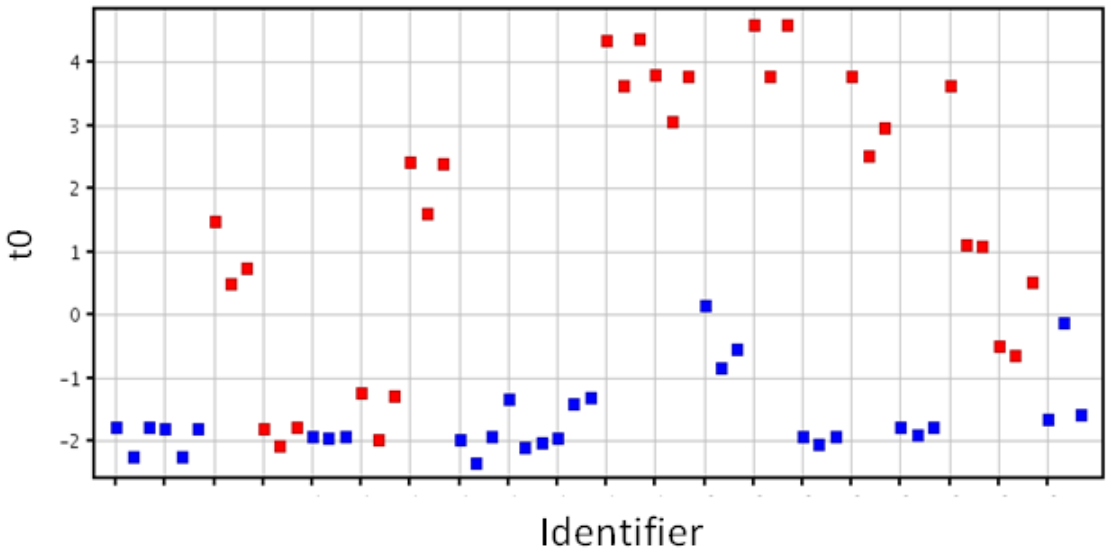


Figure 5-6. 2-PLS-DA plot of *t*-test (p = 0.05) masses in manatees in estrus and anestrus (9 masses) by RPLC/MS. Each point on the plot depicts one replicate. The x-axis (Identifier) values are the individual sample replicates. The color code is the same as Figure 5-7.

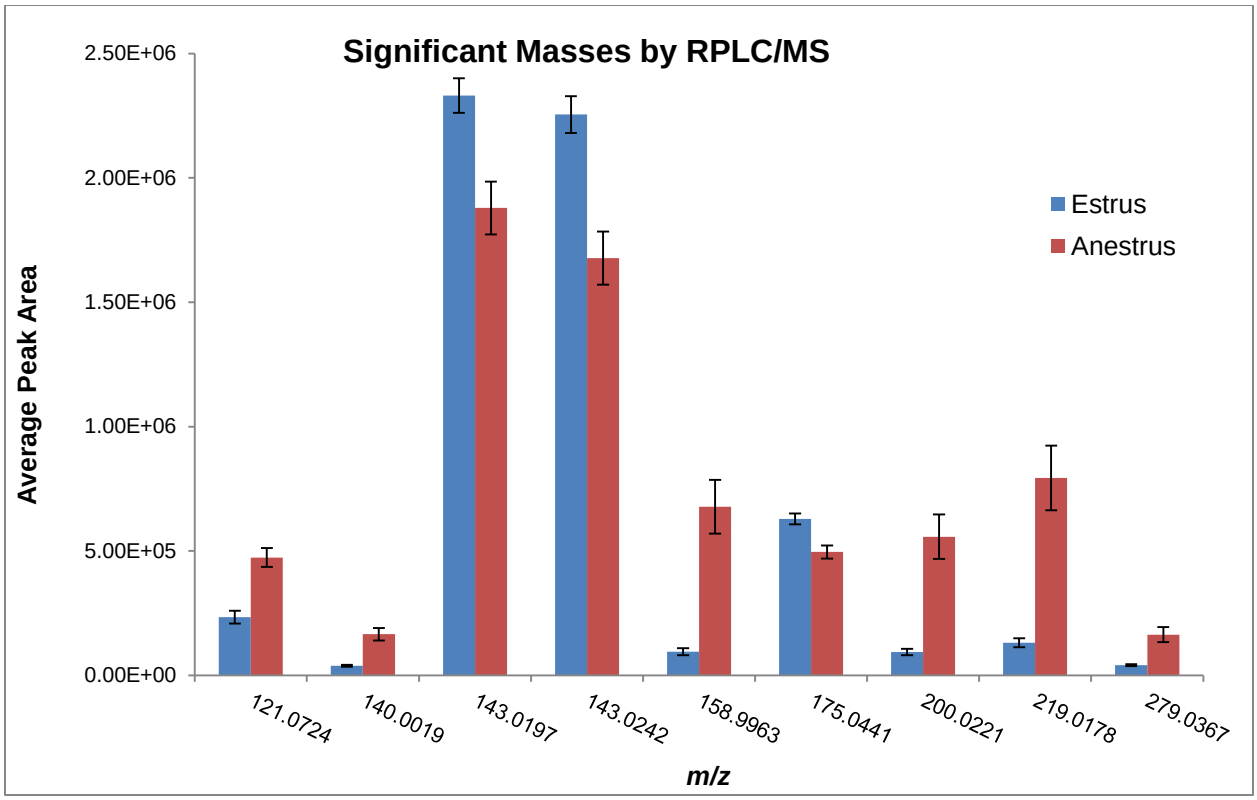


Figure 5-7. Bar graph comparing the average peak areas of the 9 significant masses in estrous and anestrous urine by RPLC/MS. A possible formula was identified in Table 5-3 for *m/z* 175.0441. Error bars indicate the standard deviation of the mean (n = 30).

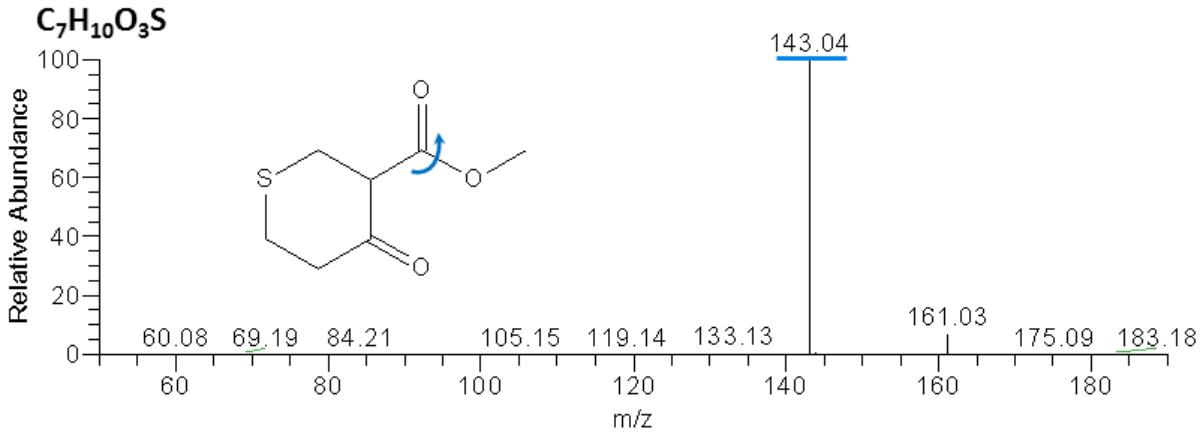


Figure 5-8. MS/MS spectrum of *m/z* 175 in estrous urine, determined by RPLC/MS. The ion at *m/z* 143 (blue arrow) is likely due to the loss of a methoxy group if the proposed structure is correct. The proposed structure is based on accurate mass.

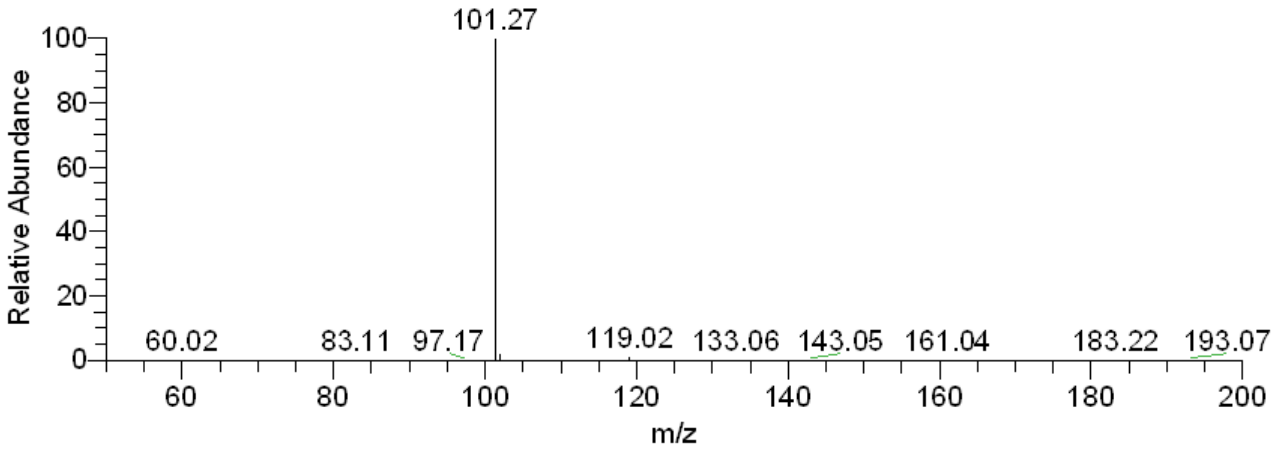


Figure 5-9. MS/MS spectrum of *m/z* 143 in estrous urine, determined by RPLC/MS. A loss of 42 mass units is observed at m/z 101.

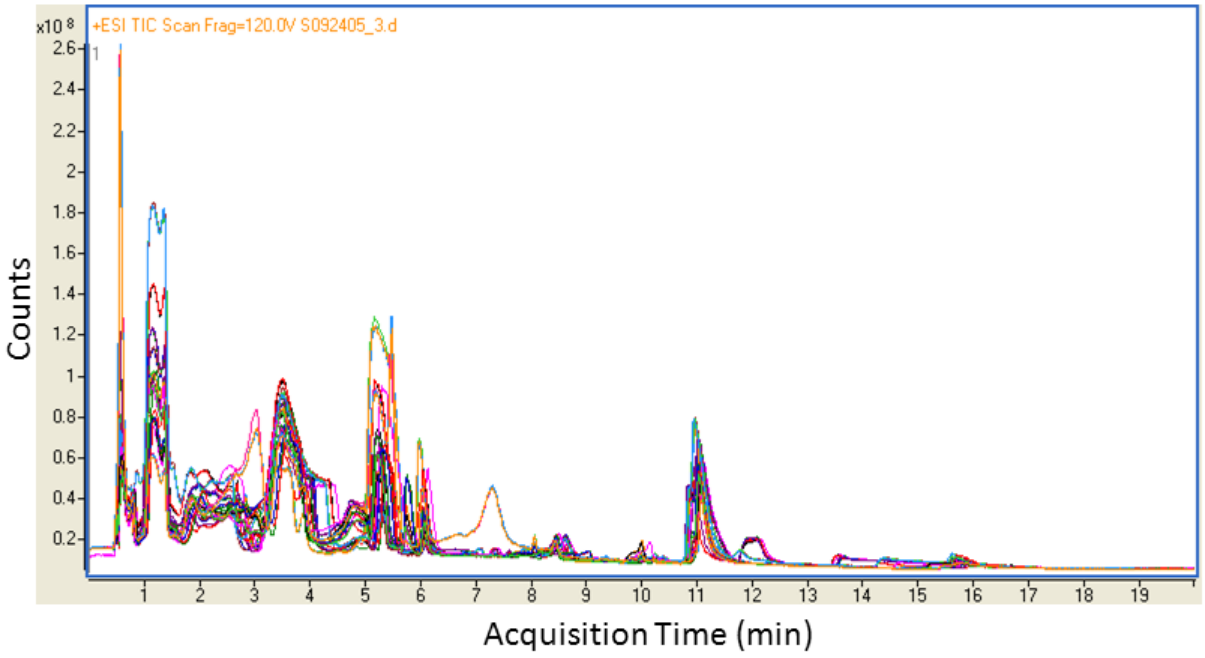


Figure 5-10. Overlay of 30 chromatograms of manatee urine from animals in estrus using HILIC.

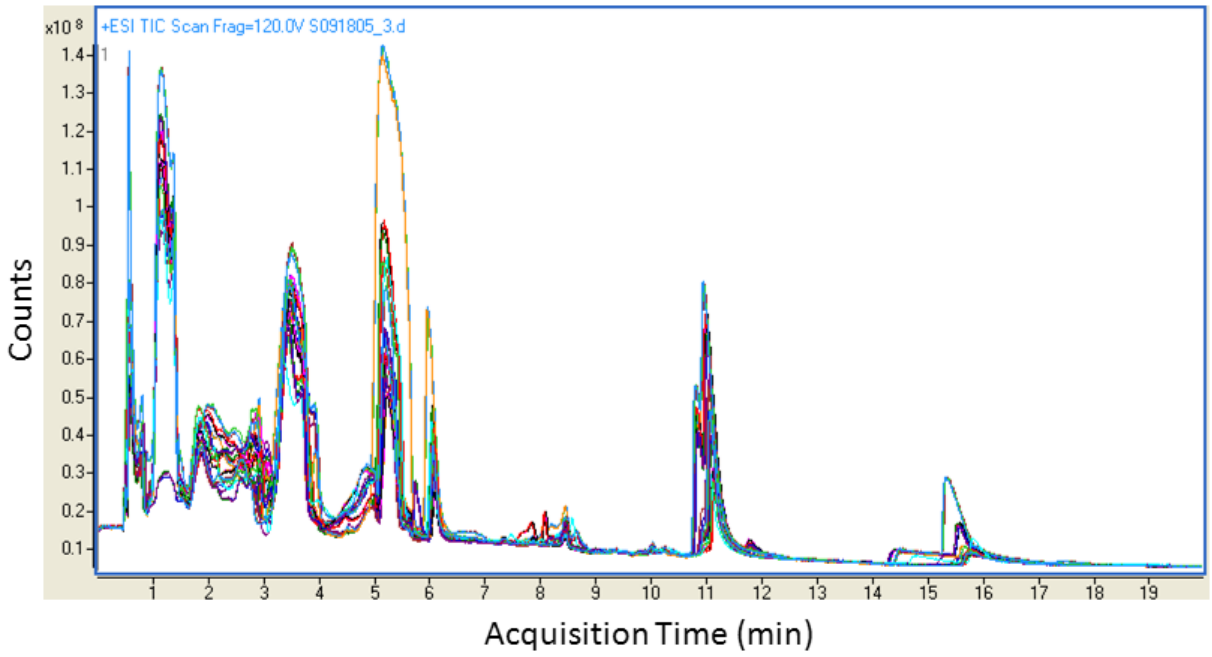


Figure 5-11. Overlay of 30 chromatograms of manatee urine from animals in anestrus using HILIC.

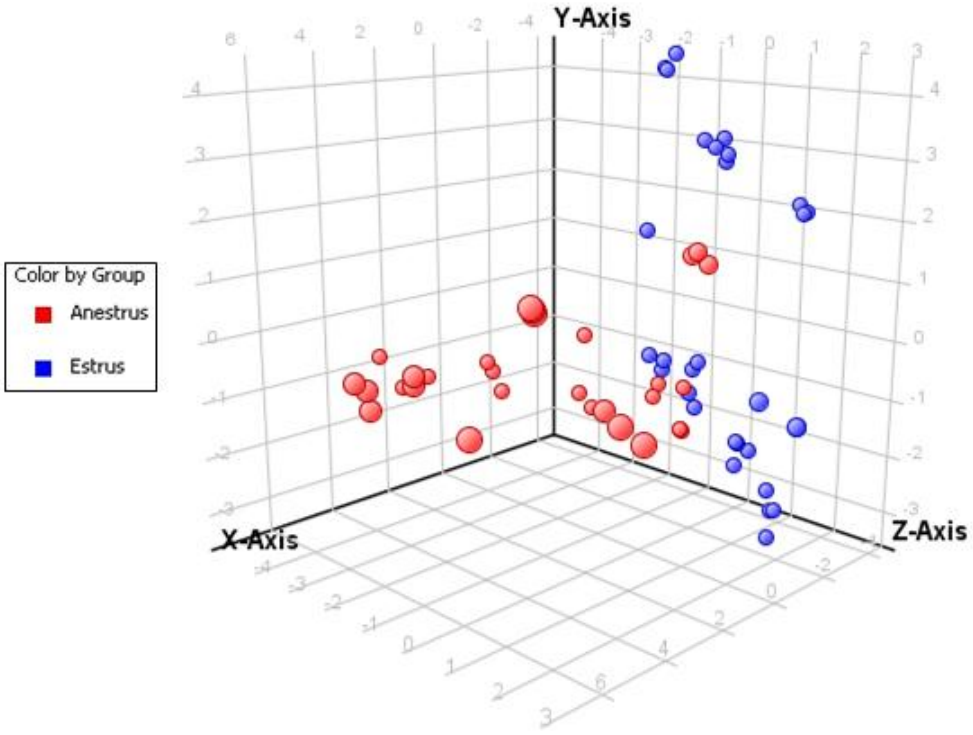


Figure 5-12. 3D-PCA plot of the significant masses in urine from manatees in estrus and anestrus by HILIC/MS (30 masses, p = 0.05). Each point represents one replicate.

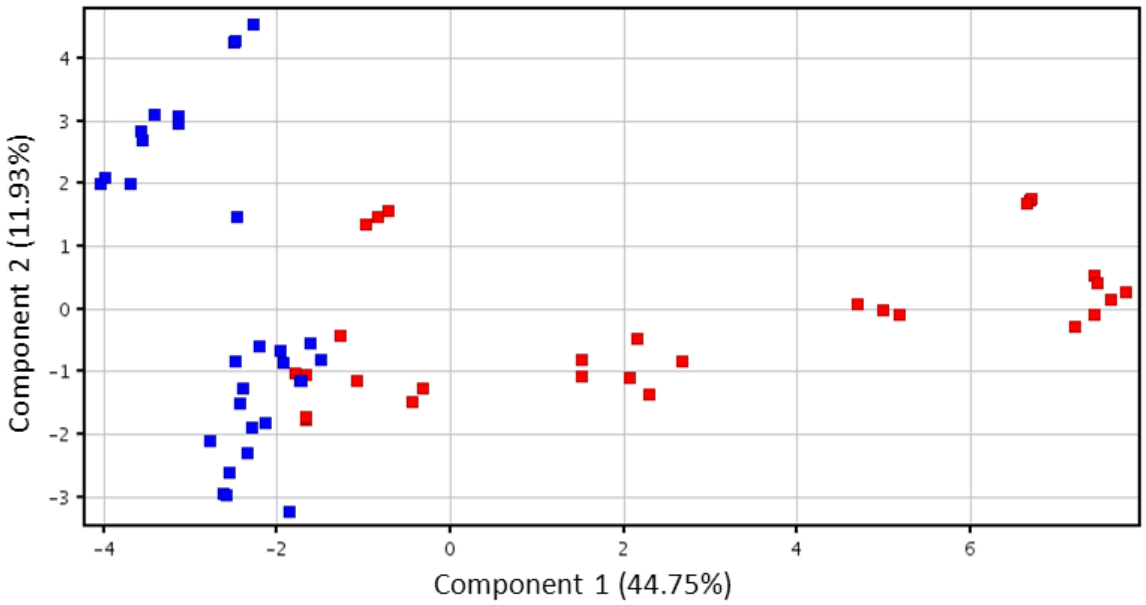


Figure 5-13. 2D-PCA plot of the significant masses in urine from manatees in estrus and anestrus (30 masses, p = 0.05) by HILIC/MS. Each point represents one replicate. The color code is the same as Figure 5-10.

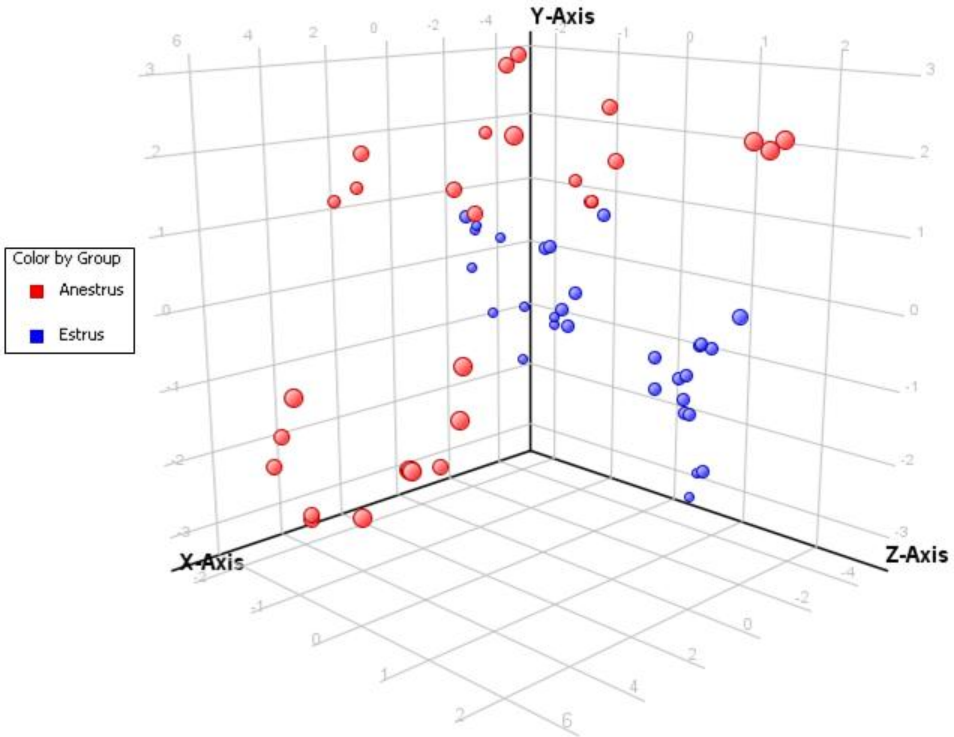


Figure 5-14. 3D-PLS-DA plot of t-test (p = 0.05) masses in urine from manatees in estrus and anestrus (30 masses) by HILIC/MS. Each point on the plot depicts one replicate.

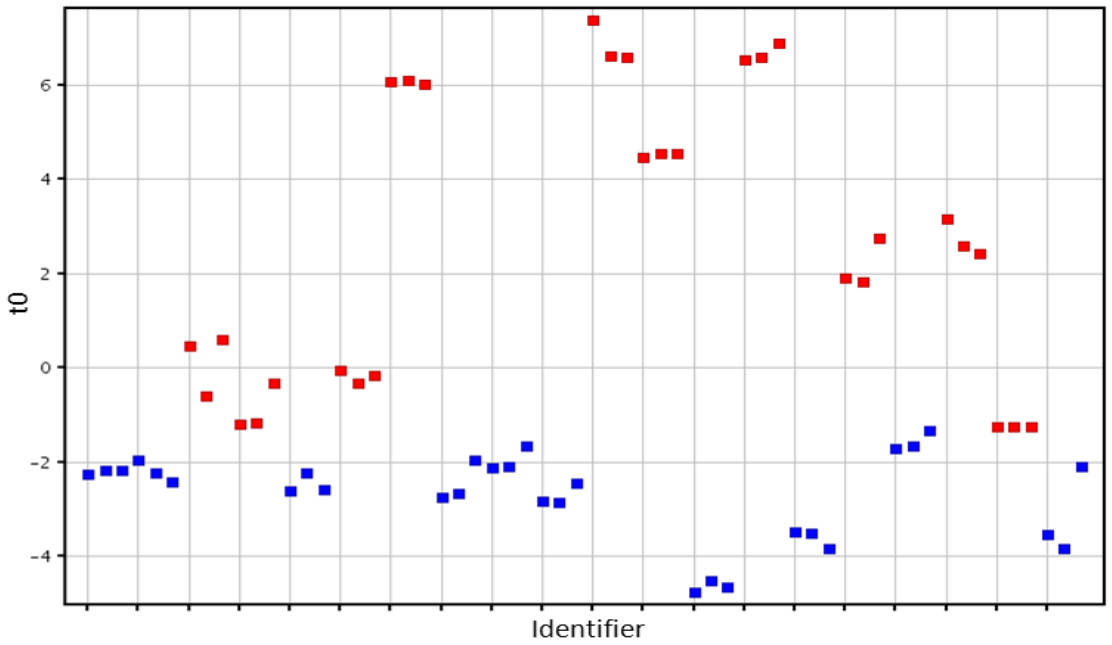


Figure 5-15. 2-PLS-DA plot of *t*-test (p = 0.05) masses in manatees in estrus and anestrus (30 masses) by HILIC/MS. Each point on the plot depicts one replicate. The x-axis (Identifier) values are individual sample replicates. The color code is the same as Figure 5-12.

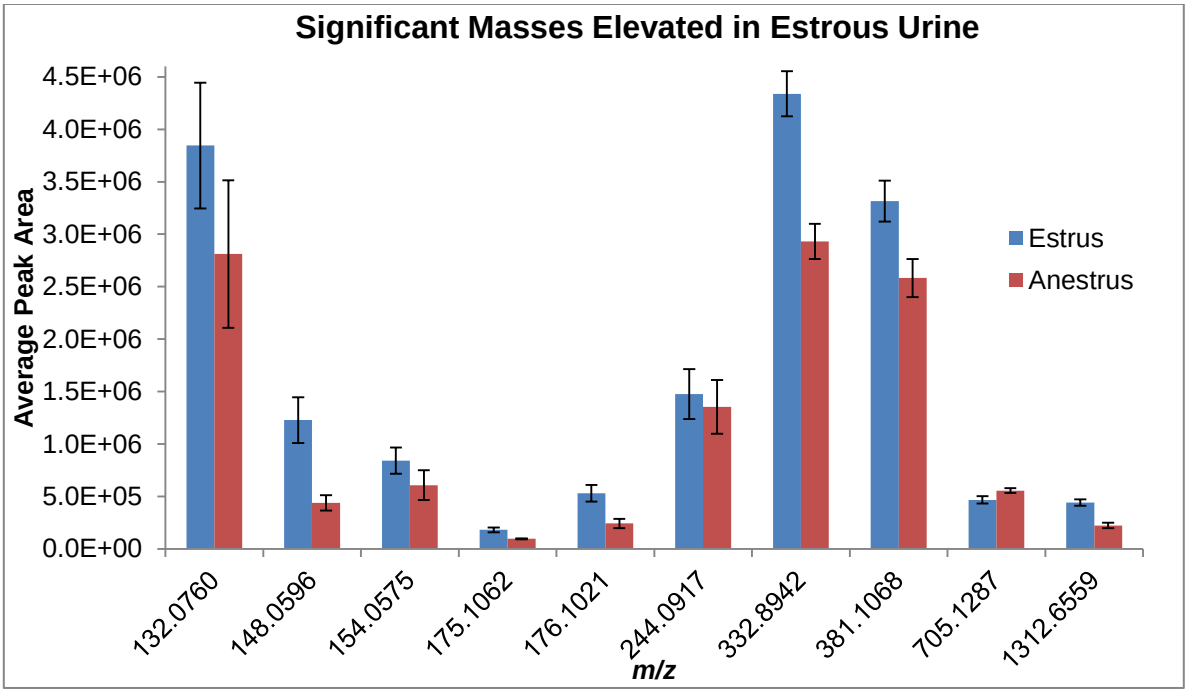


Figure 5-16. Bar graph comparing the average peak areas of the 10 significant masses elevated in estrous urine by HILIC/MS. Identifications were made for *m/z* values 132.0760, 148.0596, 176.1021, and 244.0917 in Table 5-3. Error bars indicate the standard deviation of the mean (n = 30).

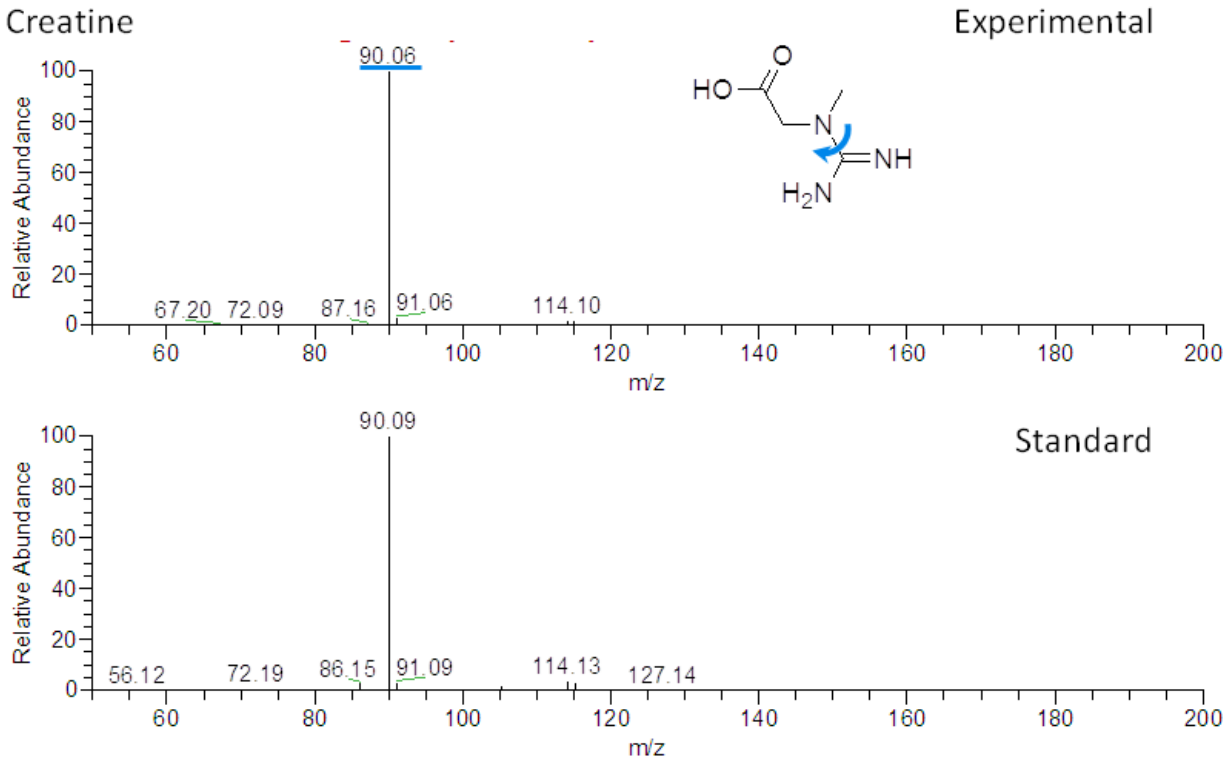


Figure 5-17. MS/MS spectra of *m/z* 132, determined by HILIC/MS, in estrous urine (top) and a synthetic standard (bottom). The ion at *m/z* 90 (loss of 42 mass units) indicates the loss of the CN_2H_2 group.

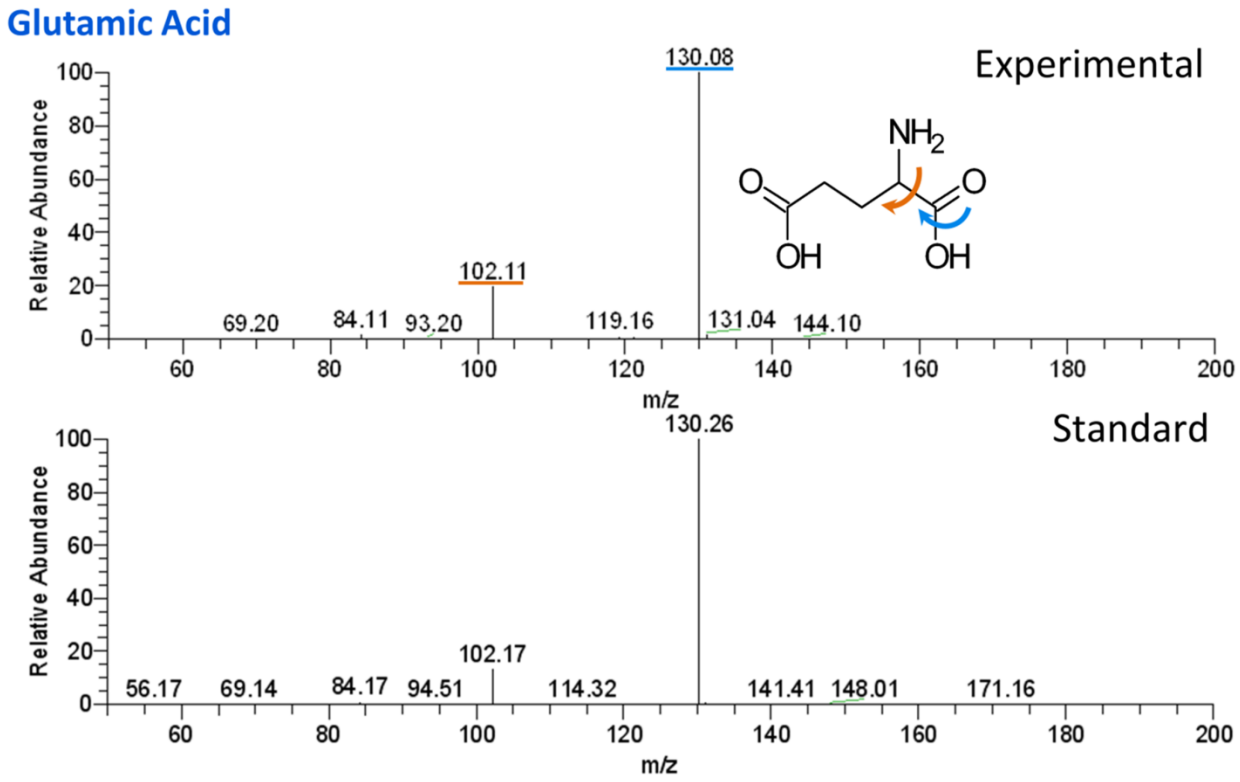


Figure 5-18. MS/MS spectra of *m/z* 148 in estrous urine (top) and a synthetic standard (bottom), determined by HILIC/MS. The ion at *m/z* 130 (blue arrow) is due to the loss of water and the ion at *m/z* 102 (loss of 46 mass units, orange arrow) is from the loss of a carboxylic acid.

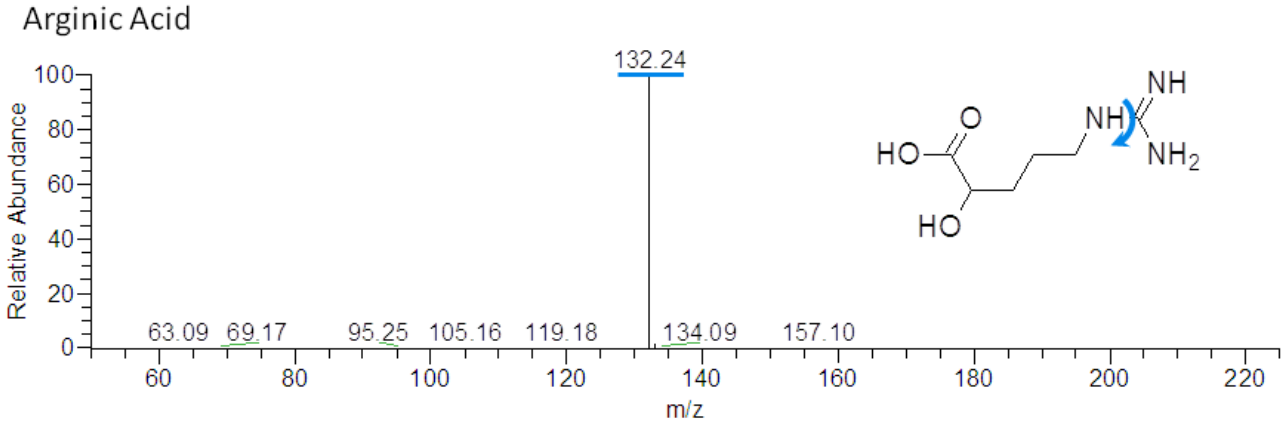


Figure 5-19. MS/MS spectrum of *m/z* 176 in estrous urine, determined by HILIC/MS. The ion at *m/z* 132 (blue arrow) is due to the loss of the CN_2H_4 group.

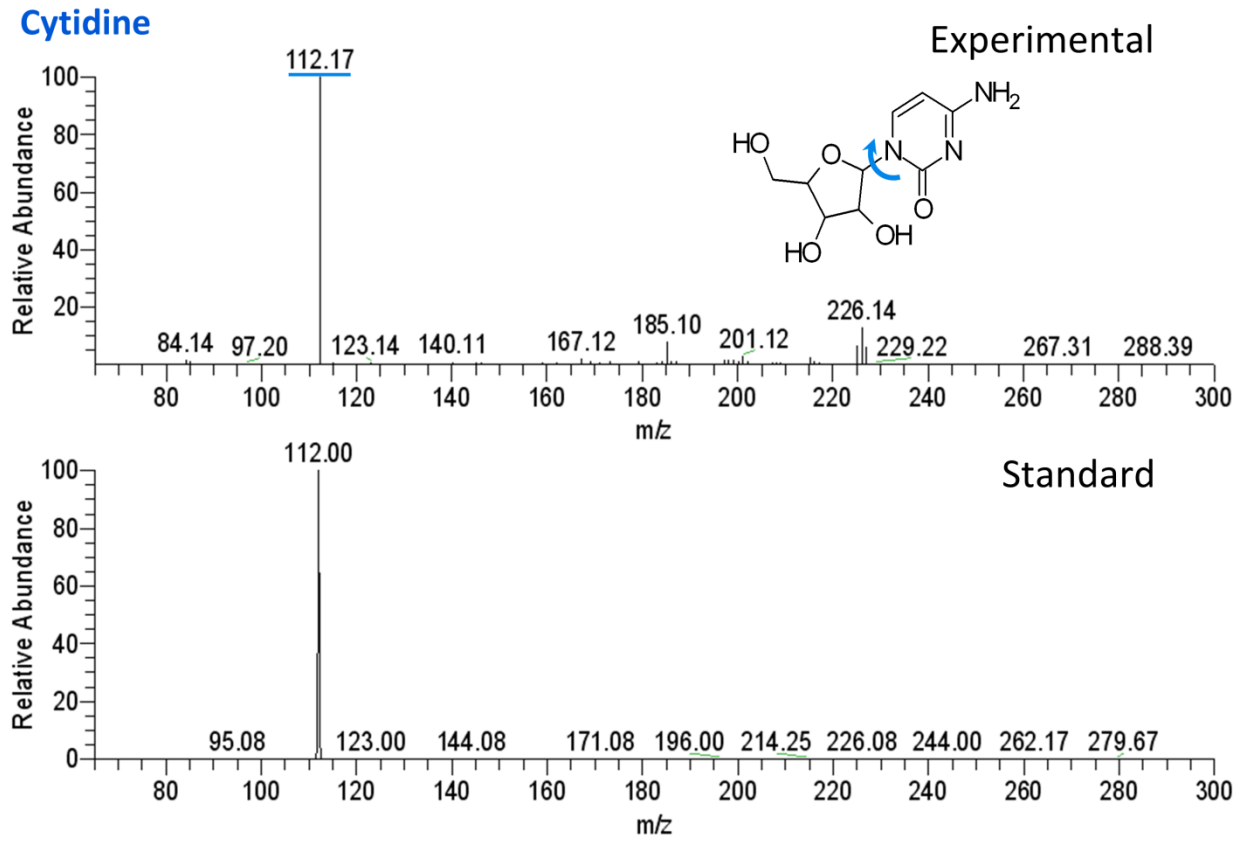


Figure 5-20. MS/MS spectra of *m/z* 244 in estrous urine (top) and a synthetic standard (bottom), determined by HILIC/MS. The ion at *m/z* 112 (blue arrow) is due to the loss of the oxygen containing ring.

CHAPTER 6
CONCLUSIONS AND FUTURE DIRECTIONS

Summary

The field of metabolomics is continually growing. However, bottlenecks in metabolomics still exist, which include data processing and metabolite identification. Robust software is necessary to process the complex data files generated from the analyses of biological samples in such large numbers. As shown in this work, identification of metabolites is very difficult. The expansion of metabolite and MS/MS libraries is necessary for easier identification using accurate mass and fragmentation patterns.

Piglet Plasma and RBC Comparison

This work has shown that the metabolomics of piglet plasma leads to more statistically significant changes over the first seven days of life compared to piglet RBCs. Piglet plasma led to 82 statistically significant masses while piglet RBCs only resulted in 23 statistically significant masses ($p = 0.01$). More carnitines were found to be present in plasma than RBCs. A general increase in carnitines was observed in plasma samples over seven days, whereas no trend was observed in the RBCs.

Future work could include this experiment being rerun with an equal number of samples and three replicates per sample. The result would be more beneficial to the metabolomics community, as (to the best of my knowledge) no one has examined the metabolomics of plasma compared to RBCs in any organism. If nothing else, it may conclude that little useful information can be determined from RBC samples compared to plasma. Additionally, work can be done to identify (MS/MS analyses) the significant masses determined by ANOVA in both the piglet plasma and RBCs.

Metabolomics of Colon Cancer

This work has compared the plasma of healthy human controls to plasma from patients with colon cancer. The potential biomarkers determined by statistics of the LC/TOF data were then analyzed by LC/LIT instrumentation. The MS/MS capabilities of the LIT, along with the accurate mass from the TOF allowed for structural elucidation and identification of five of the potential biomarkers as carnitines. No other studies have indicated these carnitines as potential biomarkers for colon cancer.

More work can be done to determine if these potential biomarkers are biologically relevant to the disease of colon cancer. Also, the identification of unknown significant masses can be investigated further.

Method Optimization for Metabolomic Applications

Methods for positive mode ESI were optimized in order to achieve the largest number of features. LC run time was decreased for a more time efficient method without injuring the resulting number of features. In addition to the LC run time being optimized, the results from this chapter were used to select the best LC column for further applications. A HILIC method was also optimized.

Future work may include the optimization of the APCI source. This would allow both ESI and APCI analysis, generating a larger number of potential biomarkers. Also, negative mode ESI and APCI can be implemented and optimized.

Metabolomics of Estrous and Anestrous Manatee Urine

Since many metabolites in urine are very polar, the RPLC/MS method resulted in all of the significant masses being at a retention time of about 0.3 minutes, or the void volume. This implies that the chromatogram was very congested in the first few minutes of analysis, leading to less sensitive MS detection, due to ion suppression, and

thus making statistical analyses more difficult. This was shown by the fact that only 9 metabolites were found to be significant by RPLC/MS whereas the HILIC/MS method led to 30 significant compounds. The HILIC/MS method allowed better separation but poorer peak shapes than the RPLC/MS method.

Cytidine, creatine, and glutamic acid were identified as significant compounds elevated in estrous urine. Two putative identifications of significant metabolites elevated in estrous urine were made with accurate mass and MS/MS, but without verification with a standard, as argininic acid and an unknown metabolite (formula calculation). However, none of these metabolites appear to be the pheromone of interest. The pheromone was expected to be present in only estrous urine, or at least very elevated. None of the elevated metabolites in estrous urine fit this description well. Although a pheromone was not found, no one has examined the metabolomics of manatee urine, thus any data collected is valuable and may help us better understand this endangered species.

Future work that would benefit this project includes introducing the male manatees to the elephant pheromone (not observed with +ESI) and noting the behavior, as well as acquisition of LC/MS by either negative mode ESI or by APCI. Additionally, targeted studies could be performed by gas chromatography/MS to see if the elephant pheromone is present in the manatee urine, although this may require special sample preparation.

.

LIST OF REFERENCES

(1) Atkinson, A. J.; Colburn, W. A.; DeGruttola, V. G.; DeMets, D. L.; Downing, G. J.; Hoth, D. F.; Oates, J. A.; Peck, C. C.; Schooley, R. T.; Spilker, B. A.; Woodcock, J.; Zeger, S. L. *Clin. Pharmacol. Ther.* **2001**, *69*, 89-95.

(2) Nass, S. J.; Moses, H. L. *Cancer Biomarkers: The Promises and Challenges of Improving Detection and Treatment*; The National Academies Press: Washington, D.C., 2007.

(3) Kaddurah-Daouk, R.; Kristal, B. S.; Weinshilboum, R. M. *Annu. Rev. Pharmacol. Toxicol.* **2008**, *48*, 653-683.

(4) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.

(5) Madsen, R.; Lundstedt, T.; Trygg, J. *Anal. Chim. Acta.* **2010**, *659*, 23-33.

(6) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155-171.

(7) Wikoff, W. R.; Gangoiti, J. A.; Barshop, B. A.; Siuzdak, G. *Clin. Chem.* **2007**, *53*, 2169-2176.

(8) Robertson, D. G. *Toxicol. Sci.* **2005**, *85*, 809-822.

(9) Nordstrom, A.; Want, E.; Northen, T.; Lehtio, J.; Siuzdak, G. *Anal. Chem.* **2008**, *80*, 421-429.

(10) Lin, Y.; Schiavo, S.; Orjala, J.; Vouros, P.; Kautz, R. *Anal. Chem.* **2008**, *80*, 8045-8054.

(11) Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2007**, *2*, 2692-2703.

(12) Viant, M. R. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943-948.

(13) Lewis, G. D.; Wei, R.; Liu, E.; Yang, E.; Shi, X.; Martinovic, M.; Farrell, L.; Asnani, A.; Cyrille, M.; Ramanathan, A.; Shaham, O.; Berriz, G.; Lowry, P. A.; Palacios, I. F.; Tasan, M.; Roth, F. P.; Min, J.; Baumgartner, C.; Keshishian, H.; Addona, T.; Mootha, V. K.; Rosenzweig, A.; Carr, S. A.; Fifer, M. A.; Sabatine, M. S.; Gerszten, R. E. *J. Clin. Invest.* **2008**, *118*, 3503-3512.

(14) Zeng, M.; Liang, Y.; Li, H.; Wang, M.; Wang, B.; Chen, X.; Zhou, N.; Cao, D.; Wu, J. *J. Pharm. Biomed. Anal.* **2010**, *52*, 265-272.

(15) Pereira, H.; Martin, J.; Joly, C.; Sébédio, J.; Pujos-Guillot, E. *Metabolomics* **2010**, *6*, 207-218.

(16) Gibney, M. J.; Walsh, M.; Brennan, L.; Roche, H. M.; German, B.; van Ommen, B. *Am. J. Clin. Nutr.* **2005**, *82*, 497-503.

(17) van Haandel, L.; Becker, M. L.; Leeder, J. S.; Williams, T. D.; Stobaugh, J. F. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3693-3702.

(18) Bruce, S. J.; Tavazzi, I.; Parisod, V.; Rezzi, S.; Kochhar, S.; Guy, P. A. *Anal. Chem.* **2009**, *81*, 3285-3296.

(19) Hewavitharana, A. K.; Hyde, C.; Thomas, R.; Shaw, P. N. *J. Chromatogr. B. Analyt Technol. Biomed. Life. Sci.* **2006**, *834*, 93-97.

(20) Villas-Bôas, S. G., Roessner, U., Hansen, M., Smedsgaard, J. and Nielsen, J. *Metabolome Analysis : An Introduction*; Wiley-Interscience: Hoboken, N.J, 2007.

(21) Wedge, D. C.; Allwood, J. W.; Dunn, W.; Vaughan, A. A.; Simpson, K.; Brown, M.; Priest, L.; Blackhall, F. H.; Whetton, A. D.; Dive, C.; Goodacre, R. *Anal. Chem.* **2011**, *83*, 6689-6697.

(22) Harris, D. C.; Freeman, W. H., Ed.; *Quantitative Chemical Analysis,* 6th ed.; Palgrave Macmillan: New York, 2002.

(23) Snyder, L. *High-Performance Gradient Elution : The Practical Application of the Linear-Solvent-Strength Model;* John Wiley: Hoboken, NJ, 2007.

(24) François, I.; Sandra, K.; Sandra, P. *Anal. Chim. Acta.* **2009**, *641*, 14-31.

(25) Horvath, C. G.; Lipsky, S. R. *Anal. Chem.* **1967**, *39*, 1893-1893.

(26) Lu, X.; Zhao, X.; Bai, C.; Zhao, C.; Lu, G.; Xu, G. *J. Chromatogr. B. Analyt Technol. Biomed. Life. Sci.* **2008**, *866*, 64-76.

(27) Svec, F.; Huber, C. G. *Anal. Chem.* **2006**, *78*, 2100-2107.

(28) Cabrera, K.; Lubda, D.; Eggenweiler, H.; Minakuchi, H.; Nakanishi, K. *Journal of High Resolution Chromatography.* **2000**, *23*, 93-99.

(29) Phenomenex. *http://www.phenomenex.com/cms400min/product.aspx?id=119* (accessed September 2, 2010).

(30) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Anal. Chem.* **1985**, *57*, 675-679.

(31) Gaskell, S. J. *Journal of Mass Spectrometry.* **1997**, *32*, 677-688.

(32) Cotter, R. E. *Time-of-Flight Mass Spectrometry*; American Chemical Society: Washington, DC; 1994.

(33) Agilent Technologies. *Publication Number 5990-9207EN*, 2011.

(34) Agilent Technologies. *Publication Number G3335-90028*, 2003.

(35) Blekherman, G.; Laubenbacher, R.; Cortes, D. F.; Mendes, P.; Torti, F. M.; Akman, S.; Torti, S. V.; Shulaev, V. *Metabolomics.* **2011**, *7*, 329-343.

(36) Wikoff, W. R.; Anfora, A. T.; Liu, J.; Schultz, P. G.; Lesley, S. A.; Peters, E. C.; Siuzdak, G. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3698-3703.

(37) Dervilly-Pinel, G.; Weigel, S.; Lommen, A.; Chereau, S.; Rambaud, L.; Essers, M.; Antignac, J.; Nielen, M. W. F.; Le Bizec, B. *Anal. Chim. Acta.* **2011**, *700*, 144-154.

(38) Dai, Y.; Li, Z.; Xue, L.; Dou, C.; Zhou, Y.; Zhang, L.; Qin, X. *J. Ethnopharmacol.* **2010**, *128*, 482-489.

(39) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779-787.

(40) Katajamaa, M.; Oresic, M. *J. Chromatogr. A.* **2007**, *1158*, 318-328.

(41) Bremer, M.; Doerge, R. W. *Statistics at the Bench : A Step-by-Step Handbook for Biologists*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y.; 2009.

(42) Smith, L. A. *www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf* (accessed on October 14, 2011).

(43) Fonville, J. M.; Richards, S. E.; Barton, R. H.; Boulange, C. L.; Ebbels, T. M. D.; Nicholson, J. K.; Holmes, E.; Dumas, M. *J. Chemometrics.* **2010**, *24*, 636-649.

(44) Miller, J.; Miller, J. *Statistic and Chemometrics for Analytical Chemistry* 5th ed.; Prentice Hall: New York; 2005.

(45) Verhoeven, K. J. F.; Simonsen, K. L.; McIntyre, L. M. *Oikos.* **2005**, *108*, 643-647.

(46) Benjamini, Y.; Hochberg, Y. *Journal of the Royal Statistical Society.Series B (Methodological).* **1995**, *57*, pp. 289-300.

(47) Hall, M. P.; Band, P. A.; Meislin, R. J.; Jazrawi, L. M.; Cardone, D. A. *J. Am. Acad. Orthop. Surg.* **2009**, *17*, 602-608.

(48) Ardrey, R. E. *Liquid Chromatography-Mass Spectrometry: An Introduction*; J. Wiley: West Sussex, England ; Hoboken, NJ; 2003.

(49) Agilent Technologies. *Publication Number G3335-90028*, 2003.

(50) Fjeldsted, J. Agilent Technologies. *Publication Number 5989-0373EN*, 2003.

(51) Schnackenberg, L.; Sun, J.; Espandiari, P.; Holland, R.; Hanig, J.; Beger, R. *BMC Bioinformatics.* **2007**, *8*, S3.

(52) Mersmann, H. J.; Pond, W. G. *Biology of the Domestic Pig*; Cornell University Press: Ithaca, N.Y., 2001.

(53) Fellman, V.; Lemmelä, S.; Sajantila, A.; Pihko, H.; Järvelä, I. *Journal of Human Genetics.* **2008**, *53*, 554-558.

(54) Chace, D. H.; Hillman, S. L.; Van Hove, J. L.; Naylor, E. W. *Clin. Chem.* **1997**, *43*, 2106-2113.

(55) Metz, T. O. *Metabolic Profiling: Methods and Protocols*; Springer: New York; 2011.

(56) Cerutti, E. S. *LC/MS-Based Targeted and Global Metabolomic Methodologies and their Application to Biomarker Discovery*; University of Florida: Gainesville, FL; 2009.

(57) The Scripps Research Institute. *http://metlin.scripps.edu/metabo_batch.php* (accessed on May, 8, 2011).

(58) Smith, C. A.; Maille, G. O.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*.

(59) American Cancer Society. *http://www.cancer.org/Research/CancerFactsFigures/CancerFactsFigures/most-requested-tables-figures-2012* (accessed on January 11, 2012).

(60) Kerr, D., Young, A. and Hobbs, R. *ABC of Colorectal Cancer*.; BMJ Books: London; 2001.

(61) Lefere, P.; Gryspeerdt, S. *Virtual Colonoscopy: A Practical Guide,* 2nd ed; Springer-Verlag: Berlin; 2010.

(62) American Joint Committee on Cancer; Edge, S. B. *AJCC Cancer Staging Manual,* 7th ed.; Springer: New York; 2010.

(63) American Cancer Society. *http://www.cancer.org/Cancer/ColonandRectumCancer/DetailedGuide/colorectal-cancer-staged* (accessed on January 17, 2012).

(64) O'Connell, J. B.; Maggard, M. A.; Ko, C. Y. *Journal of the National Cancer Institute.* **2004**, *96*, 1420-1425.

(65) National Cancer Institute. *http://www.cancer.gov/cancertopics/pdq/treatment/colon/Patient/page2* (accessed on January 13, 2012).

(66) Nambiar, P. R.; Gupta, R. R.; Misra, V. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis.* **2010**, *693*, 3-18.

(67) Chan, E. C.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C. *J. Proteome Res.* **2009**, *8*, 352-361.

(68) Mal, M.; Koh, P. K.; Cheah, P. Y.; Chan, E. C. Y. *Rapid Communications in Mass Spectrometry.* **2009**, *23*, 487-494.

(69) Warburg, O. *Science.* **1956**, *123*, 309-314.

(70) Hirayama, A.; Kami, K.; Sugimoto, M.; Sugawara, M.; Toki, N.; Onozuka, H.; Kinoshita, T.; Saito, N.; Ochiai, A.; Tomita, M.; Esumi, H.; Soga, T. *Cancer Res.* **2009**, *69*, 4918-4925.

(71) Denkert, C.; Budczies, J.; Weichert, W.; Wohlgemuth, G.; Scholz, M.; Kind, T.; Niesporek, S.; Noske, A.; Buckendahl, A.; Dietel, M.; Fiehn, O. *Mol. Cancer.* **2008**, *7*, 72.

(72) Piotto, M.; Moussallieh, F. -.; Dillmann, B.; Imperiale, A.; Neuville, A.; Brigand, C.; Bellocq, J. -.; Elbayed, K.; Namer, I. *Metabolomics.* **2009**, *5*, 292-301.

(73) Qiu, Y.; Cai, G.; Su, M.; Chen, T.; Zheng, X.; Xu, Y.; Ni, Y.; Zhao, A.; Xu, L. X.; Cai, S.; Jia, W. *J. Proteome Res.* **2009**, *8*, 4844-4850.

(74) Pekar Second, T.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. *Anal. Chem.* **2009**, *81*, 7757-7765.

(75) Lehtimäki, K. K.; Valonen, P. K.; Griffin, J. L.; Väisänen, T. H.; Gröhn, O. H. J.; Kettunen, M. I.; Vepsäläinen, J.; Ylä-Herttuala, S.; Nicholson, J.; Kauppinen, R. A. *Journal of Biological Chemistry.* **2003**, *278*, 45915-45923.

(76) McGarry, J. D.; Brown, N. F. *European Journal of Biochemistry.* **1997**, *244*, 1-14.

(77) Wright, M. H.; Heal, W. P.; Mann, D. J.; Tate, E. W. *J. Chem. Biol.* **2009**, *3*, 19-35.

(78) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263-269.

(79) Sana, T. R.; Waddell, K.; Fischer, S. M. *Journal of Chromatography B.* **2008**, *871*, 314-321.

(80) Crews, B.; Wikoff, W. R.; Patti, G. J.; Woo, H.; Kalisiak, E.; Heideker, J.; Siuzdak, G. *Anal. Chem.* **2009**, *81*, 8538-8544.

(81) Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 3289-3295.

(82) Spagou, K.; Wilson, I. D.; Masson, P.; Theodoridis, G.; Raikos, N.; Coen, M.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Nicholson, J. K.; Want, E. J. *Anal. Chem.* **2011**, *83*, 382-390.

(83) Gika, H. G.; Theodoridis, G. A.; Wilson, I. D. *J. Sep. Sci.* **2008**, *31*, 1598-1608.

(84) Cubbon, S.; Bradbury, T.; Wilson, J.; Thomas-Oates, J. *Anal. Chem.* **2007**, *79*, 8911-8918.

(85) Perrin, W. *Encyclopedia of Marine Mammals*; Elsevier Inc: Toronto, ON; 2009.

(86) Best, R. C. *Mamm. Rev.* **1981**, *11*, 3-29.

(87) Horikoshi-Beckett, C.; Schulte, B. A. *Zoo Biol.* **2006**, *25*, 285-301.

(88) Young Harper, J.; Schulte, B. A. *Zoo Biol.* **2005**, *24*, 135-144.

(89) Hartman, D. S. *Ecology and Behavior of the Manatee (Trichechus Manatus) in Florida*; American Society of Mammalogists: Pittsburgh, PA; 1979.

(90) Mann, D.; Colbert, D.; Gaspard, J.; Casper, B.; Cook, M.; Reep, R.; Bauer, G. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology.* **2005**, *191*, 903-908.

(91) Dehnhard, M. *International Zoo Yearbook.* **2011**, *45*, 55-79.

(92) Francis-Floyd, R.; White, J. R.; Chen, C. L.; Cardeilhac, P. T.; Cichra, C. E. *J. Aquat. Anim. Health.* **1991**, *3*, 70-73.

(93) Tripp, K. M.; Verstegen, J. P.; Deutsch, C. J.; Bonde, R. K.; Rodriguez, M.; Morales, B.; Schmitt, D. L.; Harr, K. E. *Theriogenology.* **2008**, *70*, 1030-1040.

(94) Carden, M.; Schmitt, D.; Tomasi, T.; Bradford, J.; Moll, D.; Brown, J. *Anim. Reprod. Sci.* **1998**, *53*, 133-142.

(95) Rasmussen, L. E. L.; Lee, T. D.; Zhang, A.; Roelofs, W. L.; Daves, G. D. *Chemical Senses.* **1997**, *22*, 417-437.

(96) Wyatt, T. D. *Nature.* **2009**, *457*, 262-263.

BIOGRAPHICAL SKETCH

Noelle Marie Elliott was born in Pottstown, Pennsylvania. While attending undergraduate school at Muhlenberg College in Allentown, Pennsylvania, she performed research under Dr. Christine Ingersoll. Noelle also participated in a Research Experience for Undergraduates (REU) program during the summer of 2007 at the University of Michigan in Ann Arbor, Michigan. At the University of Michigan, she worked on data analysis and extractions for metabolomic applications with Dr. Robert Kennedy. In May of 2008 Noelle graduated from Muhlenberg College with a Bachelor of Science in chemistry and a minor in mathematics. Noelle entered graduate school at the University of Florida (UF) in the fall of 2008, shortly after she married Nathan Elliott. At UF, Noelle performed research in the field of mass spectrometry, with applications in metabolomics, under the direction of Dr. David Powell and Dr. Richard Yost. In the summer of 2012, Noelle will graduate with a Doctor of Philosophy in chemistry.