2008

Characterization of Microorganisms of Interest to Homeland Security and Public Health Utilizing Liquid Chromatography/Mass Spectrometry

Robert A. Everley
Virginia Commonwealth University

CHARACTERIZATION OF MICROORGANISMS OF INTEREST TO HOMELAND SECURITY AND PUBLIC HEALTH UTILIZING LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry at Virginia Commonwealth University.

by

ROBERT ANTHONY EVERLEY
B.S., Eastern Kentucky University, 2003

Director: Timothy R. Croley
Adjunct Professor, Department of Chemistry

Virginia Commonwealth University
Richmond, Virginia
May, 2008

Acknowledgement

First, I would like to thank Dr. Tim Croley for his guidance and patience, and for taking me as his graduate student. I consider my time under his mentorship an honor and a blessing. In addition to the experience and knowledge of science received by working in his lab, he taught me how to *be* a scientist.

I would like to thank my committee for their input and for taking the time to challenge me and make me think about my work from different perspectives. Dr. Denise Toney is acknowledged for her advice and microbiology expertise. I would like to thank my lab mates, in particular Ricky Ciner for getting me started with liquid chromatography and for providing regular, unsolicited laboratory entertainment; and Tiffany Mott for being a patient microbiology resource, for preparing the bacterial samples and for her friendship. I would like to thank Dr. Sarah Porter for being a "big sister" in the chemistry department by giving me advice on chemometrics, cumulative exams and the proposal and dissertation writing process etc.

I would like to express gratitude to the Virginia Commonwealth University Department of Chemistry for giving me the opportunity to pursue further education and for the teaching assistantship. I am also grateful for the Virginia Division of Consolidated Laboratory Services for allowing me to conduct my research there as a part-time employee. I would like to thank the Kentucky Mafia for "taking care of one of their own". One of my first-year professors is acknowledged for telling me "you may not be

cut out for grad school". Those words, now proven false, were a constant motivation for me to do the best that I could possibly do from that point on.

I would like to express gratitude to my high school chemistry and physics teacher Arthur Hale for making science challenging and enjoyable at the same time, and inspiring me to pursue this joy professionally. Drs. Bob Fraas and Diane Vance are acknowledged for their support and assistance during my undergraduate education at Eastern Kentucky University.

I am indebted to my friends and family for their support and in particular, my parents. They taught me about hard work, discipline and respect, and have done more for me than I could ever do for them.

I am grateful for my wife, Rachel Star. She is the love of my life and has always believed in and supported me. I don't deserve her and can never repay her for her love and support, but thankfully I will spend the rest of our lives trying.

Lastly, I would like to give honor to God who made me for a purpose, gives me direction, and apart from whom I can do nothing. By His grace and wisdom He gave us an ordered universe; order allows experiments to be reproducible; reproducible experiments allow us to acquire knowledge; this knowledge we call science.

"Great are the works of the LORD; they are studied by all who delight in them."

- Psalms 111:2

Table of Contents

Page

List of Tables

Page

List of Figures

Page

List of Abbreviations

2D	Two dimensional
AAD	Antibiotic associated diarrhea
ACN	Acetonitrile
ATCC	American Type Culture Collection
BEH	Ethyl-Bridged Hybrid
BSA	Bovine serum albumin
CDC	Centers for Disease Control and Prevention
Da	Dalton
DCLS	Division of Consolidated Laboratory Services
DESI	Desorption electrospray ionization
dsDNA	Double stranded DNA
E	Electric field
EHEC	Enterohaemorrhagic *E. coli*
EIA	Enzyme immunoassay
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
ESI-MS	Electrospray ionization-mass spectrometry
FAB	Fast atom bombardment
FID	Flame ionization detector
FS	Fluorescence spectroscopy
GC/MS	Gas chromatography/mass spectrometry
GE	Gel electrophoresis
H	Plate height
HPLC	High performance liquid chromatography
HTLC	High temperature liquid chromatography
HUS	Hemolytic uremic syndrome
IAC	Immunoaffinity column
IPA	Isopropanol
LC	Liquid chromatography
LC/MS	Liquid chromatography/mass spectrometry
LC/MS/MS	Liquid chromatography/tandem mass spectrometry
LC/QTOF MS	Liquid chromatography/quadrupole time of flight mass spectrometry
LOD	Limit of detection
LOQ	Limit of quantification
LP1	Lysis protocol 1
LP2	Lysis protocol 2
M	Neutral mass
m/z	Mass/charge ratio

m_a	Mass of the adduct ion
MALDI	Matrix assisted laser desorption ionization
MaxEnt 1	Maximum Entropy
MRM	Multiple reaction monitoring
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS/MS	Tandem mass spectrometry
N	Separation efficiency
n	Number of charges attributed to each peak
NM	Non-motile
PCR	Polymerase chain reaction
PCR-MS	Polymerase chain reaction – mass spectrometry
PFGE	Pulsed-field gel electrophoresis
PSDVB	Polystyrene-divinylbenzene
PTM	Posttranslational modification
PVDF	Polyvinylidene fluoride
Q	Quadrupole
QCs	Quality controls
RF	Radio frequency
R_{FWHM}	Mass resolution at full width half maximum
RI	Relative intensity
RNase A	Ribonuclease A
Rs	Chromatographic resolution
SIM	Selected ion monitoring
SNP	Single nucleotide polymorphism
STEC	Shiga toxin producing *E. coli*
Stx	Shiga toxin
TFA	Trifluoroacetic acid
TOF	Time of flight
UHPLC	Ultra high pressure liquid chromatography
UNSCOM	United Nations Special Commission
UPLC	Ultra performance liquid chromatography
WHO	World Health Organization
z	Charge

Abstract

CHARACTERIZATION OF MICROORGANISMS OF INTEREST TO HOMELAND SECURITY AND PUBLIC HEALTH UTILIZING LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY

By: Robert Anthony Everley, Ph.D.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry at Virginia Commonwealth University.

Virginia Commonwealth University, 2008

Major Director: Timothy R. Croley
Adjunct Professor, Department of Chemistry

Liquid chromatography/quadrupole time of flight mass spectrometry (LC/QTOF MS) utilizing electrospray ionization was employed to monitor protein expression in *Escherichia coli* and *Shigella* organisms. Automated charge state deconvolution, spectral subtraction and spectral mirroring were used to reveal subtle differences in the LC/MS data. Reproducible intact protein biomarkers were discovered based on their unique mass, retention time and relative intensity. These markers were implemented to differentiate closely related strain types, (e.g. two distinct isolates of *E. coli* O157:H7) and to correctly identify unknown pathogens. Notable, was the distinction of multiple serotypes of enterohaemorrhagic *E. coli* which cannot be distinguished by clinical

manifestation alone. Additionally, speciation of *Shigella* was achieved, a task for which no commercial real-time polymerase chain reaction (PCR) primers exist.

This method was subsequently applied to two pathogenic *Clostridium* species: *C. difficile* and *C. perfringens*. Due to the increased difficulty during lysis, two new lysis protocols were developed, and each extracted a distinct set of proteins (by both mass and retention time). Extracts from both lysis procedures were utilized to discover biomarkers useful for identification and characterization at the species and strain levels. These biomarkers were successfully implemented to identify unknowns during a blind study and would enhance serological and genetic approaches by serving as new targets for detection. Two sets of the *C. perfringens* isolates that were deemed 100% similar by the gold standard for strain differentiation, pulsed-field gel electrophoresis (PFGE), were distinguished using LC/MS, demonstrating the high specificity of this approach.

The final part of this work demonstrated the application of ultra performance liquid chromatography (UPLC) to this project to improve the throughput of the method. Given that numerous small molecule applications of UPLC have been published, efforts were made to examine the potential of UPLC to enhance the separation of intact proteins. Beginning with typically employed conditions, column temperature and organic solvent were optimized followed by an HPLC vs. UPLC comparison. When applied to a mixture of ten protein standards, the optimized UPLC method yielded improved chromatographic resolution, enhanced sensitivity, and a three-fold increase in throughput. Application of this method to cell lysate analysis demonstrated no compromise in chromatographic or

mass spectral data quality; a reduction in run time from 75 minutes to 25 minutes was achieved.

CHAPTER 1 Overview and Objectives

Public health laboratories serve to aid in the event of exposure to harmful agents such as pathogenic bacteria. Exposure to these bacteria can occur naturally (e.g., foodborne illness) or intentionally (e.g., terrorist attack). Each year in the United States, an estimated 76 million foodborne illnesses cause over 325,000 hospitalizations and 5,200 deaths.[1] In 1997 alone, costs from foodborne illnesses in America were estimated at $35,000,000,000 based on medical expenses and lost productivity.[2] With this much harm occurring due to natural foodborne outbreaks, the intentional use of these agents would be more severe. The gravity of these natural occurrences and potential terrorist situations acts as a driving force for the development of improved techniques to enhance current capabilities to respond to exposure events/outbreaks, and assist with forensic and epidemiological investigations.

In the event of an infectious disease outbreak or a terrorist attack, key decisions by hospital staff, public health officials and investigating authorities will have to be made to guide medical treatment, prophylaxis and remediation. These decisions are best made when they are data directed. The method described here employs liquid chromatography/mass spectrometry (LC/MS) to find protein biomarkers useful for characterization and identification of bacteria at the species and sub-species levels (e.g. serotype and strain etc.). This approach is complementary and under certain conditions,

advantageous in comparison to current methods. Incorporating additional complementary techniques in conjunction with existing methods of detection and characterization will yield a better overall understanding of the data collected and will allow more confident data directed conclusions to be made.

Of particular interest to this study were bacterial select agents and food and waterborne pathogens. Select agents are microorganisms/toxins that are considered to have a high potential for serious illness and social disruption in the event of exposure. These agents are further categorized according to a combination of factors such as ease of dissemination, toxicity, etc. with Category A being the most potent including agents such as *Bacillus anthracis* and *Yersinia pestis*.[3] More simply, food and waterborne pathogens are pathogens which may be transmitted through contaminated food or water.

The goal of the LC/MS method was to detect proteins that were unique to certain bacterial pathogens and which could be used as biomarkers for identification and characterization. As with many biomolecular diagnostic approaches, the first step in the analysis was cultivation of the bacterial sample. After cultivation, a simple chemical lysis procedure was performed to inactivate the cells and extract proteins.

The proteins were then separated using reversed phase liquid chromatography. This type of chromatography gives hydrophobicity information in the form of retention time where the most retained (having larger retention times) proteins are more hydrophobic. Reversed phase chromatography also employs solvents that are easily evaporated and are more amenable to electrospray ionization-mass spectrometry (ESI-MS)[4] than the solvents typically employed in capillary electrophoresis and other

chromatography techniques such as ion exchange chromatography which tend to be aqueous solutions containing non-volatile salts. The effluent from the liquid chromatograph entered the directly into the ESI source of the mass spectrometer. Due to its sensitivity in full-scan mode and wide mass range, quadrupole time of flight[5] mass spectrometry was utilized for mass analysis.

ESI produces protein spectra that contain multiply charged ions and the interpretation of these spectra can be challenging even for a pure compound. To this end, an automated computer-based approach[6] was used to deconvolute the more than one hundred proteins detected per lysate. This process yielded a single spectrum containing the neutral (zero charge) masses of all of the observed proteins. Once the neutral masses are calculated, there are a total of three pieces of information that have been obtained: retention time, neutral mass and relative intensity.

During comparison, spectra from different bacteria were first mirrored along the baseline to better view the differences, and then protein masses common between the two bacteria being compared were removed using spectral subtraction. Spectral subtraction removes common masses within a given mass tolerance (± 2 Da here) leaving only the remaining unique masses which serve as potential biomarkers. Once detected, the biomarkers were evaluated for their reproducibility in terms of mass, retention time and relative intensity by repeating the entire experiment a minimum of three times on three different days. Due to the dynamic nature of the proteome and the production of artifacts during the charge state deconvolution step, not all unique masses observed after spectral subtraction were reproducible. Only proteins which were found unique on all three

repeated experiments were deemed biomarkers. Lastly, the marker proteins were further challenged for their utility during the identification of unknown pathogens in a blind study. This LC/MS method for biomarker discovery was applied in Chapters 4 and 5 and then optimized in Chapter 6.

In Chapter 4, the efficacy of the LC/MS method for characterization of bacterial pathogens was examined using *Escherichia coli* and *Shigella* species. These organisms are leading causes of the food and waterborne illness shigellosis which infects millions worldwide on an annual basis. Additionally, the toxin they produce, the Shiga toxin (Stx), is considered a category B select agent. Both qualitative (by mass and retention time) and quantitative (by relative abundance) biomarkers were discovered in this study and found to be reproducible in five repeated experiments run on five different days. These markers were implemented for *Shigella* speciation, and because no real-time PCR primers are available for this task, the gene sequence encoding these marker proteins could be used to design novel primers. Serotype and strain level discrimination was achieved allowing the distinction of three serotypes of enterohaemorrhagic *E. coli* and two different isolates of the same serotype (O157:H7), respectively. Strain level discrimination as demonstrated by this approach could be utilized during forensic traceback or food attribution investigations to pinpoint the source of an outbreak and to help determine if a group of infections is random or related. This would require examining clinical isolates from victims and isolates obtained from swabbing the crime scene or suspected food items and comparing the results to determine if a relationship

exists. The method developed and the biomarkers discovered identified over a dozen unknown bacteria in approximately two hours per sample post culture.

In Chapter 5, this work was expanded to include other types of bacteria (e.g., Gram positive and endospore forming). The *Clostridium* species *C. difficile* and *C. perfringens* were chosen because they are Gram positive, sporulate, and are of interest to public health and homeland security. *C. difficile* is the leading cause of the nosocomial illness antibiotic associated diarrhea (AAD) and has been stockpiled by a terrorist organization with the intent of deliberate use. *C. perfringens*, a well known foodborne pathogen, is also a cause of AAD and was stockpiled by the Iraq government's biological weapons program. Additionally, the epsilon toxin of *C. perfringens* is a category B select agent. Because they are Gram positive, the *Clostridia* have greater structural rigidity and therefore are more resistant to lysis than Gram negative bacteria such as *E. coli*.[7] For this reason, two new lysis methods were developed, with each extracting distinct proteins. The first protocol utilized a combination of lysozyme digestion and repeated cycles of freezing and thawing. The second utilized less lysozyme during the digestion step, but was followed by a treatment with acetonitrile and trifluoroacetic acid after the freeze/thaw step.

After lysis, the spores were removed by filtration and the extracts were analyzed using the LC/MS method described above. Reproducible biomarkers were observed that could distinguish both *Clostridium* species, something that in the case of AAD can not be achieved by symptoms alone. Markers from both lysis protocols were utilized to identify unknown isolates and to distinguish strains of each species. Strain level distinction of *C.*

difficile could be used to direct treatment, as highly virulent drug-resistant strains have been reported. In cases where isolates whose PFGE restriction patterns were indistinguishable, differentiation was achieved using LC/MS. These findings encourage modification of the current PFGE protocols used in this laboratory to include different or multiple restriction enzymes for added information and specificity.

Finally, the work described in Chapter 6 incorporates the latest advancement in liquid chromatography, i.e., ultra performance liquid chromatography (UPLC)[8], in an attempt to improve the overall method throughput. UPLC employs small (sub-2 μm diameter) porous particles at very high back pressures to generate more efficient separations in less time. As predicted from Van Deemter theory, the plate height minimum in the plot of height equivalent to a theoretical plate vs. linear velocity is lower for sub-2 μm particles than for larger particles sizes while the linear velocity region at that minimum is also much longer.[9] This means the use of sub-2 μm particles (1.7 μm here) allows for improved resolution over a wide range of flow rates. The runtime can then be significantly reduced because the flow rate can be significantly increased, at no expense in resolution. However, since the column diameter remains constant, as the flow rate is increased, there is a concomitant increase in back pressure. Thus, UPLC is a combination of sub-2 μm particles used with a column and system that can withstand much higher pressures (e.g. 1000 bar) than those utilized in conventional high performance liquid chromatography (HPLC) experiments.

One novel aspect of this work is the application of UPLC to the separation of intact proteins. Due to their slow diffusivities and greater structural complexity relative

to small molecules, intact proteins have been problematic to chromatograph.[10] This work utilized ten protein standards of varying size and reversed phase retention behavior as a model set for method development. In addition to particle size, column temperature and organic modifier were adjusted for optimal performance. The optimal method involved a combination of high temperature, a solvent of high eluotropic strength and small particles at ultra high pressure. This method was then applied to an *E. coli* cell lysate and demonstrated improved sensitivity, and a threefold reduction of data collection and data analysis time compared to that of the original HPLC method.

The work performed and the ensuing results from these three projects will be discussed at greater length in Chapters 4-6. Necessary background discussion of the existing methods for bacterial identification and characterization is described in Chapter 2. Chapter 3 provides an introduction to and the rationale behind the protocols within the LC/MS method which was later employed in Chapters 4 and 5, and optimized in Chapter 6. Finally, insights, implications and potential future work resulting from this dissertation are detailed in Chapter 7. The work presented here represents advancement in the study of diagnostic microbiology as it relates to public health and homeland security, built upon a foundation in analytical chemistry and achieved through the application of novel, cutting edge technology.

CHAPTER 2 Current Methods of Bacterial Characterization

2.1. Cultural Methods

Since the initial observation of bacteria in 1683 by Antoni van Leeuwenhoek[11], the means by which bacteria are observed and characterized has experienced tremendous advancement. In the modern public health domain, three approaches dominate: culture-based, serological and genotypic. The culture of bacteria is a way of purifying the bacteria from its matrix and increasing its concentration by growth. The time required to reach the desired concentration depends on the generation time (doubling time) and growing conditions (temperature, atmosphere and nourishment etc.). At optimal conditions, the generation time can vary widely depending on the bacterium. *Escherichia coli* for example has a generation time of 17 minutes compared to 1980 minutes (33 hours) for *Treponema pallidum.*[12] Techniques that have been traditionally used to identify bacteria post-culture include morphologic, chemotaxonomic, and biochemical tests, most of which require several days to perform. The morphology of bacteria (shape, size, stain response, etc.) viewed with a microscope is a useful tool for grouping bacteria. Chemotaxonomic methods involve analysis of the chemical content of bacterial cells. Biochemical approaches monitor the bacteria's ability to ferment, metabolize, or cleave certain compounds.[13]

While it is desirable for detection methods to be sensitive enough so that culture is not required (in order to improve assay speed), the majority of accepted methods in diagnostic bacteriology culture the bacteria prior to analysis. In addition to purification and concentration, there are other distinct advantages to culture such as the differentiation of viable and non-viable cells; and cell stress or injury resulting from food processing or the environment can be repaired.[14]

The culture process itself has diagnostic value. There are two principal methods of culture-based diagnosis, both dependant upon the type of media used for growth – selective or differential. The selective technique attempts to grow unknown samples on different types of media, with each type designed to enhance the growth of a particular kind of organism while containing inhibitors to inhibit the growth of others. The ensuing results are then compared to the growth trends of known bacteria for identification. The differential approach cultivates bacteria on a nonselective media but in the presence of several indicator dyes so that different types of bacteria will be identified based upon the different colors they produce.[13] One very common medium, MacConkey agar, functions as both selective and differential. This medium contains inhibitors for the growth of Gram positive bacteria, thus selecting for Gram negative; and contains lactose and a pH indicator dye that turns red under acidic conditions. Since acid is produced during lactose fermentation, the ability of unknown bacteria to ferment lactose is discerned based on the color (pH) of the medium after growth. Thus, common foodborne pathogens such as *E. coli* which do ferment lactose are readily distinguished from *Salmonella* which do not. While culture-based techniques are useful for classifying

unknown bacteria and can serve to direct further testing, they are generally not specific enough to differentiate between species or strains of bacteria and are not considered confirmatory techniques.

2.2. Serological Approaches

Serological/immunological assays employ antibodies that react with antigens produced by, or located on, the surface of bacteria. These antigens may be lipopolysacharides or proteins for example. Bacteria of the same species with different antigens are known as serotypes. Bacteria are often described by their serotype as in the case of *E. coli* O157:H7 which is O (lipopolysacharide) antigen number 157 and H (flagellar protein) antigen number 7. Since only certain serotypes of bacteria such as *E. coli* may be toxic, serological methods are often used to distinguish between toxic and non-toxic bacteria. There are two main serological techniques used to identify bacteria: particle agglutination and enzyme-linked immunosorbent assays.

Particle agglutination assays adsorb antibodies to particles such as latex and as multiple particles react with antigens on the same bacterium, the bound particles clump together (agglutinate) to form a mass of particles that is visible to the naked eye. The greater the size of the particles or beads (typically > 800 nm), the greater the ease in which the agglutination reaction can be observed which may take up to 24 hours. Often culture is not required for this technique; but contaminates can adversely effect the reaction so often sample clean up is necessary.[15]

Unlike agglutination assays which are used to detect intact bacterial cells using their surface antigens, enzyme-linked immunosorbent assay (ELISA) approaches can also

be used to detect individual proteins such as toxins expressed by bacteria. ELISA, is an indirect immunofluorescence assay where the bacteria adhere to a microtiter plate before reacting with an unlabeled primary antibody. Next, the secondary antibody which binds to the primary antibody, is coupled to an enzyme which after removal of unbound antibody, cleaves a dye substrate added to the solution resulting in a color change in the dye.[13] Unlike agglutination assays which are used to detect intact bacterial cells using their surface antigens, ELISA approaches can also be used to detect individual proteins such as toxins expressed by bacteria.

The serological approaches are sufficiently sensitive such that culture is often not required and analysis time is rapid (≤ 1 day). However, as with all antibody approaches, cross reactivity between closely related samples can occur, and these techniques are only applicable for bacteria to which antibodies have been made.[16]

2.3. Genetic Approaches

2.3.1. Real-Time Polymerase Chain Reaction

While culture and serology yield useful information, ultimately a genetic approach is desired for confirmation of the microorganism's identity. First presented by Siaki et al. in 1988[17], polymerase chain reaction (PCR), is a method used to amplify small amounts of DNA by making repeated copies of specific strands that can then be detected by gel electrophoresis or fluorescence spectroscopy.[13] This technique is used to characterize and identify microorganisms by monitoring sequences of DNA unique to a particular genus or species. Starting at the location of the primer, a replicate double stranded DNA (dsDNA) molecule identical to the original one is produced. The process

is then repeated, but this time with double the starting material and is thereby an exponential process. This feature makes PCR a very sensitive technique and led to a Nobel Prize in chemistry in 1993. In theory culturing is not required, however in practice, the task of isolating the cells so that only bacterial DNA is used and purifying of the cells from inhibitors in the matrix which can interfere with the polymerase reaction is required. Purification can be accomplished by using immunoaffinity beads[18] but these require antibodies and for specificity reasons stated previously, most often cultural enrichment is employed.

The cycling process can continue until a desired concentration is obtained or until the reactants are consumed and the reaction plateaus. Some attempts to quantify the DNA produced at the end of the reaction have been made but it is difficult to correlate the amount of DNA during the plateau stage with the amount of starting material. For this reason, Higuchi et al. used a fluorescent probe that emits signal only when bound to dsDNA.[19] The fluorescence intensity and therefore the reaction can be monitored in 'real-time' as dsDNA is continually produced. From the number of cycles required to reach the plateau phase, the amount of DNA at the beginning of the reaction can be calculated. This quantitative PCR is commonly referred to as real-time PCR.

Real-time PCR is categorized as a rapid technique because it can often be performed after culturing in 0.5 - 4 hours. While sensitivity and speed are advantages, some limitations are also present. One, this method requires that primers be available for use. Bacteria with out available primers are not suitable for detection by real-time PCR. Two, since it is a targeted approach the information obtained from the primers may be

unreliable or insufficient if bacteria were genetically engineered for such purposes as drug or heat resistance or had undergone natural or environmental mutagenesis. Finally, small genetic differences unique to one strain of bacteria such as single nucleotide polymorphisms (SNPs) could occur between the primer locations and be missed during the comparison of two closely related strains.[20]

2.3.2. Pulsed-field Gel Electrophoresis

Another genomic approach that is widely used to characterize bacterial pathogens is pulsed-field gel electrophoresis (PFGE). First presented in 1983, this technique separates DNA fragments using gel electrophoresis in a pulsed electric field and has three main steps.[21] First, after culture of bacteria, the cells are lysed and the intact chromosome is removed. Next, using a restriction endonuclease enzyme, the chromosome is digested into fragments (usually 10-20) of DNA that range in size from 10-1000 kb (~ 3.13 – 313 MDa). Finally, these fragments are electrophoretically separated according to size using a pulsed-field to create a pattern unique to that isolate of bacteria.[22, 23]

The electric field is pulsed in varying angles at different time intervals to better separate the large DNA fragments which are not efficiently resolved using traditional electrophoretic techniques. Using a traditional (continuous) field, these large chromosomal fragments will elute together as a single band. But upon varying (pulsing) the field at different angles and at different time intervals, the enough change in the mobility the fragments occur to effect separation. Computer images of the gel patterns, also known as restriction patterns, are then compared to historical samples or samples from other laboratories etc. Highly similar or identical patterns indicate two isolates are

related and differences in the restriction patterns signify differences in nucleotide sequence.[22]

PFGE is considered a non-targeted screening approach and is therefore able to detect subtle differences missed by targeted approaches such as serology and real-time PCR. The specificity achieved by PFGE has made it the 'gold standard' technique for strain typing and it is widely used for epidemiological purposes. Outbreaks are detected and monitored through a program known as PulseNet which compares and certifies PFGE patterns from public health laboratories around the country.[24]

The advantages of PFGE are that primers or antibodies are not required for analysis and that strain-level typing is straightforward. However, because optimal digestion for different types of bacteria require different restriction enzymes, prior knowledge of the genus of the sample is required. Another limitation common to all genotypic approaches is that the mere presence of a gene doesn't necessarily correlate to protein expression. This consideration is of particular importance for bacterial pathogens which are known to contain genes that are not expressed[25-27] which could result in a false positive indication of pathogenicity. Furthermore, PFGE is not a rapid technique since it requires a minimum of two days to perform after culturing.

2.4. Mass Spectrometric Approaches

2.4.1. Small Molecule Analysis

Mass spectrometry was first applied as a tool for the identification and characterization of microorganisms in the 1970's.[28] Early approaches typically involved gas chromatography/mass spectrometry (GC/MS) both for the information provided by

chromatography as well as the added specificity and sensitivity afforded by MS compared to a flame ionization detector (FID). Furthermore, using tandem mass spectrometry (MS/MS) or selected ion monitoring (SIM) the background noise was greatly reduced because only ions of interest were monitored increasing sensitivity. Additionally, a GC/MS technique employing high temperature pyrolysis of intact bacteria to introduce volatile compounds into the gas phase has also been utilized.[28] The early mass spectrometric approaches typically were chemotaxonomic in nature and monitored small volatile molecules such as lipids and metabolites.[20] For compounds that were less volatile or were too polar, such as carbohydrates, derivatization reactions could be employed to make them more amenable for GC analysis. Some specific techniques utilized for speciation include monitoring the fatty acid content from phospholipid backbones and detecting carbohydrate profiles.[20] These fatty acid or sugar profiles were then compared with other bacteria either manually or with the use of spectral libraries.[28] Additionally muramic acid, one of the constituents of bacterial peptidoglycan, which is not synthesized by mammalian enzymes, can be monitored by GC/MS to detect the presence of bacteria in environmental or clinical samples and can also distinguish bacterial from viral infections.[20] To avoid the time consuming step of derivatization and to monitor phospholipids directly, fast atom bombardment (FAB) mass spectrometry was used in the early 1990's due to its ability to analyze polar compounds.[29] More recent techniques such as desorption electrospray ionization (DESI) have been used to discover sub-species distinctions in *Salmonella typhimurium* and *E. coli* by monitoring lipid profiles.[30]

The analysis of small molecules such as lipids and carbohydrates can be problematic because many of those compounds are not unique to a particular strain of bacteria (unlike certain DNA or protein sequences) therefore limiting their overall discriminatory power as biomarkers. Examples include fatty acids, some which are shared over a wide range of bacteria, phospholipids which occur throughout nature therefore possibly leading to background contamination or false positives, and hydroxy fatty acids which are found in normal blood and tissues and could interfere during clinical analysis.[20] Furthermore, techniques that deal more directly with the function of the organism such as genomic or proteomic information are more likely to gain acceptance in the microbiology community.

2.4.2. Large Molecule Analysis

The advent of matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) allowed soft ionization of highly polar compounds and both have become the preeminent ionization techniques for the analysis of large biomolecules e.g., DNA, by mass spectrometry.[20] One example of the use of mass spectrometry to classify bacteria by their DNA, is polymerase chain reaction – mass spectrometry (PCR-MS).[20] As described earlier, when PCR amplifies DNA, the products (amplicons) are monitored by fluorescence spectroscopy (FS). However, the low level of specificity obtained by FS makes it difficult to detect small differences in the nucleic acid sequence such as a single A to T or G to C switch (SNPs) which yields mass differences of 9 and 40 Da, respectively. Mass spectrometry offers enhanced specificity thereby minimizing

false positives. This technique has been used to distinguish two species of *Bacillus* using the same PCR primers.[31]

MALDI and ESI have furthered the mass spectrometric analysis of proteins as well. Interestingly, key differences between closely related strains may occur at the proteome level that do not occur at the genome level. In fact, the data collected from proteomic approaches is much more than just an indirect genetic analysis. This is because the proteome is one to two orders of magnitude larger than what the genome encodes.[32] If an organism's genome contains 20,000 genes, the number of unique proteins in that organism may be as many as 200,000 to 2 million – information lost by detecting DNA only. This large difference between the size of a genome and its respective proteome is due to the various isoforms of proteins which are largely a result of posttranslational modification (PTM). As their name implies, PTMs occur after the genetic contribution to protein structure is complete. Examples of PTMs include the adding of carbohydrates (glycosylation) or lipids (prenylation). One of the dominant PTMs in bacteria is removal of the *N*-terminal methionine and is estimated to occur in 50% of bacterial proteins.[33] In total, the differences that can be observed during the comparison of bacterial proteins can be qualitative: the presence/absence of proteins, protein sequence mutations and PTMs, or quantitative, i.e., varying levels of protein expression.[34]

Often proteomic analysis by MS is coupled with a separation technique such as liquid chromatography (LC) and two LC/MS workflows are commonly employed. The first, known as the bottom-up approach, uses a protease such as trypsin to digest all of the

proteins in the cell lysate.[35] The resulting peptides are then analyzed by two dimensional (2D) LC and MS and subsequent database searches are used to match the peptide masses to known peptide sequences. The second approach is top-down, and first isolates intact proteins of interest either by LC/MS or 2D gel electrophoresis so that only the proteins of interest are sequenced.[36] The last step of sequencing is performed either inside the mass spectrometer or by analyzing peptides after proteolytic digestion.[35, 36]

The bottom-up approach is more time consuming because of the digestion (about 2 hours), the 2D LC separation and the added data analysis time due to the large number of peptides produced. The method described here is a top-down approach where LC/MS was used to identify the unique proteins rather than 2D gel electrophoresis. LC/MS is automated and has improved dynamic range, resolution and reproducibility and is less time consuming than gel electrophoresis.[37] Moreover, the combination of mass and retention time information means LC/MS could be considered a 2D approach. The top-down approach has the advantage of detecting the mass of the intact protein prior to sequencing. Knowing the intact mass allows immediate detection of modified proteins during comparative proteomic investigations.

The additional step of sequencing and database searching was not employed in the method described here. The reason being unique and reproducible biomarkers and protein profiles were observed allowing for identification at the species and strain levels without knowing the actual identity of the proteins involved. Therefore, this approach could potentially be applied to bacteria whose genomes have not been sequenced. In contrast, proteomic approaches that rely upon database results for

identification purposes are of little to no use for such bacteria. Finally, not sequencing the proteins avoids a timely digestion step which resulted in a reduced analysis time.

CHAPTER 3 Overview of Analytical Procedures

3.1. Cultivation of Bacterial Cells

The bacterial pathogens that were investigated are: *E. coli*, *Shigella* species and *Clostridium* species. Shiga toxin producing *E. coli* and *Shigella* are foodborne pathogens, and the toxin they produce is a Category B select agent. These two pathogens cause the enteric disease Shigellosis. Ubiquitous in the environment, *Clostridium* species are toxin producers and cause illnesses such as gangrene and pseudomembranous colitis, and are a common cause of nosocomial infection.[38] In addition to the characteristics used to identify these microorganisms as select agents and or foodborne pathogens; other traits commonly used for their characterization are listed in Table 1.

Table 1. Characteristics used to describe the bacteria of interest to this study.

Bacteria	Gram +/-	Aerobic	Shape	Sporogenic	Motile
Shigella spp.	-	Yes	Bacillus	No	No
E. coli	-	Yes	Bacillus	No	Yes
Clostridium spp.	+	No	Bacillus	Yes	Yes

The bacteria were obtained from clinical isolates stored at the Virginia Division of Consolidated Laboratory Services (DCLS) or were purchased from American Type

Culture Collection (ATCC, Manassas, VA). Since no official nomenclature for describing strains exists, in-house accession numbers were used to describe the clinical isolates as different strains and the commercial strains were distinguished by their ATCC product number Table 2.

Table 2. The isolates used in this study. NM = non-motile, NA = not applicable, ND = not determined.

Family	Genus	Species	Serotype	Strain
Enterobacteriaceae	*Escherichia*	*coli*	O157:H7	06-1464
Enterobacteriaceae	*Escherichia*	*coli*	O157:H7	06-1439
Enterobacteriaceae	*Escherichia*	*coli*	ND	06-0004
Enterobacteriaceae	*Escherichia*	*coli*	ND	06-0006
Enterobacteriaceae	*Escherichia*	*coli*	O111:NM	06-1440
Enterobacteriaceae	*Escherichia*	*coli*	O26:H11	06-1418
Enterobacteriaceae	*Shigella*	*sonnei*	NA	06-1364
Enterobacteriaceae	*Shigella*	*sonnei*	NA	06-1362
Enterobacteriaceae	*Shigella*	*flexneri*	ND	04-0497
Enterobacteriaceae	*Shigella*	*flexneri*	ND	06-0967
Clostridiaceae	*Clostridium*	*perfringens*	ND	06-0385
Clostridiaceae	*Clostridium*	*perfringens*	ND	06-0387
Clostridiaceae	*Clostridium*	*perfringens*	ND	05-0025
Clostridiaceae	*Clostridium*	*perfringens*	ND	05-0070
Clostridiaceae	*Clostridium*	*perfringens*	ND	05-0076
Clostridiaceae	*Clostridium*	*perfringens*	ND	04-1464
Clostridiaceae	*Clostridium*	*perfringens*	ND	04-1672
Clostridiaceae	*Clostridium*	*perfringens*	ND	04-1665
Clostridiaceae	*Clostridium*	*difficile*	A	43594
Clostridiaceae	*Clostridium*	*difficile*	B	43593
Clostridiaceae	*Clostridium*	*difficile*	C	43596
Clostridiaceae	*Clostridium*	*difficile*	F	43598
Clostridiaceae	*Clostridium*	*difficile*	ND	700792

All bacteria were grown on trypticase soy agar plates containing 5% sheep's blood. The tryptone, soytone and sheep's blood cells provide a source of nutrients (sugar, nitrogen, minerals) to enhance cellular growth. Depending on the bacteria, either

anaerobic or aerobic conditions were employed (Table 1). A temperature of 37°C was maintained for all bacteria. Efforts were made to ensure that the growing conditions and concentration of cells analyzed were kept constant. The *Shigella*, *Escherichia* and *Clostridium perfringens* cells were grown for 24 hours, however the more fastidious *Clostridium difficile* required 48 hours. To circumvent this, twice the number of plates was used to cultivate *C. difficile* and after 24 hours, the cells from two plates were combined and added to form one cell suspension.

After the 24 hour growth period, the cells were removed from the plate using a sterile disposable swab and placed in a test tube containing 1 mL of sterile water until the optical density reading reached 1.0 using a MicroScan Turbidity Meter (Dade Behring West Sacramento, CA). The water used had been autoclaved and purified using the RiOs 5 Water Purification System (Milipore Billerica, MA). A 500 µL aliquot of the cell suspension was added to a 1.5 mL protein LoBind micro-centrifuge tube (Eppendorf, Westbury, NY) and washed three times with 500 µL of sterile water followed by centrifugation (6000 × *g* at room temperature for 5 minutes) to remove residual media.

3.2. Cell Lysis

3.2.1. Gram Negative Bacteria

Bacteria are classified in various ways, one of them as being Gram positive or Gram negative. This difference is based on the amount of peptidoglycan (linear polysaccharide chains cross-linked by tetrapeptides) in the cell structure. Gram negative cells have three layers. First, an outer membrane consisting of protein, lipopolysaccharide and phospholipids, next, a peptidoglycan layer, followed by the

cytoplasmic membrane. The cell envelope of Gram negative bacteria contains 10-20% peptidoglycan compared to 50-80% in Gram positive.[7] The name Gram positive/negative arises from the results of a test known as the Gram stain. This test employs a dye that turns red upon reacting with peptidoglycan. Since Gram negative cells have significantly less peptidoglycan, they produce a faint pink color and are assigned a negative result, while the Gram positive cells produce a deep red color indicating a positive result.

Initial work in this study involved Gram negative cells. After the final wash step described above, the cells were resuspended in 150 μL of the lysis solution (1:1 H_2O: acetonitrile, 0.1% v/v trifluoroacetic acid (TFA)). Unlike detergents, which are sometimes used for the chemical lysis of cells, the lysis solution (1:1 H_2O:acetonitrile, 0.1% v/v TFA) is amenable to both MALDI and ESI analysis. After the lysis solution is added, the samples are vortexed for a few seconds and lysis is complete. The suspected mechanism of lysis involves the solubilization of cell wall lipids which may result in swelling and bursting of the cells as well as denaturation of the cell wall and other proteins. That the lysis solution denatures proteins is advantageous in that it will likely inactivate protein toxins such as the Shiga toxin in addition to inactivating proteases which will degrade other proteins of interest. It is expected that many of the proteins extracted using this procedure are highly soluble cytosolic proteins.[37] Many of the less soluble proteins from the outer membrane and other cellular material precipitate immediately after vortexing the samples. In order to prevent some of the cellular debris from clogging the LC column, the samples are centrifuged at 4100 × *g* for 4 minutes at room temperature prior to analysis, making a pellet of the debris and clarifying the

supernatant. Following centrifugation, 65 µL of supernatant was removed and placed in an autosampler vial containing a microvial insert for analysis.

3.2.2. Gram Positive Bacteria

Gram positive bacteria lack the outer membrane layer found in Gram negative cells but have a significantly larger peptidoglycan layer. Since it is this layer that provides the most structural strength and rigidity to the cell, Gram positive cells are much more resistant to disruption and lysis. Applying the Gram negative lysis method to the Gram positive *Clostridia* resulted in very few proteins being extracted. New methods were attempted, and two protocols were successful with each extracting a distinct (by both mass and retention time) set of proteins. After washing the cells as described above, the first method involved adding 150 µL of 1 mg/mL lysozyme (HEWL) (Sigma-Aldrich St. Louis, MO) in 20 mM NH_4OAc and incubating at 37°C for 30 minutes. Lysozyme disrupts peptidoglycan structure by hydrolyzing β-1,4 linkages of the polysaccharide chains. This was followed by four cycles of freeze/thaw in liquid nitrogen and a 37°C water bath respectively. The second lysis method began with 75 µL of 1 mg/mL lysozyme in 20 mM NH_4OAc and incubating at 37°C for 30 minutes, followed by four cycles of freeze/thaw. Then, 75 µL of 1:1 H_2O: acetonitrile, 5% v/v TFA was added. The combination of these steps was required to recover a similar amount of proteins as seen in the Gram negative protocol. Each individual step of lysis protocols resulted in insufficient lysis, requiring a combination of disruption techniques.

In addition to Gram positive bacteria being more difficult to lyse, some Gram positive bacteria (from the phylum Firmicutes) are known to sporulate posing greater

difficulty in lysis and inactivation. Endospores are resistant to extremes of pH, temperature, certain disinfectants and radiation. For this reason, endospores are regarded as the most resistant of any known biological structure.[39] Indeed, viability studies using the lysis procedures described above revealed that neither lysis method was capable of inactivating the endospores. Given that under idealized conditions (such as laboratory cell culture) endospore forming bacteria may produce only a small amount of spores, it was more straightforward to remove the endospores rather than attempting to lyse them. To this end, after centrifugation at 4100 x g for 4 minutes at room temperature to pellet the debris and clarify the supernatant, the supernatant was filtered using a disposable syringe and blunt tip needle (BD, Franklin Lakes, NJ), and a 0.22 μm, 4 mm polyvinylidene fluoride (PVDF) low protein binding GV filter (Milipore, Billerica, MA). Subsequently, 65 μL of the filtered supernatant was transferred to an autosampler vial for analysis.

3.3. Liquid Chromatography

3.3.1. High Performance Liquid Chromatography

Prior to detection by mass spectrometry, the proteins from the lysate are separated using high performance liquid chromatography (HPLC). The purpose of this prior separation is to simplify the detection process by reducing the number of proteins that enter the mass spectrometer at any given time. Reversed phase chromatography was chosen because of its use of volatile, water-miscible solvents. These solvents allow for easy removal which is advantageous for down stream applications such as fraction collection and detection by mass spectrometry. Gradient elution is most often used to

separate protein mixtures and employs two mobile phases. The amount of organic solvent (% B) is increased until a sufficient amount is present to displace the protein from the site of adsorption causing elution from the column. Since the % B is increased slowly over time and proteins with the larger amount of or more easily accessible hydrophobic regions will stay adsorbed to the column longer, hydrophobicity is generally proportional to retention time.[40] However, other factors such as charge and conformation of the protein can play a small role in the separation as well.

The liquid chromatograph used was an Acquity (Waters, Milford, MA). The gradient conditions are listed in Table 3, where A = 1% formic acid in deionized water and B = 1% formic acid in 2-propanol. The water was purified in house to yield organic-free 18.3 MΩ × cm using an E-pure purification system (Barnstead International Dubuque, IA). The HPLC grade solvents 99% formic acid and 2-propanol were purchased from Fisher Scientific (Fairlawn, NJ) and Honeywell Burdick and Jackson (Morristown, NJ), respectively. The column used was a nonporous polymeric column made from polystyrene-divinylbenzene (PSDVB), Prosphere P-HR 2.1 x 150 mm, 4 μm particle size (Alltech, Columbia, MD) and was operated at a temperature of 50°C. To help preserve the samples, the autosampler was maintained at a temperature of 15°C prior to administering the injection volume of 20 μL. A typical chromatogram of an *E. coli* O157:H7 isolate is shown in Figure 1.

Table 3. LC gradient conditions used in this study. All steps in the gradient were linear.

Time (min)	A%	B%
0.00	95	5
60.00	45	55
65.00	5	95
70.00	5	95
70.01	95	5
75.50	95	5

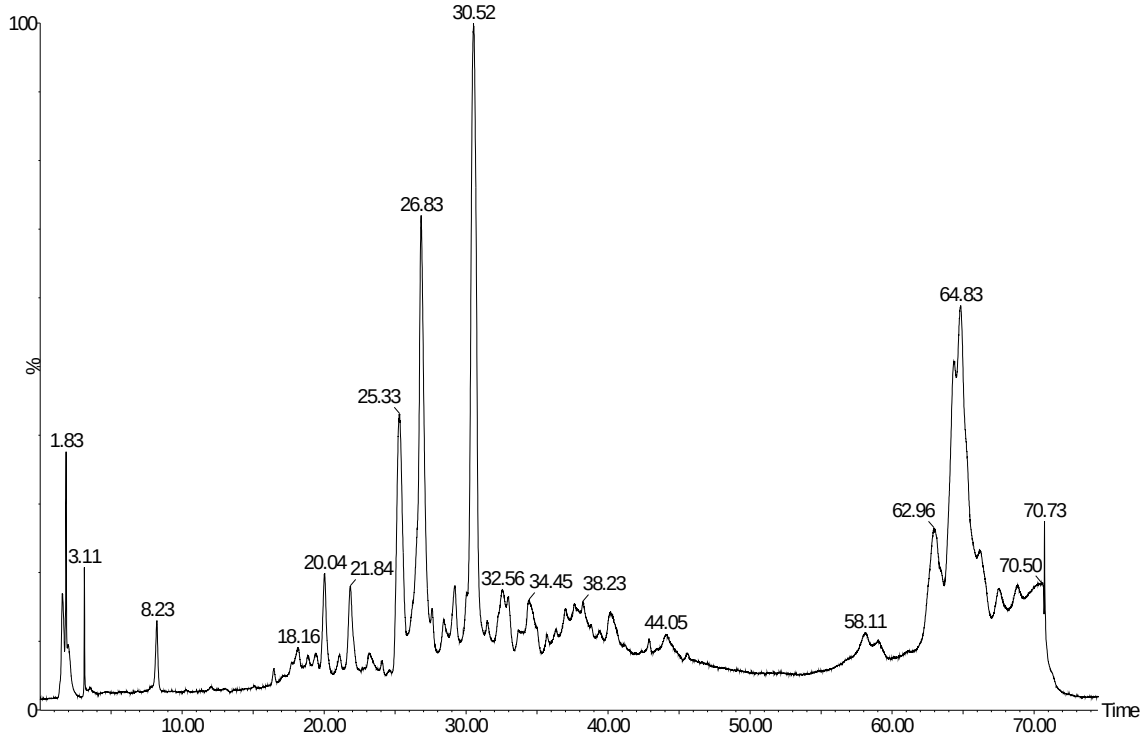


Figure 1. A typical HPLC/MS chromatogram from an *E. coli* cell lysate. Shown is *E. coli* O157:H7, accession # 06-1439. The y-axis represents % relative intensity and the x-axis is time in minutes.

The gradient conditions in Table 3 are a compromise between improved resolution and maintaining good peak shape. The slope of the gradient is one of the simplest ways to control the peak shape of proteins with the larger the % B/min yielding the sharpest peaks. However, it is the difference in % B at the point of elution that

distinguishes the proteins so increasing the rate of change of organic content results in a decrease in resolution. The majority of proteins elute prior to 55% B with only a few eluting during the ramp to 95% B. The hold at 95% B is to help clean the column of lipids and contaminates in the mobile phase such as polyethylene glycol, and is followed by a hold at 5% B to allow the column to equilibrate to initial conditions prior to the next injection.

3.3.2. Ultra Performance Liquid Chromatography

Ultra performance liquid chromatography (UPLC) involves the separation of compounds using sub-2 μm particle sizes at pressures that exceed the capacity of standard high performance liquid chromatographs. Van Deemter curves (plate height vs. linear velocity) for applications utilizing sub-2 μm particles predict lower plate heights (H) over a wider range of linear velocity than with larger particle sizes.[8] The lower H values occur because H is proportional to the particle diameter (d_p). The optimum linear velocity is higher and occurs over a wider range due to the reduced resistance to mass transfer (C term in the Van Deemter equation); a consequence of using particles with a smaller diameter.[9] Lower plate heights result in better resolution (see Equations 1 and 2) and the wide range of linear velocities at this plate height means the flow rate can be increased while maintaining equal chromatographic resolution (Rs), yielding shorter runtimes. Increased column back pressure has also been shown to improve protein recovery from the column.[41] To better ascertain the benefit received from UPLC vs. HPLC, nearly identical columns were utilized. The HPLC column was the narrowbore X-bridge C18 BEH 300 Å 2.1 × 150 mm, 3.5 μm (Waters, Milford, MA). According to

the manufacturer, the UPLC column was identical (e.g., ligand density, pore size, carbon load) to the HPLC column except in particle size (1.7 μm).

Chromatographic resolution (Rs) is described by the following equation:

$$\mathrm{Rs} = \frac{\sqrt{N}}{4}\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k}{k+1}\right) \qquad (1)$$

Where, N = separation efficiency (plate number), α = selectivity, and k = retention factor (capacity factor).

The separation efficiency (N) equals:

$$N = \frac{L}{H} = \frac{L}{hd_p} \qquad (2)$$

Where, L = column length, H = plate height, h = reduced plate height, and d_p = particle diameter.

Therefore, since N is proportional to $\sqrt{N}$, Rs is inversely proportional to d_p:

$$\mathrm{Rs} \propto N \propto \frac{1}{d_p}$$

While chromatographic separation helps to simplify the sample prior to detection and yields information on individual protein reversed phase retention behavior, chromatography alone is rarely enough to distinguish two closely related strains of bacteria. Occasionally a unique peak can be observed in the chromatogram, but often differences in the proteins are very minute such as a PTM or a single amino acid substitution, either of which will likely not have a significant impact on the retention of the protein. Furthermore, even with optimized conditions, in mixtures as complex as

bacterial cell lysates, co-elution of proteins will occur. A sensitive, and specific detection system such as mass spectrometry is needed that can further resolve co-eluted proteins.

3.4. Mass Spectrometry

The mass spectrometer for this project was a Q-TOF Premier (Waters, Milford, MA). The Q-TOF has two quadrupoles (Q) prior to the orthogonal acceleration time of flight (TOF). Table 4 lists the source conditions for the mass spectrometer. These and the other parameters used in the MS method were optimized for best desolvation at 0.225 mL/min; and optimal sensitivity and resolution for bovine serum albumin. Positive ion ESI was used to create ions that were monitored over a mass range of 620-2450 Da and resolved in single reflectron (V) mode.

Table 4. Source conditions of the mass spectrometer.

Capillary	+ 3.9 kV
Cone voltage	40 V
Source Temperature	115^{o}C
Desolvation Temperature	500^{o}C
Desolvation Gas Flow	900 L/hr

ESI is an atmospheric pressure technique where liquid is passed through a charged capillary causing the liquid to form an aerosol spray of fine charged droplets. As the droplet sizes are reduced, the analyte molecules inside the droplets are removed and thereby enter the gas phase.[42] In full-scan mode (the mode used here), the two quads are in radio frequency (RF) only mode and act as ion guides rather than mass analyzers. Once the ions enter the quads, they have a trajectory that is hyperbolic in motion that is governed by the Mathieu equation. Mass analysis in a TOF is based on the mass

dependence of ion velocity[43] and, that for a given velocity over a fixed distance, the time taken to travel that distance is easily obtained. The set of equations[44] that govern this process is as follows:

The kinetic energy of an ion accelerating in an electric field (E) is given by:

$$\frac{mv^2}{2} = zeE \qquad (3)$$

Where, m = mass, v = finally velocity achieved during acceleration, z = number of charges, e = charge of an electron.

It then follows that:

$$v = \sqrt{\frac{2zeE}{m}} \qquad (4)$$

If d, is the distance that an ion travels in the flight tube, then the time, t, taken to travel this distance is given by:

$$t = d/v = \frac{d}{\sqrt{\frac{2zeE}{m}}} = d\sqrt{\frac{(m/z)}{2eE}} \qquad (5)$$

Since d is fixed, e is a universal constant, and E is held constant in the mass spectrometer, t is then directly proportional to the square root of (m/z) times a constant, C.

$$t = C\sqrt{(m/z)} \qquad (6)$$

Since t is proportional to (m/z), ions of a larger (m/z) take longer to travel to the detector and vise versa.

When comparing results from different days, it is crucial that the instrument is providing optimal and reproducible results. For this reason, prior to each analysis, the instrument is calibrated and the mass resolution at full width half maximum (R_{FWHM}) is recorded as a quality control measure. The calibrant used was NaI because it produces a

wide range of ions allowing the instrument to be calibrated from 300 – 2500 Da. The daily specifications required of the instrument are: a mass error of no more than approximately 5 ppm and a $R_{FWHM} \approx 10{,}000$. The mass accuracy is calculated as follows:

$$\left| \frac{M_A - M_I}{M_A} \right| \times 10^6 = \text{mass error (ppm)} \qquad (7)$$

Where M_A and M_I are the actual mass and indicated mass respectively, in units of Da. The resolution is calculated as shown in equation 8:

$$\frac{M_I}{\Delta M_I} = R_{FWHM} \qquad (8)$$

Where ΔM_I = the full peak width at half maximum height (in Da) of the ion of interest. The $Na_8I_7^+$ cluster ion at 1072.249483 Da was utilized for both calculations.

3.5. Data Analysis

3.5.1. Manual Interpretation

A critical aspect of this project was the data analysis and there are two key issues pertaining to it: data complexity and the amount of data collected. Both hamper the data analysis step by decreasing the throughput and increasing the time required to report results to medical staff/investigating authorities. The complexity of the data largely arises from the fact that as proteins undergo the electrospray process, multiply charged ions are produced (Figure 2). While this aspect has been a major advantage of ESI by allowing large biomolecules to be observed on inexpensive mass analyzers of moderate mass range (< 3000 Da) such as quadrupoles, the envelope of multiply charged ions can

be challenging to interpret even for a pure compound. Furthermore, a mixture of proteins of different mass may have overlapping charge states in their envelopes and which charge states belong to which protein may be unclear, making the spectral interpretation of a protein mixture significantly more difficult. Even with the use of chromatography, this scenario occurs often, for with complex mixtures like cell lysates, co-elution of proteins will be common.

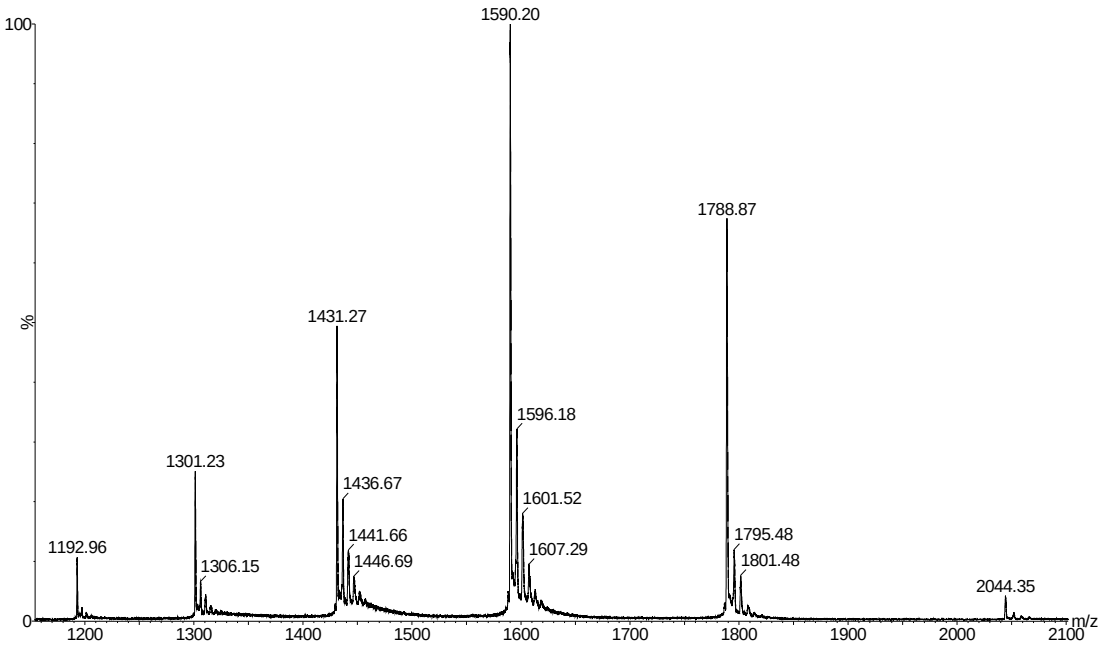


Figure 2. A mass spectrum depicting the envelope of multiply charged ions produced by electrospray ionization of a protein. The y-axis represents percent relative intensity and the x-axis indicates mass/charge ratio (*m/z*).

The most common method of addressing the complexity of interpreting ESI mass spectra of proteins is to employ computer-based algorithms. To evaluate their efficacy for charge state deconvolution, one can initially test the method on standards of known

mass. However for quality assurance purposes, some manual interpretation skills are beneficial when proteins of unknown mass are being analyzed, as in a cell lysate analysis. Some guidelines that can assist with the manual interpretation of protein mass spectra acquired using ESI are given below. The mass/charge ratio (m/z) is obtained from mass spectral analysis and the mass spectrum is a plot of intensity vs. m/z (Figure 2). However, there are three unknowns in the mass spectrum: the number of charges (z), the neutral (zero charge) mass (M), and the mass of the adduct ion (m_a). The charges received by a protein in positive ion mode are most often solvent adducts bonded by ion/dipole force. The mobile phase used in this study is acidic (1% v/v formic acid, pH ≈ 2.5). For this reason, the most common adduct is a proton (H^+) weighing 1 Da. The number of charges attributed to each peak (n) is discrete (where n is used for the number of charges to avoid confusion with z), having only integer values with each peak in the envelope being one charge away from its neighboring peak. The charge state can be determined by using two adjacent peaks and solving a simultaneous equation beginning with:

$$m/z = \frac{(M + nm_a)}{n} \qquad (9)$$

Given the mass spectrum in Figure 2, and letting m_a be equal to H^+ (1 Da), the two adjacent *m/z* peaks at 1590.20 and 1431.27 (the latter having one more charge than the former) can be used to determine the charge state as follows:

$$1590.20 = \frac{(M + nH^+)}{n} \qquad 1431.27 = \frac{(M + (n+1)H^+)}{(n+1)}$$

Letting $H^+ = 1$, solve for M:

$$1590.20n = M + n \qquad 1431.27(n + 1) = M + (n + 1)$$

$$1589.20n = M \qquad 1430.27(n + 1) = M$$

Since $M = M$,

$$1589.20n = 1430.27(n + 1)$$

Solve for n:

$$1589.20n = 1430.27n + 1430.27$$

$$158.93n = 1430.27$$

$$n = 9$$

Recall that:

$$1590.20 = \frac{(M + nH^+)}{n}$$

Since $H^+ = 1$, and $n = 9$, it then follows that:

$$1590.20 = \frac{(M + 9)}{9}$$

Now solve for M:

$$14311.80 = M + 9$$

$$M = 14302.80$$

Since there is a certain amount of error with each peak measurement, after the charge state for two of the peaks has been determined, the charge states can then be assigned to all peaks in the envelope. Once each peak has been assigned a charge state, each one can be used to determine *M* and then determine the mean *M* value. The unweighted average of the *M* values in Table 5 is 14302.89 Da.

Table 5. The M values determined from each peak in the envelope of multiply charged ions in Figure 2.

m/z	***n***	***M***
1192.96	12	14303.52
1301.23	11	14302.53
1431.27	10	14302.70
1590.20	9	14302.80
1788.87	8	14302.96
2044.35	7	14303.45

A simpler way to determine the charge state for a particular peak is to examine the mass difference between the ^{12}C and ^{13}C isotopes. A singly charged species will have an isotopic difference of 1 Da, 0.5 for doubly charged, 0.33 for triply charged, and so on. As mentioned previously, once a charge state is determined, the neighboring peaks will differ by one charge, allowing all peaks to be assigned a charge state. For a protonated species, when the charge state (n) of a peak is known, the neutral mass (M) can be determined by:

$$M = n[(m/z) - 1] \qquad (10)$$

The problem with this method is that determining the isotopic difference is dependant on the resolution of the mass spectrometer. Therefore, the utility of this

method will decrease with decreasing MS resolution: ion cyclotron resonance > Orbitrap > TOF > quadrupole. For the instrument used in this study, having a $R_{FWHM} \approx 10{,}000$ at *m/z* 1000, the highest charge state that can be accurately be determined at m/z 1000 on a centroided spectrum is +10 ($\Delta = 0.1$). Figure 3 depicts a centroided version of the +10 charge state of the protein shown in Figure 2. This means that the above method of charge state determination is only applicable to proteins weighing approximately 10,000 Da or less.

Figure 3. Close inspection of the 1431.27 ion in Figure 2 allows charge state determination. The spectrum is first smoothed (bottom) then centered (top). The isotopic separation of 0.1 Da indicates a +10 charge state for this ion. The M value from this peak corresponds to 14302.70 Da.

The arduous task of manually determining the neutral masses of hundreds of proteins per isolate during cell lysate analysis and the limited ability of most commercial instruments to allow visual charge state determination make clear the need for using computer based approaches.

3.5.2. Chemometric Methods

Not long after the first protein was analyzed by ESI-MS, computer based chemometric methods were used to deconvolute multiply charged protein ions into their neutral masses (Figure 4). The first method was developed by Mann et al.[45] This algorithm satisfied the two main goals of identifying the charge state of each ion in the envelope and determining the neutral mass of the parent ion, but had the disadvantages that in the deconvoluted spectrum, there were artifact peaks other than the protein of interest and the baseline increased with mass range. Other algorithms were soon developed[46, 47] notably one developed by Hagen and Monnig which corrected the rising of the baseline with mass and filtered out more noise giving cleaner deconvoluted spectra, but all quantitative information was lost.[48] Today, the most commonly used method is known as Maximum Entropy (MaxEnt 1) developed by Ferrige et al.[49, 50] This method produces few artifacts, is quantitative and can effectively resolve mixtures of proteins (Figure 4). While this algorithm has been in use for years, it can only process a single or few chromatographic peaks at a time. For this reason a recently developed software known as Protrawler6 (Bioanalyte, Portland, ME) can be employed. This software automatically deconvolutes entire chromatograms by dividing the data into time intervals, performing sequential deconvolutions, and then summing the centered data

together into one text file containing neutral masses, intensities and retention times.[6] As seen in Figure 4, the calculated mass for the protein in Figure 2 is 14303.83 Da, which is very similar to the unweighted average that was obtained by manual calculation, 14302.89 Da. The results from MaxEnt1 are slightly high (+ 0.94 Da), but reasonable (mass difference < .01%). Protrawler6 uses an algorithm that is similar to MaxEnt 1 for deconvolution, but requires less time for processing and adjusts for the slightly higher mass estimates. For this reason, the values obtained from this algorithm are typically lower, and after processing with Protrawler6, the neutral mass determined for the protein in Figure 2 was 14303.20 Da.

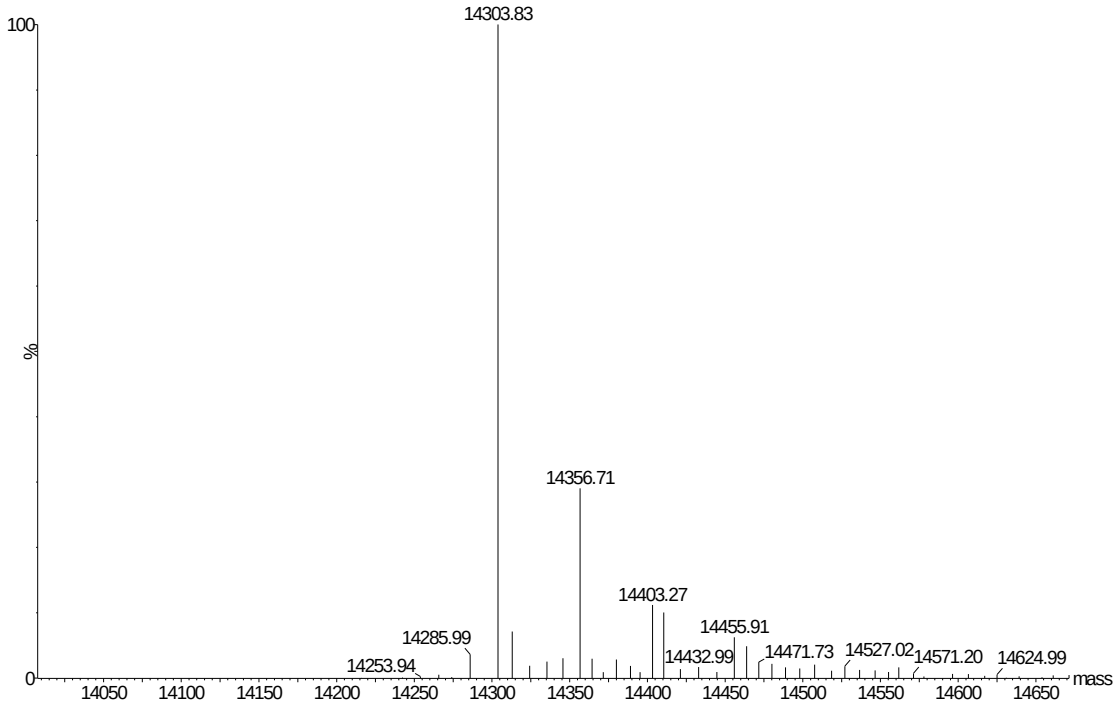


Figure 4. After processing with MaxEnt 1, the multiply charged spectrum in Figure 2 is deconvoluted to a zero charge (neutral) mass spectrum.

To apply Protrawler6 to LC/MS data, one must first build a model that accounts for the resolution of the mass spectrometer, whether the deconvolution will be used for peptides or proteins, and the typical background noise of the instrument. For this reason, an analysis on peptides (which would require deisotoping and charge state deconvolution) using an ion trap instrument (lower resolving power) would require a different model than an analysis of intact proteins (requiring only charge state deconvolution) using a TOF instrument (higher resolving power). The model is then used to create a mock protein spectrum of known mass. This mock spectrum is used to fit the raw data collected from the instrument, and is adjusted iteratively until the fit reaches convergence. Since the mass of the mock spectrum is known (because it was created by the software) once the mock spectrum satisfactorily fits the raw data, the mass of the protein will be determined.

When an instrument of the same or highly similar resolution and background noise is used and the model is applied to the same type of analysis, i.e., intact proteins rather than peptides, the same model can be successfully applied to any type of protein analysis on that instrument. The model is not dependant on the source of the proteins (e.g. standards, *Shigella* lysates, *Clostridium* lysates). For this project, one model was built and applied to all of the analyses thereafter. Since the proteins to be analyzed will vary in size, models that are built for a large protein are more successful because they more readily encompass small proteins than a small protein model can large proteins. The model used here was applicable from 5 – 75 kDa. This mass range covered proteins typically observed with this method and if after routine manual inspection of the data,

proteins outside this range were observed, the model could be adjusted accordingly, but that was not necessary here.

To verify the validity of the model, one can apply it to protein standards of known mass or calculate the mass manually or use MaxEnt1 as described above. An ideal model produces accurate charge state deconvolution and neutral mass determination, low noise and few artifact peaks. Although parameters during model development were adjusted to accurately account for the background noise and to minimize artifacts, the production of some artifacts was unavoidable. All data compiled from Protrawler6 that was < 1% relative intensity compared to the base peak was removed. This helps prevent artifacts from subtracting out real peaks during spectral subtraction.

Once a model has been developed and its utility verified, it can then be applied. Since the length of the chromatogram and width of the typical chromatographic peak may change depending on which chromatographic method was used, the width of the time slice and the trawl start and stop times are entered individually for each chromatogram. The typical width used was 30 seconds. Because the first few minutes of the chromatogram were from the solvent peak or system dead volume, and the last ~10 minutes were for column cleaning and re-equilibration, no proteins eluted during this time. Therefore, to minimize processing times, Protrawler6 was typically set to process from 2.5 – 62.5 minutes.

It is important to note the distinction between summing the MaxEnt data iteratively as opposed to summing the entire chromatogram and doing a single MaxEnt analysis. By summing time slices of MaxEnt data, the noise level is reduced because

only formerly multiply charged ions are present in the MaxEnt results. This is in contrast to summing the entire chromatogram which would be a summation of multiply and singly charged data and given that the majority of background noise in LC/MS is singly charged, this summation would drown out nearly all low abundance multiply charged ions. The retention time information is tabulated in order to distinguish two proteins of the same mass that have different retention times and if a protein is found to be interesting, its retention time could then be utilized to isolate the protein for further study by collecting fractions as they elute off of the column.

In addition to lessening the complexity of the data, the application of Protrawler6 also results in a significant reduction of the overall amount of data. The mass spectrometer is set to acquire spectra at a rate of 2 spectra/second and the chromatographic run time is 75.5 min (4,530 seconds) so each data file contains 9,060 summed spectra. But, after processing with Protrawler6, the data is reduced to one text file which is then converted into a single mass spectrum representing all of the proteins observed in the isolate.

3.5.3. Processing of Deconvoluted Spectra

Mass spectra are created from text files produced after Protrawler6 analysis using MS Manager software (Advanced Chemistry Development Laboratories, Ontario, CA). The data analysis process is then further simplified by using the spectral mirroring and spectral subtraction tools in MS Manager. Spectral mirroring allows the comparison of two spectra with one of them being inverted (Figure 4). Viewing the spectra in this manner simplifies the search for unique masses by placing the baseline (where

differences occur) at the center of view rather than overlaying them which places the baseline at the bottom of the screen. The second tool, spectral subtraction, is useful for comparisons of two spectra that have a high degree of similarity; such as two different serotypes of the same species, referred to as spectrum A and B in Figure 5. Spectral subtraction involves removing the peaks common in both spectra from one of the spectra within a user-defined mass accuracy window (± 2 Da is used here), leaving a new spectrum containing only unique masses (spectrum A2). This process is then reversed to produce a spectrum containing masses only unique to spectrum B2. These two subtracted spectra can then be mirrored, greatly simplifying the identification of unique masses (Figure 6).

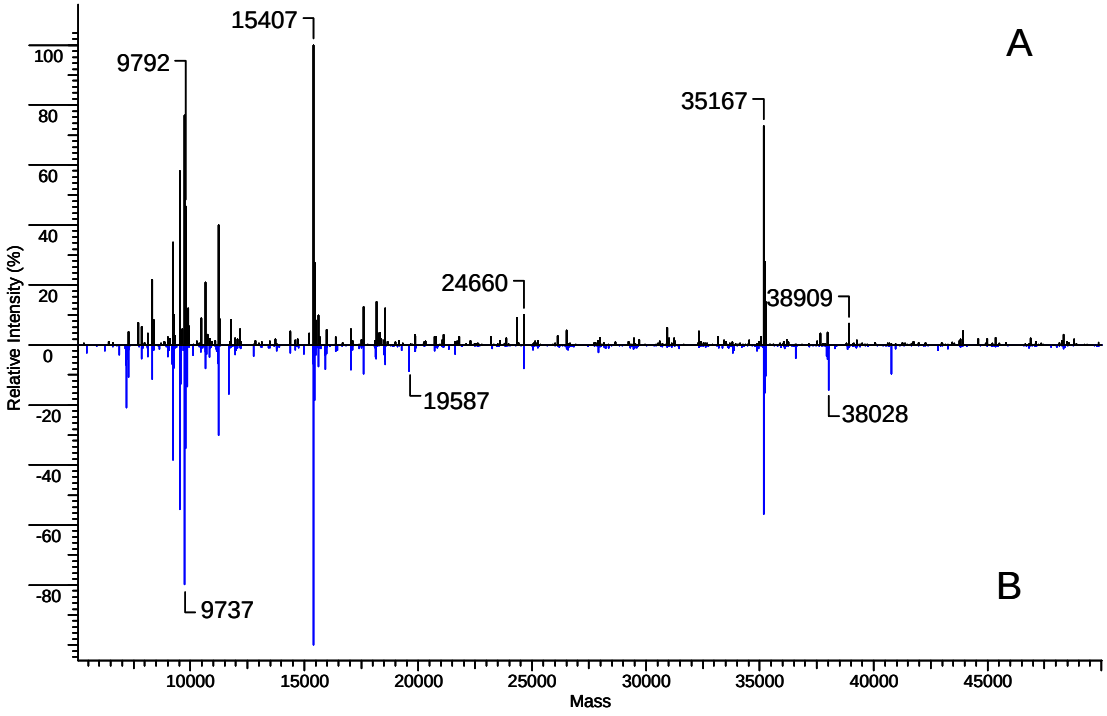


Figure 5. Mirrored spectrum of *E. coli* O157:H7, accession # 06-1439 (top - A) and non-pathogenic *E. coli*, accession # 06-0004 (bottom - B) showing many common peaks.

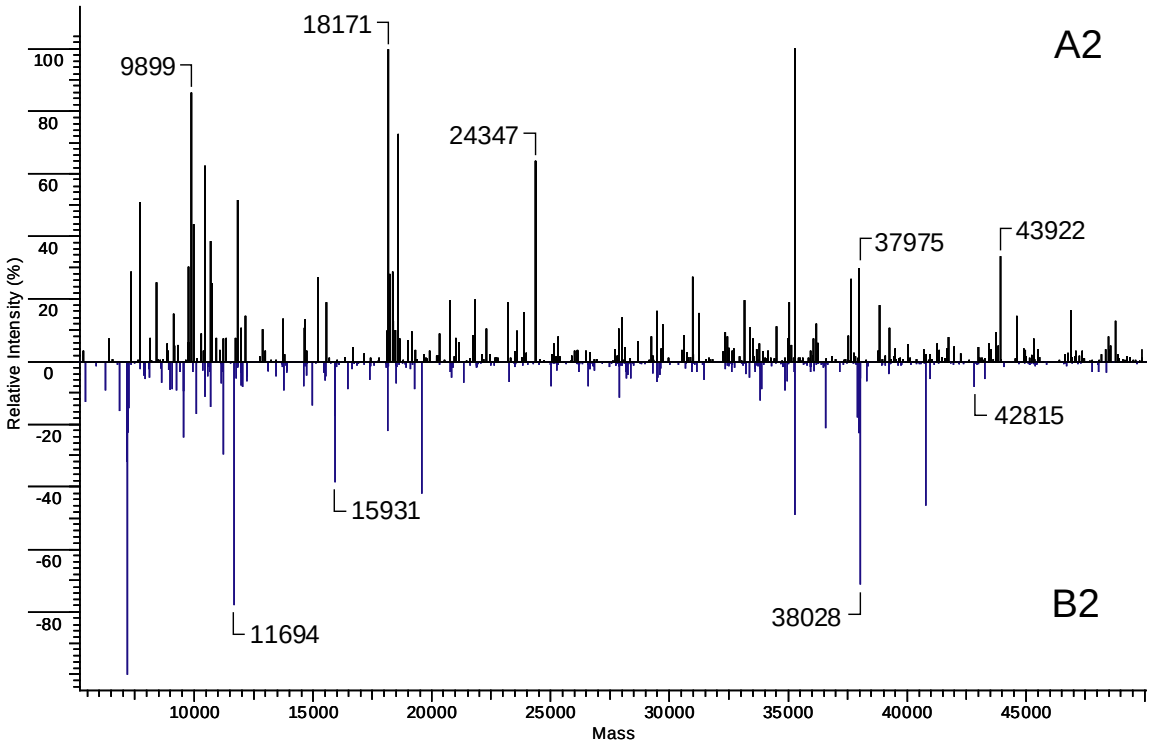


Figure 6. Subtracted spectra of *E. coli* seen in Figure 5. Following subtraction, unique peaks become much more obvious. A subtraction window of ± 2 Da was used.

One difficulty with the MS Manager program is that mirroring and subtraction can only be performed with two spectra at a time. This is a problem if one needs to compare groups of spectra such as comparing two of the O157:H7 spectra against the eight other *E. coli* and *Shigella* spp. studied. One possible solution would be to compare each *E. coli* O157:H7 spectra to each non *E. coli* O157:H7 spectra and tabulate the results for each individual comparison; however, with 5 replicates of 10 isolates this would be very time-consuming. To address this problem, theoretical spectra are made, i.e., the text files for all non O157:H7 isolates are combined, and then this text file is converted into a single mass spectrum. Since the hypothetical text file contains a large amount of masses, some of which are actually artifacts from the MaxEnt 1 process, the hypothetical spectra

are filtered by removing all peaks that are < 1% relative intensity of the most abundant (base) peak in the mass spectrum. "Hypothetical" is used because analysis of a mixture of all non O157:H7 isolates at roughly the same concentration as the individual isolates was never performed. However, since all of the masses of each individual isolate are included into the hypothetical spectrum's text file, this new spectrum is valid for comparison purposes. Furthermore, if a mixture as complex as the one described above were to be analyzed, it is likely that many of the lower abundance proteins would be missed and much poorer chromatographic resolution would occur, thereby yielding poorer quality data than that present in the hypothetical spectrum.

CHAPTER 4 Characterization of *Escherichia coli* and *Shigella* Species Utilizing Liquid Chromatography/Mass Spectrometry of Intact Proteins

4.1. Introduction

4.1.1. Natural Routes Exposure

Foodborne illnesses are a significant cause of social and financial burdens even in industrialized nations. In 1997, costs from foodborne illnesses in the United States alone were estimated at $35 billion and were attributed to medical expenses and lost productivity.[2] Two of the top ten leading causes of food and waterborne illness outbreaks reported to the Centers for Disease Control and Prevention from 1972 - 2000 were Shiga toxin producing *E. coli* (STEC) and *Shigella* organisms.[51] A particularly dangerous subset of STEC are the enterohaemorrhagic *E. coli* (EHEC), which include serotypes O157:H7, O26:H11 and O111:NM.[52, 53] Annually, *E. coli* O157:H7 alone is responsible for 73,000 cases of infection, 2,100 hospitalizations and 61 deaths in the U.S..[54] Notable was the recent outbreak of *E. coli* O157:H7 caused by the consumption of contaminated spinach which infected 199 persons resulting in 102 hospitalizations and 3 deaths.[55]

Moreover, *Shigella* organisms cause an estimated 400,000 cases of shigellosis each year in the U.S. More severe is the impact in developing countries where 163 million *Shigella* infections occur annually resulting in over 1 million deaths. Of the four species of *Shigella*, *S. sonnei* and *S. flexneri* are the two most commonly implicated in human illness.[56, 57] Children, the elderly and the immunocompromised are most

susceptible to the severe sequelae of *Shigella* and EHEC infections such as hemolytic uremic syndrome (HUS), the principle cause of renal failure for U.S. children.[54, 56]

4.1.2. Potential for Deliberate Use

Since natural outbreaks of these organisms have had considerable public health impact, of even greater concern is the threat of these pathogens being used intentionally as biowarfare agents. Cultures of these organisms could easily be obtained from the stools of animals or humans and only a small inoculum (as few as 10 organisms) is required for infection. Additionally, both *E. coli* and *Shigella* infections can be transmitted person to person allowing the impact to extend beyond those initially exposed.[54, 56] STEC and some species of *Shigella* produce a toxin that is classified as a Category B select agent[58] and deliberate use of these bacteria has been documented. *Shigella* organisms were used in battle by the Japanese in WWII[51] and by a disgruntled hospital employee in Texas to infect co-workers.[59] Furthermore, *E. coli* cultures were found in possession of a Wisconsin man later arrested for possession of a toxin for use as a weapon.[60]

4.1.3. Review of Current Methods of Analysis

In the event of an infectious disease outbreak or terrorist attack, key decisions by hospital staff, public health officials and investigating authorities will have to be made to guide medical treatment, prophylaxis and remediation. These decisions are best made when they are data directed. Clinical manifestations (malaise, abdominal pain, diarrhea etc.) of exposure do not unambiguously identify their cause. Therefore, rapid, sensitive and specific analytical methods are needed to gain further insight. Rapid diagnoses of

infections, such as those leading to HUS, are crucial as early treatment with intravenous volume expansion has been shown to decrease kidney damage and augment patient outcome.[61] Sensitive and specific methods are desirable to minimize false negatives and positives respectively. Three commonly used analytical methods in diagnostic microbiology laboratories are: enzyme immunoassay (EIA), real-time polymerase chain reaction (PCR), and pulsed-field gel electrophoresis (PFGE).

EIA methods are sensitive often eliminating the need for cultural enrichment, but lack the specificity to be considered confirmatory. This was evident when public health response was misguided during an outbreak of gastroenteritis in which two independent laboratories found stool samples positive for Shiga toxin (Stx) by EIA, but further investigation revealed norovirus as the cause.[62] Real-time PCR is a rapid and sensitive approach requiring 0.5 - 4 hours post culture to perform. However, strain specific and often species specific primers are unavailable or impractical and for this reason the specificity required in outbreak investigations is not typically afforded by this method. Secondly, the mere presence of a gene does not guarantee that protein is being expressed. This is an important consideration as bacterial pathogens have been shown to contain genes that are not expressed.[25-27] Finally, the gold standard for subtyping of bacteria during outbreak investigations is PFGE. While this technique is capable of providing strain level discrimination, it is not easily automated, is labor intensive and requires a minimum of 2 days post culture.

4.1.4. Specific Aims

Several different reviews on the techniques employed to analyze bacteria all had one common conclusion – the strongest approach is polyphasic, combining information from several different yet complimentary techniques.[13, 16, 28, 63] The approach described here is unique in that it utilizes liquid chromatography/mass spectrometry (LC/MS) of intact proteins, to monitor protein expression in bacterial cells. Key differences between closely related strains may occur within the proteome to which genetic approaches are insensitive (e.g., posttranslational modifications, PTMs). One of the dominant PTMs in bacteria, removal of the *N*-terminal methionine, is estimated to occur in 50% of bacterial proteins.[33] In total, the differences that can be observed during the comparison of proteins can be qualitative: the presence/absence of proteins, protein sequence mutations and PTMs, or quantitative, i.e., varying levels of protein expression.[34]

To examine the efficacy of this approach as a tool in diagnostic microbiology, a brief comparison with MALDI/TOF-MS was performed, then a model set of ten of *Shigella* and *E. coli* clinical isolates were studied (Table 6). From these ten isolates, biomarkers based on protein mass, retention time and relative intensity were discovered and evaluated for their reproducibility by performing five replicate analyses. Finally, the validity of these markers was challenged by applying them to a blind test of clinical isolates.

Table 6. The ten known isolates examined in this study. ND = not determined, NM = non-motile and NA = not applicable.

Family	**Enterobacteriaceae**					
Genus	***Escherichia***				***Shigella***	
Species	*E. coli*				*S. flexneri*	*S. sonnei*
Serotype	ND Stx (-)	O111:NM	O26:H11	O157:H7	ND	NA
Accession Number	06-0004	06-1440	06-1418	06-1439	04-0497	06-1362
	06-0006			06-1464	06-0967	06-1364

4.2. Experimental

4.2.1. Materials

HPLC grade solvents (acetonitrile, formic acid and trifluoroacetic acid) were purchased from Fisher Scientific (Fairlawn, NJ) and 2-propanol was purchased from Honeywell Burdick and Jackson (Morristown, NJ). The water utilized for HPLC analysis was purified in house to yield organic-free 18.3 MΩ × cm using an E-pure purification system (Barnstead International Dubuque, IA). Sterile water that had been autoclaved and purified with a RiOs 5 Water Purification System (Milipore Billerica, MA) was used during bacteria preparation.

4.2.2. Growth and Lysis

Bacterial isolates were obtained from the Virginia Division of Consolidated Laboratory Services. Cells were grown for 24 hours at a temperature of 37^{o}C. Trypticase soy agar plates containing 5% sheep's blood in the presence of oxygen with 5% CO_2 were used as the growth medium. After this growth period, cells were removed from the plate and placed in a test tube containing 1 mL of water until the optical density

reading reached 1.0 using a MicroScan Turbidity Meter (Dade Behring West Sacramento, CA). A 500 µL aliquot of this suspension was washed three times with 500 µL of water followed by centrifugation (6000 × *g* at room temperature for 5 minutes) to remove residual media. Finally, the cells were resuspended in 150 µL of the lysis solution (1:1 H_2O: acetonitrile, 0.1% v/v trifluoroacetic acid). After chemical lysis, the sample was again centrifuged (4100 × *g* for 4 minutes) at room temperature. Following centrifugation, 65 µL of supernatant was removed and placed in an autosampler vial for analysis.

4.2.3. LC/QTOF MS Analysis

Intact proteins were separated by reversed phase chromatography using an Acquity UPLC (Waters, Milford, MA). Gradient elution (5-55% B in 60 min) was used at a flow rate of 0.225 mL/min where A = H_2O (1% formic acid) and B = 2-propanol (1% formic acid). The column was a non porous Prosphere P-HR 2.1 x 150 mm, 4 µm particle size (Alltech, Columbia, MD) operated at 50°C. The autosampler was maintained at 15°C prior to administering the injection volume of 20 µL.

A Q-TOF Premier (Waters, Milford, MA) utilizing positive ion electrospray ionization was used for mass analysis. Ions were monitored over a mass range of 620 - 2450 Daltons (Da) and resolved in single reflectron (V) mode. The parameters employed in the MS method were optimized for sensitivity and resolution using bovine serum albumin.

4.2.4. Data Processing

The LC/MS data was processed using two software packages: Protrawler6 and MS Manager. Protrawler6 (Bioanalyte, Portland, ME) software provided automated deconvolution of multiply charged ions by first dividing the full-scan data from the chromatogram into time intervals (30 seconds) and summing the data from each interval. Sequential deconvolutions were then performed to obtain neutral masses of the proteins that eluted during each interval. A text file containing the neutral masses, intensities and retention times was then created summarizing the results for each chromatogram.[6] Retention time information can be used for further study (e.g., fraction collection) of proteins of interest or to distinguish proteins of the same mass that differ in retention. The masses and intensities were used to create a single spectrum representing all of the proteins observed in the lysate using MS Manager (Advanced Chemistry Development Laboratories, Ontario, CA).

To further facilitate biomarker discovery, MS Manager was employed for spectral mirroring and spectral subtraction. Spectral mirroring allowed spectra to be mirrored along the abscissa, placing the baseline at the center of view. Spectral subtraction removed all common peaks between two spectra within a given mass accuracy so that only unique ones remained. For group and strain level comparisons involving multiple spectra, the text files of all isolates not in that group or strain were combined to create a hypothetical spectrum which could then be used for subtraction. A subtraction window of $\pm$ 2 Da was utilized.

4.3. Results and Discussion

4.3.1. LC/MS vs. MALDI-TOF/MS Comparison

Since nearly all of the early work concerning the MS analysis of bacterial proteins has centered on MALDI-TOF[20], a brief comparison with the LC/MS technique described here was made. MALDI-TOF exhibits certain advantages over LC/MS. One, by producing primarily singly charged ions, data interpretation is greatly simplified relative to ESI-MS. MALDI-TOF is also better suited for the analysis of complex mixtures therefore prior separation (e.g., chromatography) is not required. Consequently, MALDI-TOF has a considerable throughput advantage ~two min (after deposition and drying) compared to ~two hours (data acquisition and deconvolution) over LC/MS.

However, as can be seen in Figure 7, after automated charge state deconvolution with Protrawler6, spectra from LC/MS are as simple to interpret as MALDI data, and are much richer. In addition to providing more proteins (particularly >15 kDa), other advantages to using LC/MS exist. These include improved mass resolution and mass accuracy, reproducibility, and more reliable quantitative data. The last two advantages stem from the uneven distribution of the sample across the spot on the MALDI target and variance in the placement and number of laser shots acquired on the sample.

Having retention time information allows more to be known about the biomarker candidates. MALDI-TOF data is analogous to that obtained from a 1D gel, while LC/MS data is comparable to that acquired from a 2D gel (with obvious improvements in mass resolution and mass accuracy over gel-based approaches). The LC/MS approach also allows for distinctions of proteins of the same mass that differ in retention time. If

the effluent from the LC is split, simultaneous fraction collection and MS analysis can be performed and the collected fractions can be used for further study (e.g., sequencing). Protein isolation for further study can not be performed by MALDI-TOF; either LC or tandem mass spectrometry would be required.

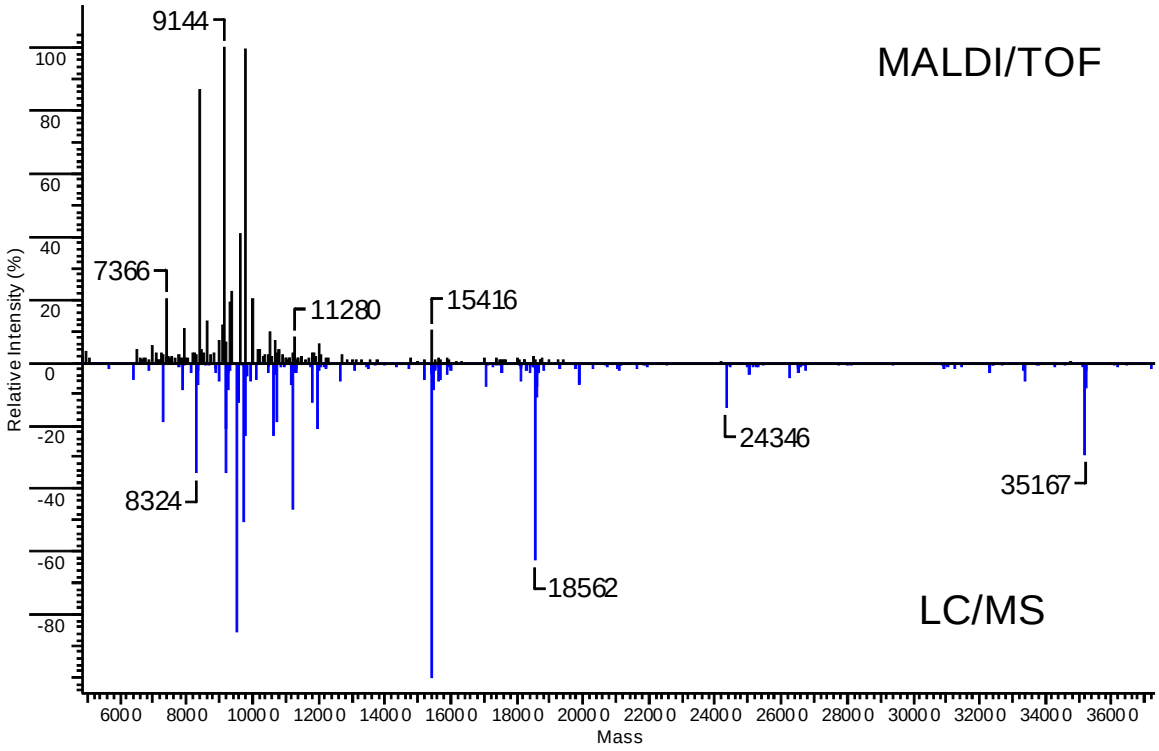


Figure 7. Comparison of MALDI/TOF-MS and LC/MS using the same *S. sonnei* lysate. More proteins, particularly >15 kDa were observed using LC/MS.

Figure 7 shows a comparison of MALDI/TOF and LC/QTOF data using the same sample preparation and protein extraction procedures for a *S. sonnei* isolate. Reasons for the differences in observed proteins may include difficulty in optimizing MS conditions over such a wide *m/z* range (4,000-20,000 Da, Δ = 16,000 Da) with MALDI-TOF

compared to (620-2,450 Da, Δ = 1,830 Da) during ESI, and the complexity of the lysate which may lead to ion suppression and/or detector saturation. Although the LC step causes decreased throughput, this step is likely part of the reason more proteins are observed. Often, distinctions between closely related strains may involve only one or a few proteins and for this reason the increased information content and protein yield observed by the LC/MS approach is likely advantageous and was deemed worthy of further investigation.

4.3.2. Biomarker Discovery

The process of biomarker discovery is shown in Figure 8. First, chromatographic data is collected in full-scan mode (8a). Next, automated charge state deconvolution is performed to yield a single mass spectrum representing all of the proteins observed in the chromatogram (8b). The spectra are then mirrored (8c) and subtracted revealing unique masses (8d). As seen in Figure 8d, numerous peaks appear to be unique to each isolate after subtraction. However, many of these peaks were not reproducible and may have been artifacts from the deconvolution process. For this reason, a protein was deemed a biomarker only if its unique mass, retention time and or relative intensity was observed in each of the five repeated experiments. Spectra from the ten isolates listed in Table 6 were examined to find reproducible biomarkers whose presence or absence could be used to identify unknown samples. To determine the specificity of the technique, a search for biomarkers was made at each taxonomic level (e.g., genus, species, strain etc.).

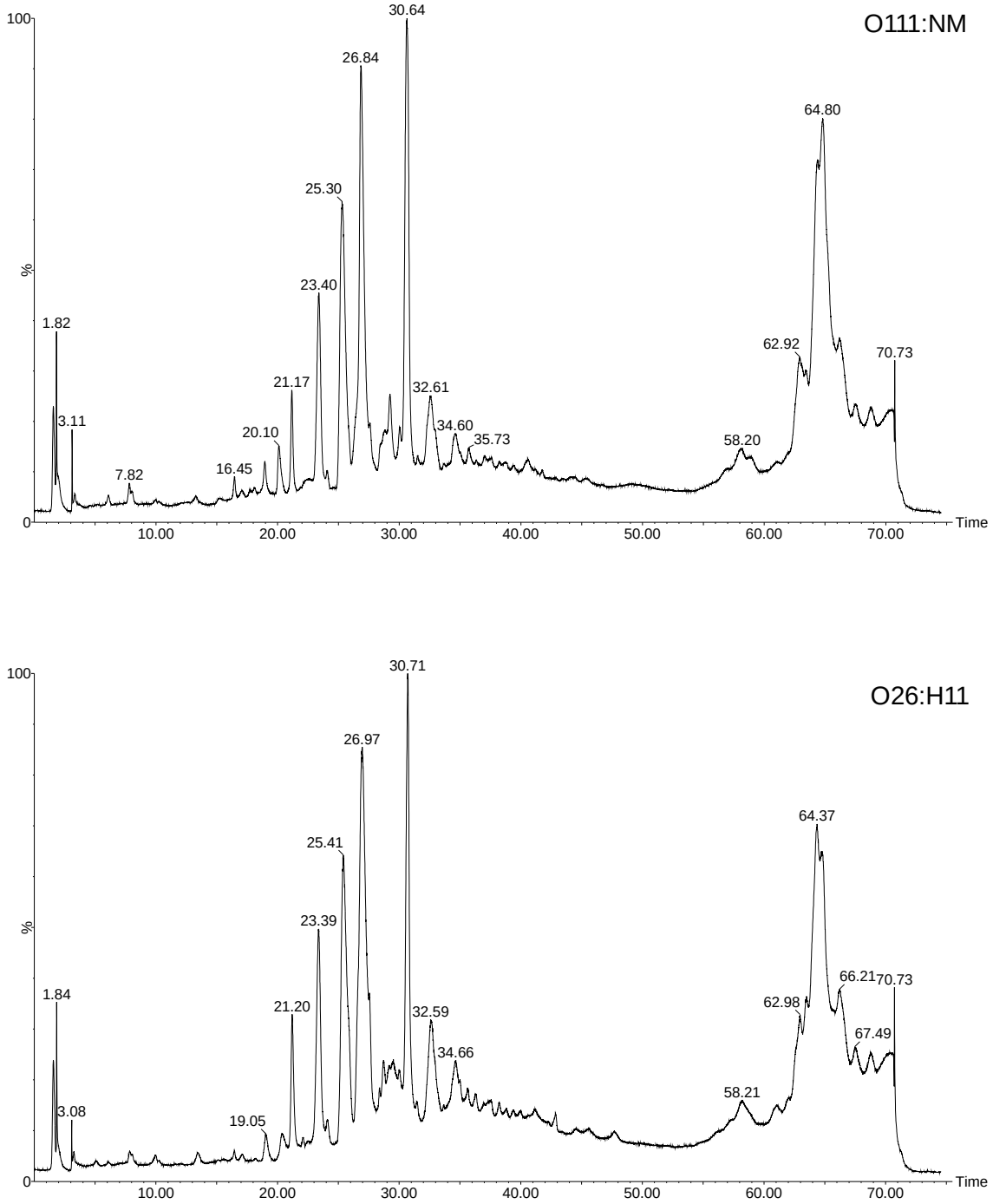


Figure 8a.

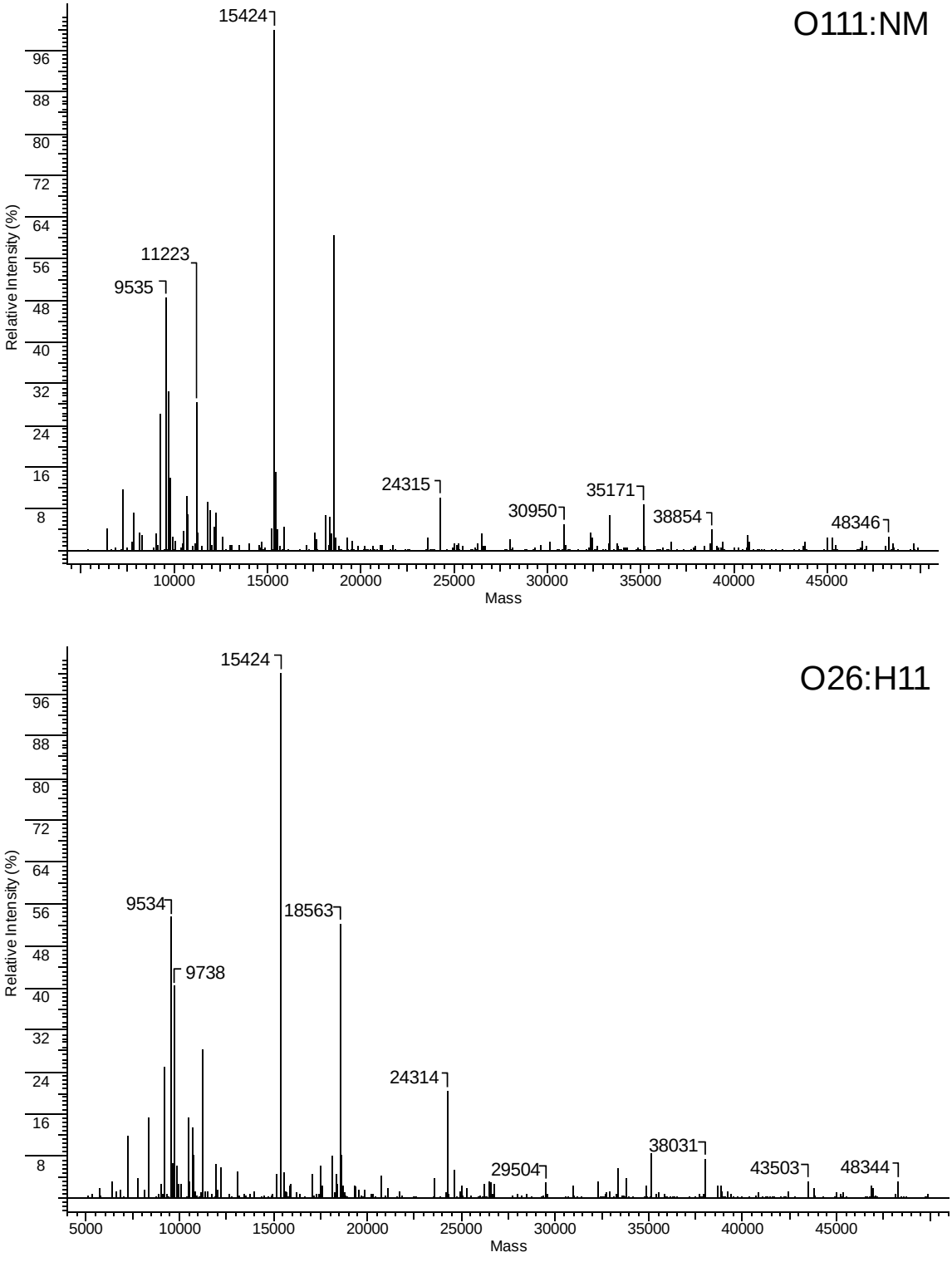


Figure 8b.

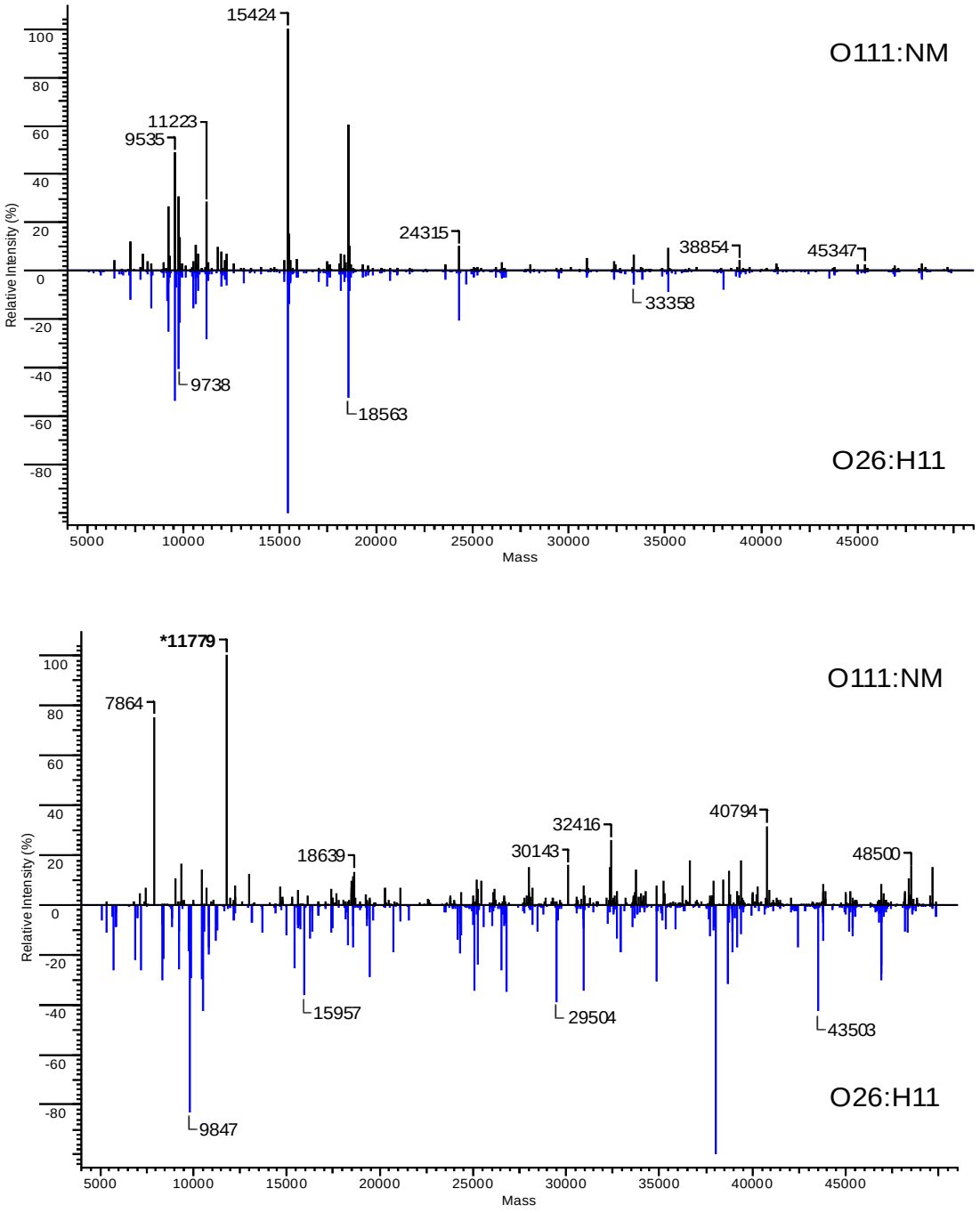


Figure 8. During biomarker discovery, LC/MS chromatograms were obtained in full-scan, positive ion mode (8a). After automated charge state deconvolution using ProTrawler6, a single spectrum showing all of the proteins observed in each chromatogram was generated (8b) and mirrored along the abscissa using MS Manager (8c - top). As can be seen in 8c, a number of similar masses were observed between the two serotypes. Mass spectral subtraction using a ± 2 Da window was subsequently employed leaving only unique masses for each serotype (8d - bottom). The protein at 11,779 Da was reproducibly unique to *E. coli* O111:NM during the comparison of these two serotypes.

4.3.3. Qualitative Markers: Mass and Retention Time

Historically, it has been difficult to distinguish *Escherichia* and *Shigella* by genetic approaches due to cross-reactivity with primers and probes.[64] In fact, it has been argued that these two actually comprise one genus due to considerable overlap between genomes.[65] The protein expression profiles of either *Shigella* or *Escherichia* isolates were easily distinguished from other genera (data not shown), but distinctions between *Shigella* and *Escherichia* were few (Figure 9). While, no genus specific biomarkers were observed, *Shigella* and *Escherichia* were distinguishable using the proteins unique to either *Shigella* species.

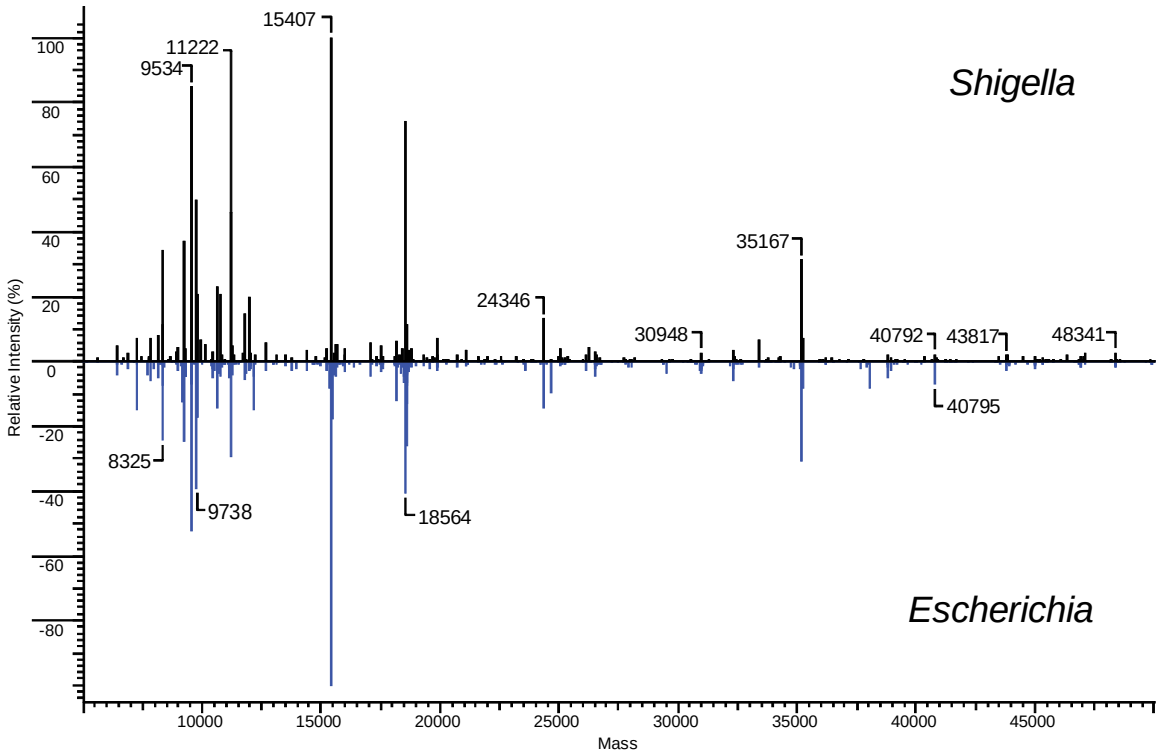


Figure 9. Mass spectral comparison of *Shigella* and *Escherichia* showed significant overlap between these closely-related genera.

PCR primers for *Shigella* speciation are commercially unavailable. However, with LC/MS, distinctions between *Shigella* species were observed. Figure 10 depicts a spectral comparison of *S. flexneri* and *S. sonnei*. The masses in bold marked with asterisks were found in both isolates of that *Shigella* species, yet were not observed in any of the other eight isolates studied. The protein at mass 7,287 unique to *S. flexneri* has the same retention time (27.9 min) and nearly the same mass as a 7,273 Da protein present in all of the *E. coli* and *S. sonnei* isolates studied. This mass difference of 14 Da could be due to a PTM (e.g., methylation), an amino acid substitution (e.g., I for V) or some combination of the two. Either way, such a small difference would likely go unnoticed in a gel based approach or when using a detector with less specificity such as ultra-violet or fluorescence spectroscopy.

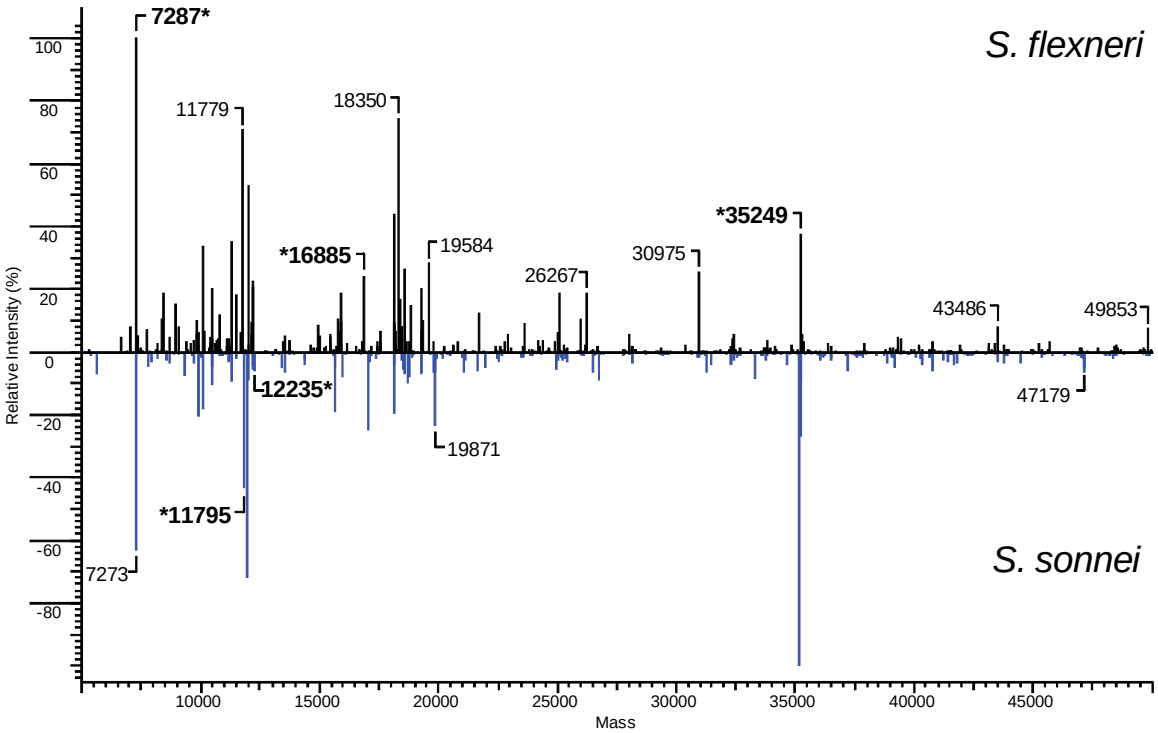


Figure 10. The mass spectra from two species of *Shigella* were both mirrored and subtracted to detect biomarkers for speciation. Unique masses to each species are denoted in bold-type and with an "*".

As an example of serotype differentiation by this approach, the two *E. coli* O157:H7 isolates were compared against the other eight isolates. During this comparison a protein at mass 18,996 eluting at 43.3 minutes was discovered unique to this serotype thereby demonstrating the ability of this method to distinguish EHEC serotypes which are otherwise indistinguishable by clinical symptoms.[53] Table 7 contains the masses and retention times for proteins that were found unique to a group such as to both O157:H7 or to both *S. sonnei* isolates etc.

Table 7. Group specific qualitative biomarkers. Mass (± 2 Da) is listed first followed by retention time (± 0.5 min) in parentheses. * indicates a mass tolerance of ± 3 Da. These markers were present in all 5 replicates.

Group	*E. coli* O157:H7	non O157:H7 EHEC	*S. flexneri*	*S. sonnei*
Unique Proteins	18,996 (43.3)	15,478 (27.1) 24,315 (38.5)	*35,250 (31.4) 16,886 (26.8) 7,287 (27.9)	11,795 (27.3) 12,235 (45.4)

The two non-O157:H7 EHEC have peaks that identify them as a group as well (Table 7), but when each individual isolate (06-1440 or 06-1418) was compared against the other 9, no unique peaks were found. It was suspected however, that one of the isolates might share a genetic similarity with some of the other eight isolates that was not shared with the other non-O157:H7 EHEC. For this reason, the O111:NM (non-motile) and O26:H11 spectra were subtracted only against each other. During this comparison, a protein at 11,779 Da having a retention time of 27.0 minutes was found unique to O111:NM (Figure 8). A protein of this same mass and retention time has also been observed in *S. flexneri* and *E. coli* O157:H7 isolates. Accordingly, during a blind test these two *E. coli* serotypes could be distinguished first by looking for the group specific peaks listed in Table 7 which would classify them as a non-O157:H7 EHEC, then observing a protein at 11,779 Da with a retention time of 27.0 minutes would indicate the sample was *E. coli* O111:NM.

In epidemiological and forensic investigations, techniques that can characterize bacteria at the strain level are desirable for establishing cluster or outbreak relationships via strain relatedness. Highly specific characterization is needed to detect and pinpoint the source of an outbreak such as a particular produce manufacturer or suspected

bioweapons facility. To this end, strain level comparisons between *E. coli* O157:H7 isolates were made. One O157:H7 isolate studied, accession # 06-1464, has shown a reproducible protein at 14,880 Da eluting at 26.9 minutes not observed in the other O157:H7 isolate, accession # 06-1439, or any of the other *E. coli* or *Shigella* samples. The differences observed between these two O157:H7 isolates indicates that the method described here is not only capable of identifying bacteria, but also of discerning small phenotypic differences which could be indicative of the pathogen's origin and growth environment. With the exception of PFGE, which indicated ~ 98% similarity, other established techniques (e.g., serology) found these two isolates to be identical. In addition to the value of establishing strain relatedness during outbreak investigations, the ability to distinguish two strains (such as the ones described above) that while genetically similar are epidemiologically unrelated, is also significant. Figure 11 depicts the comparison of the two *E. coli* O157:H7 spectra with PFGE results in the inset.

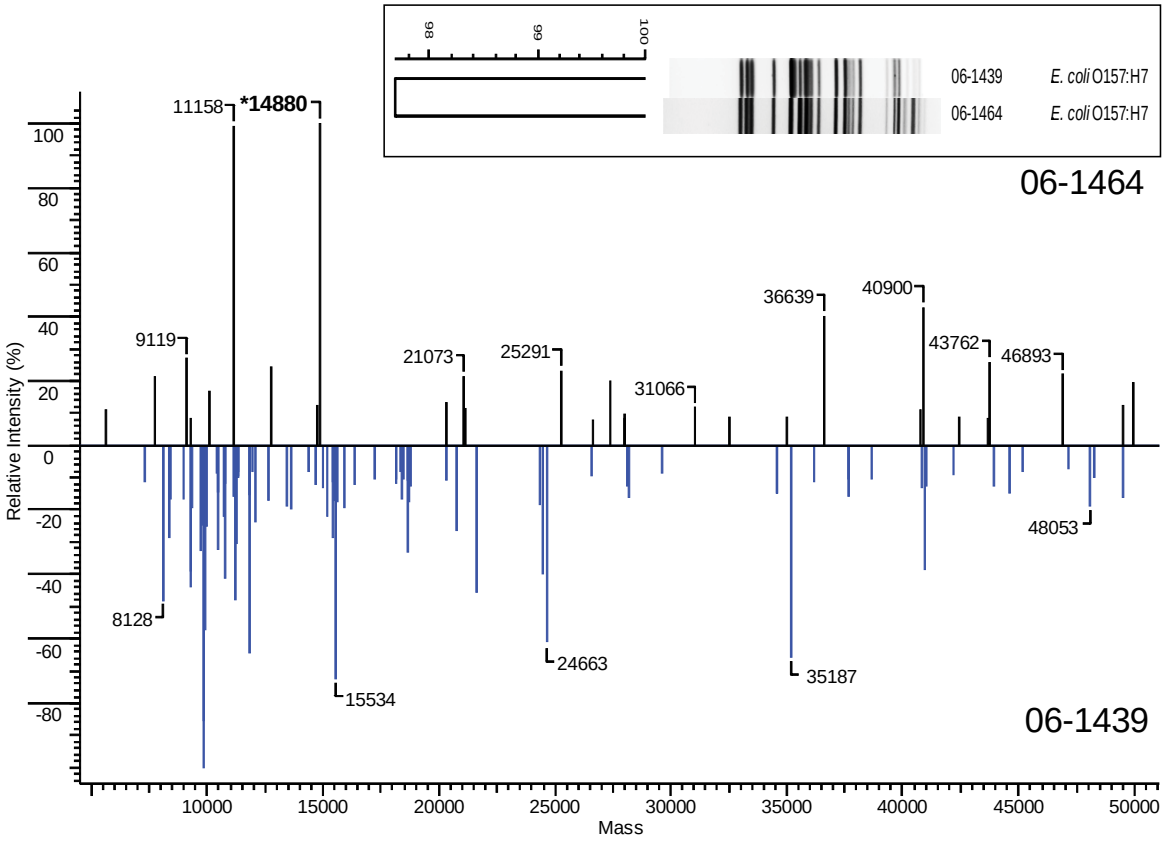


Figure 11. Strain-level comparison of two *E. coli* O157:H7 isolates revealed a reproducible difference at mass 14,880 that was unique to accession # 06-1464. This distinction is noteworthy considering serological and PCR analysis would not be able to distinguish the two isolates from one another. The gold standard method, PFGE analysis, determined ~ 98% similarity between the two isolates after 48 hours (inset).

Analogous to PFGE, in which sequencing of the chromosomal fragments is not performed[66], this approach does not involve sequencing of the biomarkers. The justification being reproducible biomarkers have been observed allowing for characterization at the strain level without knowing the actual identity of the proteins involved. Therefore, this approach could potentially be applied to bacteria whose genomes have not been sequenced. In contrast, proteomic approaches that rely upon database results for identification purposes would have little utility for such bacteria.

Finally, circumventing protein sequencing eliminates a timely digestion step resulting in a reduced analysis time.

4.3.4. Distinction of Isobars Differing in Retention

Since proteins with larger quantities of, or more easily accessible, hydrophobic regions will stay adsorbed to the column longer,[40] retention time can therefore be used to distinguish two different isobaric proteins. This is critical when a sample has two or more different proteins of approximately the same mass. Such was found to be the case for *S. flexneri* 04-0497. This strain of *S. flexneri* has two proteins within 2 Da of mass 18,121 that differ in retention time by nearly 16 minutes (13% B). One protein which eluted at 37.1 minutes had been observed in both *E. coli* O157:H7 and both *S. flexneri* isolates studied. The other protein however, eluted at 21.3 minutes and was present only in *S. flexneri* 04-0497. Techniques yielding only mass and intensity information (e.g., MALDI-TOF) would likely not detect this protein or mistake it as common. Table 8 contains masses and retention times for proteins found unique to an individual isolate and absent from the other nine isolates. Tables 7 and 8 could be used to identify unknowns based on the presence or absence of these proteins.

Table 8. Strain level qualitative biomarkers. Mass (± 2 Da) is listed first followed by retention time (± 0.5 min) in parentheses. These markers were also present in all 5 replicates. ND = not determined.

Species/ Serotype	***E. coli* O157:H7**	***E. coli* ND**	***E. coli* ND**	***S. flexneri* ND**	***S. flexneri* ND**
Accession Number	06-1464	06-0004	06-0006	04-0497	06-0967
Unique Proteins	14,880 (26.9)	11,694 (28.5) 15,931 (30.2) 9,066 (16.3)	11,707 (28.0)	10,119 (27.7) 18,860 (41.9) 18,121 (21.1)	10,089 (28.9) 18,151 (21.5) 18,874 (40.7)

4.3.5. Quantitative Markers

In addition to qualitative aspects (e.g., mass and retention time) that signify biomarkers, proteins that differ in intensity are also informative and may be caused by up or down regulation or possibly genetic engineering (to produce more toxin etc.). The utility of quantitative biomarkers was evident during the analysis of the two non-O157:H7 EHECs. In the other eight samples, the intensity of a protein at 15,406 Da is much greater than one at 15,423 Da, but the trend was reversed for the two non-O157:H7 EHECs. Interestingly, this difference involved two of the most abundant proteins in the lysate. For this reason, the quantitative difference was immediately obvious and no spectral subtraction of common peaks was required.

Strain level quantitative differences were also observed. In *S. flexneri* 04-0497, a protein at mass 9,737 eluting at 26.4 minutes and highly abundant in all 9 other isolates, was barely detected. Additionally, in *E. coli* 06-0006 a protein which elutes at 30.6 minutes weighing 35,171 Da, common to other *E. coli* and *S. sonnei* isolates, is

completely absent– most likely underexpressed below the limit of detection. Alternatively, the gene for this protein could be damaged or turned off or possibly absent so that no protein is being expressed at all making this a qualitative distinction.

4.3.6. Analysis of Unknowns

To challenge the validity of the biomarkers discussed above, a blind study of thirteen isolates distinct from the original ten was performed. In an attempt to identify each unknown, the mass spectra obtained from each of the thirteen isolates were individually screened for the biomarkers listed in Table 7. Upon inspection of the blind study data, one initial observation was shifting retention times for the markers. During the early investigation of known isolates, a retention time window of ± 0.5 minutes was observed. However, during the blind study, analyte retention times seemed more variable indicating an average window of ± 1.0 minute was more suitable. Possible explanations for this variation include degradation of the column or minor differences in the mobile phase composition. This variation was consistent within each run however, thus not affecting the relative retention times of the analytes. When used in conjunction, the retention time, mass and relative intensity (RI) information allowed the biomarkers to be detected with confidence.

Another observation was made concerning two of the three biomarkers for *S. flexneri*, one at 7,287 and one at 35,250 Da. These markers have counterparts in *E. coli* and *S. sonnei* exhibiting the same retention times but at decreased masses of 7,273 and 35,170 Da. During the blind study, these two *S. flexneri* proteins were observed in small amounts (2 - 5% RI) in some of the *E. coli* and *S. sonnei* isolates. There were two

possible reasons for this. One, there was a small amount of *S. flexneri* present in these isolates and they were therefore technically a mixture. Two, the mass spectrometer displayed higher total ion counts during the blind study than in any of the five previous replicates of the known isolates. However, even with the greater ion counts, the *E. coli* and *S. sonnei* counterparts were not observed in any of *S. flexneri* isolates. Since the exact reason(s) was not determined, these two *S. flexneri* proteins were at least for the unknown isolates examined here, best used as quantitative biomarkers rather than qualitative.

Using the two *S. flexneri* proteins as quantitative markers, all biomarkers were present *and* absent as expected allowing all thirteen unknown isolates to be correctly identified. In total, there were three *S. sonnei*, three *S. flexneri*, four *E. coli* O157:H7, one *E. coli* O26:H11 and two *E. coli* O111:NM isolates identified. The time required to collect, process and examine the data to determine the identity of the unknown isolates was approximately two hours per sample post culture.

4.4. Conclusions

Using LC/MS, a technique commonly found in analytical chemistry laboratories but rarely found in diagnostic microbiology laboratories, a complementary approach with certain advantages over typical microbiological methods has been developed. This study used as a proof of concept ten isolates of *Shigella* and *Escherichia* including the species most often implicated in human disease to examine the efficacy of the LC/MS approach to characterize microorganisms. This approach has demonstrated greater specificity than obtainable using current real-time PCR protocols, allowing for distinctions at the strain

level. Furthermore, this method is automated and is less labor intensive compared to PFGE, the gold standard for subtyping. Analogous to PFGE data, LC/MS results from various public health laboratories could be uploaded into a public database such as PulseNet which is utilized to detect outbreaks around the country.[67]

Using a combination of automated charge state deconvolution and spectral subtraction, reproducible intact protein biomarkers were observed at the species and sub-species (serotype, strain, etc.) levels, and were successfully implemented for the identification of unknown pathogens. These biomarkers and their corresponding protein expression profiles could be used to enhance public health response (treatment, remediation, etc.) by yielding insight into the identification of unknown bacteria. Highly specific and accurate identification of microbial pathogens in a timely manner is essential to guide the data directed decision making of hospital staff, public health officials and investigating authorities. No primers, antibodies, or proteomic database searches were required for this study. This was of particular interest for *Shigella* speciation for which PCR primers are commercially unavailable. Although no sequencing was performed here, these protein biomarkers could be sequenced, and that information could be used to reverse-engineer novel PCR primers.[37, 68, 69] Likewise these biomarkers could be purified for the production of antibodies to enhance serological investigations (e.g. protein microarrays).[70]

Ultimately, this method would be ideal as a complementary technique adding another dimension to the polyphasic approach of bacterial identification. This enhanced polyphasic approach would then lead to more confident results which are required to

cease production and distribution during an outbreak, or for the prosecution of suspected terrorists. Additionally, this LC/MS method could be expanded to monitor biomarkers for other foodborne pathogens and implemented to screen food items before they enter the market to prevent possible outbreaks from occurring.

CHAPTER 5 Characterization of *Clostridium* Species Utilizing Liquid Chromatography/Mass Spectrometry of Intact Proteins

5.1. Introduction

5.1.1. Natural Routes Exposure

Ubiquitous in the environment, *Clostridium* species can be found in soil, dust and in the intestinal flora of humans and animals.[71] They are anaerobic, rod-shaped sporeformers[72] and produce more toxins than any other genus of bacteria.[73] In addition to their role in human disease (e.g. gastroenteritis), pathogenic *Clostridia* also cause illness in both domestic and wild animals.[74] Of the approximately forty-five species that cause clinical conditions,[75] of particular interest to this study were the species *C. difficile* and *C. perfringens*.

C. difficile is the leading cause of the nosocomial illness antibiotic associated diarrhea (AAD), a significant cause of morbidity and mortality particularly among hospitalized elderly and immunocompromised.[76] *C. difficile* related AAD is responsible for increased hospital stays, resulting in an annual economic burden of $1.1 billion in the U.S. alone.[77] Outbreaks in both long-term care facilities and hospitals have been reported.[78] Although not typically recognized as a foodborne pathogen, *C. difficile* has been found in retail ground meat[79] and commercially vacuum-packed raw meat.[80] Additionally, PCR ribotypes 017 and 027, which have been implicated in outbreaks, were

isolated from both humans and cattle raising concerns that *C. difficile* may be an emerging foodborne pathogen and indicating the possibility of zoonotic transmission.[81]

Although less common, *C. perfringens* is also a causative agent of AAD,[76, 82] responsible for up to 15% of the cases.[83] More known for its role in foodborne disease, *C. perfringens* was listed in the top ten by the Centers for Disease Control and Prevention (CDC) for microbial causes of foodborne outbreaks from 1972-2000.[51] Additionally, from 1993-1997 alone, fifty-seven outbreaks of *C. perfringens* food poisoning were reported causing an estimated 250,000 cases per year, yielding costs of approximately 500 million dollars. Although *C. perfringens* has a high infectious dose (~ 10^8 cells), it is fast growing, with generation times in ground beef at 41°C, as small as 7.1 minutes being reported.[84] At this rate, a single *C. perfringens* cell can grow to over the infectious dose in just 3.5 hours. Beyond the problems associated with ready-to-eat foods, infection control is challenging, as normal cooking temperatures which kill most bacteria, are insufficient for killing *C. perfringens* spores.[85]

5.1.2. Potential for Deliberate Use

If natural outbreaks of these organisms have had such considerable public health impact, of even greater concern is the threat of these pathogens being used intentionally. The ease of access, and the toxin producing and spore forming capabilities of *Clostridia* make them viable options as bioweapons, and their cultivation for deliberate use has been documented. The epsilon toxin of *C. perfringens* is considered by the CDC a category B select agent[51] and spores of *C. perfringens* were stockpiled by the Iraqi government

during the first Gulf War.[86] Additionally, cultures of *C. difficile* were grown by the Japanese terrorist organization Aum Shinrikyo in the mid 1990s.[87]

5.1.3. Review of Current Methods of Analysis

In the event of an infectious disease outbreak or terrorist attack, key decisions by hospital staff, public health officials and investigating authorities are required to guide medical treatment, prophylaxis and remediation. These decisions are best made when they are data directed. Since individuals with *C. difficile* and *C. perfringens* infections would exhibit similar symptoms, molecular differentiation is needed. Two common molecular approaches for diagnosing *Clostridia* are enzyme immunoassay (EIA) and real-time PCR.

EIA methods are rapid and sensitive, and are typically used to target toxins. However, cross reactivity is a common problem leading to a high false positive rate which can misguide public health response. Additionally, the toxins of *C. perfringens* and *C. difficile* are unstable and can degrade quickly, which can lead to false negatives if stool samples are not analyzed soon after collection or properly refrigerated.[88, 89] Real-time PCR methods, are typically employed to amplify toxin genes. Difficulties with this method can arise since many of the toxin genes in *Clostridium* spp. reside on extrachromosomal elements (e.g. plasmids or phages) and can be horizontally transferred to other types of bacteria or even within different *Clostridium* species.[90] The transfer of genetic material between *C. perfringens* and *C. difficile* has been reported[91] and this could be problematic for methods targeting only one or few genes of a single species. Furthermore, the mere presence of toxin genes does not guarantee that the genes are

actually being expressed, which could result in a false positive indication of pathogenicity. This is noteworthy since many bacterial pathogens possess genes that are not expressed.[25-27]

Both EIA and real-time PCR are targeted approaches and not capable of providing strain level typing which is desired during outbreak investigations. Due to its high degree of specificity, the gold standard during outbreak investigations is pulsed-field gel electrophoresis (PFGE). Although highly specific, PFGE is labor intensive, not easily automated and time consuming - requiring a minimum of two days post culture. Since each of these molecular methods has its strengths and weaknesses, several reviews regarding the identification and characterization of bacteria have all reported a common conclusion: the highest confidence approach is polyphasic i.e., combines information from several complementary yet distinct techniques.[13, 16, 28, 63] One molecular technique which can add a dimension to the polyphasic approach is mass spectrometry.[92]

5.1.4. Using Mass Spectrometry to Characterize Bacteria

Mass spectrometry was first applied to the problem of microbial characterization in the 1970s[93] and subsequent books and review articles have been dedicated to the subject.[28, 94, 20] Since that early application, mass spectrometry has been used to characterize bacteria by detecting lipids,[95] carbohydrates,[96] and nucleic acids.[97] The approach described here is distinct in that it uses reversed phase liquid chromatography/mass spectrometry (LC/MS) to monitor intact proteins and is capable of providing information regarding protein mass and relative hydrophobicity (retention time). Key differences between closely related bacteria may occur within the proteome

to which genetic approaches are insensitive (e.g. posttranslational modifications, PTMs) making this a complementary approach. PTMs are common in bacteria and one of the more dominant PTMs, removal of the *N*-terminal methionine, results in a mass shift of -131 Dalton (Da)[98] and is estimated to occur in 50% of bacterial proteins.[33]

5.1.5. Specific Aims

The advantages of the polyphasic approach to characterizing bacteria, the significant socio-economic burden posed by pathogenic *Clostridia*, and the need of public health officials to have sound data to assist with decision making, all act as a driving force for the continuing research and development of novel molecular assays. The aim of this study was to examine the efficacy of the LC/MS approach for characterizing pathogenic *Clostridium* species. The goal was to develop a method that can provide complementary information compared to existing protocols and yield information that may also be used to enhance current approaches. The method described here is threefold: discovery of reproducible protein biomarkers, implementation of those biomarkers for the speciation of *C. perfringens* and *C. difficile* and characterization of each species at the strain level. Finally, the validity of the species specific markers was further challenged by applying them for the identification of unknowns in a blind study.

5.2. Methods

5.2.1. Materials

HPLC grade solvents (acetonitrile, formic acid and trifluoroacetic acid (TFA)) were purchased from Fisher Scientific (Fairlawn, NJ) and 2-propanol was purchased from Honeywell Burdick and Jackson (Morristown, NJ). The water utilized for HPLC analysis

was purified in-house to yield organic-free 18.3 MΩ × cm using an E-pure purification system (Barnstead International, Dubuque, IA). Sterile water that had been autoclaved and purified with a RiOs 5 Water Purification System (Milipore, Billerica, MA) was used during bacteria preparation.

5.2.2. Growth and Lysis

C. perfringens clinical isolates were obtained from the Virginia Division of Consolidated Laboratory Services and *C. difficile* isolates were purchased from American Type Culture Collection (ATCC, Manassas, VA). Since no official nomenclature for describing strains exists, in-house accession numbers were used to describe the eight *C. perfringens* isolates as different strains and the five *C. difficile* strains were distinguished by their ATCC product number. The following strains were utilized during the portion of this study where the isolates' identity was known: *C. perfringens* (06-0385, 06-0387, 05-0025, 05-0070, 05-0076, 04-1664, 04-1672 and 04-1665); *C. difficile* (43593, 43594, 43596, 43598 and 700792).

The *C. perfringens* cells were grown for 24 hours; however, the more fastidious *C. difficile* required 48 hours. To shorten the time required to reach the desired cell concentration, twice the number of plates were used to cultivate *C. difficle* and after 24 hours, the cells from two plates were combined and added to form one cell suspension. Both species were grown at a temperature of 37^{o}C. Trypticase soy agar plates containing 5% sheep's blood under anaerobic conditions were used as the growth medium. After this growth period, cells were removed from the plate and placed in a test tube containing 1 mL of sterile water until the optical density reading reached 1.0 using a MicroScan

Turbidity Meter (Dade Behring, West Sacramento, CA). A 500 µL aliquot of this suspension was washed three times with 500 µL of water followed by centrifugation (6000 × *g* at room temperature for 5 minutes) to remove residual media.

Two protocols for lysis were employed. After washing the cells as described above, the first method (lysis protocol 1, LP1) involved adding 150 µL of 1 mg/mL lysozyme (HEWL) (Sigma-Aldrich, St. Louis, MO) in 20 mM NH_4OAc and incubating at 37°C for 30 minutes. This was followed by four cycles of freeze/thaw in liquid nitrogen and a 37°C water bath respectively. The second lysis protocol (LP2) began with 75 µL of 1 mg/mL lysozyme in 20 mM NH_4OAc and incubating at 37°C for 30 minutes, followed by four cycles of freeze/thaw. Then, 75 µL of 1:1 H_2O: acetonitrile, 5% (v/v) TFA was added.

Neither lysis protocol was capable of inactivating spores, so they were removed by filtration. After centrifugation at 4100 × *g* for 4 minutes at room temperature to pellet the debris and clarify the supernatant, the supernatant was filtered using a disposable syringe, a blunt tip needle (BD, Franklin Lakes, NJ), and a 0.22 µm, 4 mm polyvinylidene fluoride (PVDF) low protein binding GV filter (Milipore). Following centrifugation, 65 µL of supernatant was removed and placed in an autosampler vial for analysis.

5.2.3. LC/QTOF MS Analysis

Intact proteins were separated by reversed phase chromatography using an Acquity liquid chromatograph (Waters, Milford, MA). Gradient elution (5-55% B in 60 min) was used at a flow rate of 0.225 mL/min where A = H_2O (1% formic acid) and

B = 2-propanol (1% formic acid). The column was a non porous Prosphere P-HR 2.1 × 150 mm, 4 μm particle size (Alltech, Columbia, MD) operated at 50°C. The autosampler was maintained at 15°C prior to administering the injection volume of 20 μL.

The LC was directly interfaced to a Q-TOF Premier (Waters) mass spectrometer utilizing positive ion electrospray ionization. The capillary and cone voltages were 3900 and 40 V respectively. The desolvation gas (900 L/h) was heated to 500°C and the source temperature was 115°C. Ions were monitored over a mass range of 620 - 2450 Da and resolved in single reflectron (V) mode. An acquisition rate of two spectra/sec was utilized. Other parameters employed in the MS method were optimized for sensitivity and resolution using bovine serum albumin. Data was collected using MassLynx software version 4.1 (Waters).

5.2.4. Data Processing

LC/MS data was processed using two software packages: Protrawler6 and MS Manager. Protrawler6 (Bioanalyte, Portland, ME) software provided automated deconvolution of multiply charged ions by first dividing the full-scan data from the chromatogram into time intervals (30 seconds) and summing the data from each interval. Sequential deconvolutions were then performed to obtain neutral masses of the proteins that eluted during each interval. A text file containing the neutral masses, intensities and retention times was then created summarizing the results for each chromatogram.[6] Retention time information can be used for further study (e.g. fraction collection) of proteins of interest or to distinguish proteins of the same mass that differ in retention. The masses and intensities were used to create a single spectrum representing all of the

proteins observed in the lysate utilizing MS Manager (Advanced Chemistry Development Laboratories, Ontario, CA).

To further facilitate biomarker discovery, MS Manager was employed for spectral mirroring and spectral subtraction. Spectral mirroring allowed spectra to be mirrored along the abscissa, placing the baseline at the center of view. Spectral subtraction removed all common peaks between two spectra within a given mass accuracy so that only unique ones remained. For group and strain level comparisons involving multiple spectra, the text files of all isolates not in that group or strain were combined to create a hypothetical spectrum which could then be used for subtraction. A subtraction window of ± 3 Da was utilized.

5.3. Results and Discussion

5.3.1. Development of Gram Positive Lysis Protocols

Early work in this laboratory using LC/MS to discover biomarkers for pathogenic bacteria involved Gram negative enterics (e.g. *Escherichia coli* and *Shigella* spp.). The lysis procedure used in that study was straightforward and involved reconstituting the cell pellet in a solution of 50% acetonitrile with 0.1% (v/v) TFA. This lysis procedure yielded complete inactivation of the cells and efficient protein extraction. However, upon applying the same method to the Gram positive *Clostridia*, the cells were neither inactivated nor allowed sufficient protein recovery for characterization. For this reason, a new lysis procedure was developed. Several attempts were made using methods such as lysozyme incubation, freeze/thaw cycles and adding more acid to the Gram negative protocol. Individually neither of the methods yielded protein recovery similar to that

found with the Gram negative bacteria, but when used in conjunction, two protocols emerged that provided sufficient protein recovery.

The first protocol (lysis protocol 1, LP1) began with a 30 minute incubation at 37^{o}C after suspending the cells in 150 μL of lysozyme. This was followed by four cycles of freeze/thaw in liquid nitrogen and a 37^{o}C water bath respectively. The second protocol (lysis protocol 2, LP2) was the same as LP1, only 75 μL of lysozyme was used and after the freeze thaw cycles, 75 μL of 50% acetonitrile (50% water) with 5% (v/v) TFA was added for a total volume of 150 μL. Both of these protocols enabled sufficient protein recovery, but interestingly, each method extracted a distinct set of proteins (i.e., by mass and retention time). As depicted in Figure 12, the proteins from LP1 were generally larger in mass and retained longer on the column (were more hydrophobic), while the proteins from LP2 were generally lower in mass and exhibited less retention (were more hydrophilic).

After failed attempts to combine the lysates prior to injection, it was clear that most of the proteins from LP1 were not soluble in the acid/organic solution used in LP2 as a precipitate was formed after combination. This explains why few proteins from LP1 were observed in LP2. Since the proteins from LP2 were absent from LP1, LP2 must be further altering the cellular structure in a way that LP1 is not. LP1 employs lysozyme incubation which hydrolyzes the β-1,4 linkages of the polysaccharide chains within the peptidoglycan layer and when combined with four freeze/thaw cycles, weakens the overall structural integrity of the cell. LP2 utilizes essentially the same two initial steps, but by adding a denaturing solvent containing 50% organic and 5% acid, further disrupts

intracellular components and/or complexes (e.g., ribosomes) and extracts only acid and organic soluble proteins. Three of the four marker proteins found for each species using LP2 (vide infra) matched masses in the ribosomal protein list within the Rapid Microorganism Identification Database (http://www.RMIDb.org/), while none of the species marker masses from LP1 matched. Sequencing however, would provide more conclusive information regarding the types of proteins extracted by each protocol.

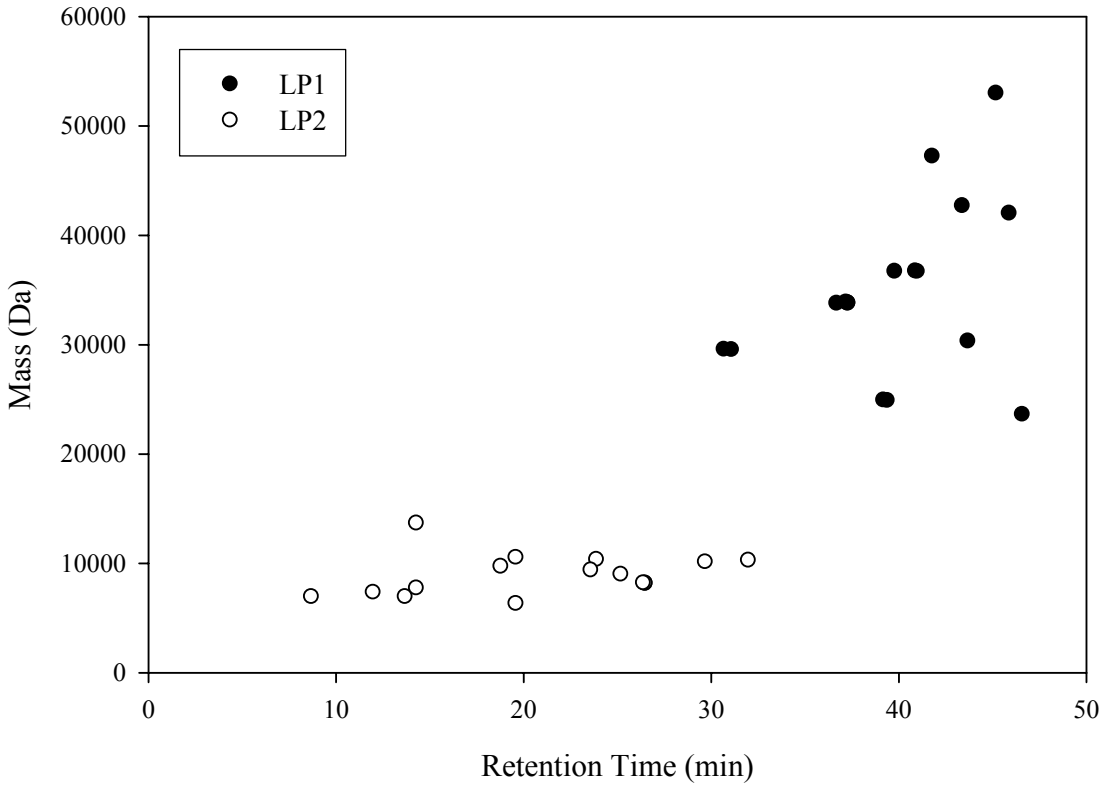


Figure 12. Mass vs. Retention Time plot demonstrating the differences in proteins extracted by the two lysis protocols: LP1 and LP2. Data from *C. perfringens* 05-0070 is shown. For clarity, all proteins below 10% relative intensity were not included in this figure.

Given that the Gram negative protocol was unsuccessful for Gram positives and the two newly developed Gram positive protocols LP1 and LP2 had not been tested on

Gram negative bacteria, prior to the analysis of an unknown, a Gram stain would be required. To address this issue, both of the Gram positive lysis protocols were applied to the Gram negative *E. coli* and *S. sonnei*. Both protocols enabled sufficient lysis and inactivation and yielded similar protein extraction results in terms of mass and retention behavior as seen with the Gram positive *Clostridia*. New proteins were observed that were not previously observed using the Gram negative lysis protocol. While neither method displayed all of the biomarkers found during the Gram negative study, when the results from both methods were combined, all previously observed biomarkers were present (data not shown).

Combined, LP1 and LP2 can be employed regardless of the type of bacteria, precluding the need for a Gram stain. While not very time consuming (~ 10 min) and informative, during an outbreak or terrorist event, the elimination of any unnecessary steps would be beneficial. In addition, these lysis methods would be beneficial for Gram variable bacteria or ones that are unresponsive to the Gram stain. It also important to note that unlike some lysis procedures that employ detergents (e.g. SDS) or chaotropes (e.g. urea), the lysates from LP1 and LP2 are amenable to mass spectrometry analysis without further preparation. Furthermore, these methods are simple and can be easily performed under a hood preventing issues with containment and aerosolization that may occur with mechanical techniques (e.g. French press).

5.3.2. Biomarker Discovery

As mentioned previously, the work described herein has three elements: biomarker discovery, speciation and strain level characterization. After lysis, the process

of biomarker discovery begins with chromatographic separation of the intact proteins. Reversed phase chromatography separates proteins based on their ability to interact with a hydrophobic stationary phase, with the more hydrophilic proteins being less retained. This type of chromatography was employed rather than ion exchange or size exclusion chromatography because the mobile phases are amenable to mass spectrometry analysis and are easily removed during fraction collection.

Since electrospray ionization was used in this study, multiply charged ions were produced. In order to determine the neutral mass of the proteins, the mass spectra containing multiply charged ions were deconvoluted using Protrawler6 software. This process is performed throughout the entire chromatogram and the information is summed forming a single mass spectrum depicting the neutral masses of all of the observed proteins from the lysate. A mass spectrum shows the mass of the proteins on the X-axis and the intensity or abundance of the proteins on the Y-axis. The mass spectrum contains the same information as a one dimensional gel, only that the abundance is indicated by the height of the peak rather than the color intensity of the spot. It should be noted however that the mass accuracy and precision using a TOF mass spectrometer is significantly greater than using a 1D gel, and the improved mass resolution allows proteins to be confidently distinguished that are 3 Da apart.[99]

After a single mass spectrum was created from the cell lysate, the mass spectra were compared by spectral mirroring and spectral subtraction using MS Manager software. Mirroring allows viewing two spectra along the same baseline to better observe differences. Then, proteins common to both spectra were removed by subtraction using a

user defined mass window (± 3 Da was used here). After subtraction, only unique proteins remain that may qualify as biomarkers. A protein was deemed a biomarker only if its combined mass (± 3 Da) and corresponding retention time (± 0.5 min) were found unique after three repeated experiments performed on three different days indicating it was reproducibly unique.

5.3.3. Speciation

Both lysis protocols enabled the two *Clostridium* species to be distinguished. As seen in Figure 13a and 13b, all of the markers from LP2 weigh < 14 kDa, but when LP1 was used the markers for each species contained a unique protein weighing > 30 kDa. Additionally, each species specific marker from LP1 eluted after 28.6 minutes while all markers from LP2 eluted prior to 28.6 minutes. Although both lysis protocols enabled the species to be distinguished, if rapid speciation is the goal, only LP2 could be used due to the earlier elution of its proteins. Since patients with AAD would exhibit similar clinical presentation regardless which species of *Clostridium* was responsible, identification of the cause by symptoms alone is not possible. While similar symptoms are produced, speciation is important as the risk factors for each species are different and for this reason species specific control measures have been recommended.[76]

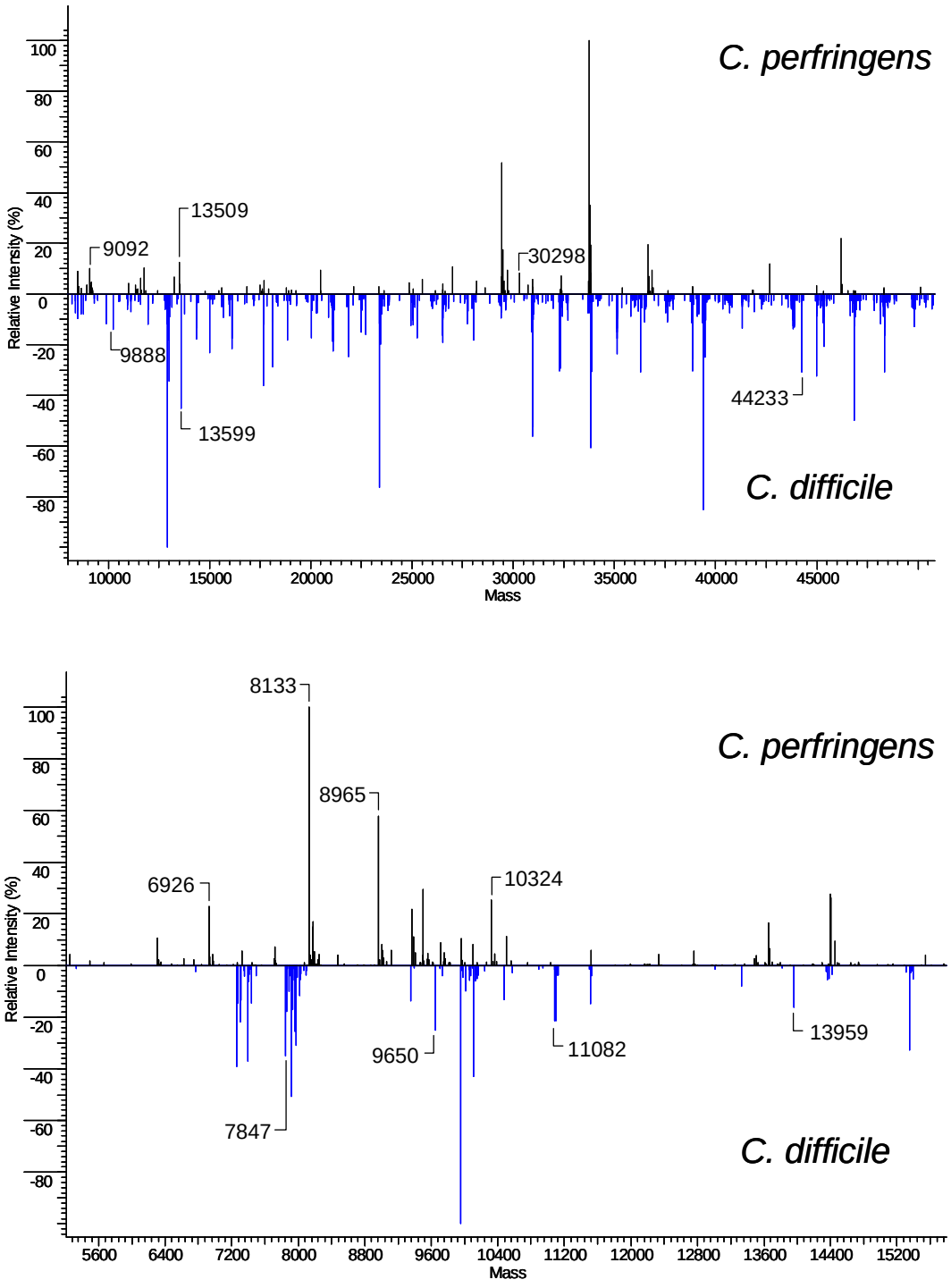


Figure 13. Speciation using both lysis protocols LP1 (top) and LP2 (bottom). Only the biomarker masses are labeled. All markers extracted from LP2 eluted prior to 28.6 minutes, while all markers from LP1 eluted after 28.6 minutes. While many other proteins appear unique after subtraction, only the labeled masses were reproducibly unique.

5.3.4. Sub-species Determinations

In addition to inter-species comparisons, intra-species comparisons were also performed. One example using *C. difficile* 700792 is shown in Figure 14. In this figure, the results from both lysis protocols were combined in order to view them simultaneously. The bottom shows the remaining four *C. difficile* isolates with 700792 on top. The masses marked in bold and denoted with an asterisk were reproducibly unique to this strain. The lower molecular weight protein at 8,933 Da eluted at 16.4 minutes and was extracted using LP2, while the larger protein at 19,548 Da eluted at 40.7 minutes was obtained using LP1. Strain level discriminations were achieved for both species. This type of analysis is beneficial especially for *C. difficile* since highly virulent, drug-resistant strains have been reported. These strains have been deemed 'superbugs' similar to methicillin-resistant *Staphylococcus aureus* (MRSA), but in the UK have caused more illness and mortality than MRSA.[100] Additionally, *C. difficile* superbugs are potentially more challenging to disinfect since unlike *S. aureus*, *C. difficile* is sporogenic. Identification of these strains would direct treatment towards specific antibiotics and due to the virulence of these strains, would likely expedite infection control measures such as quarantining infected individuals etc.

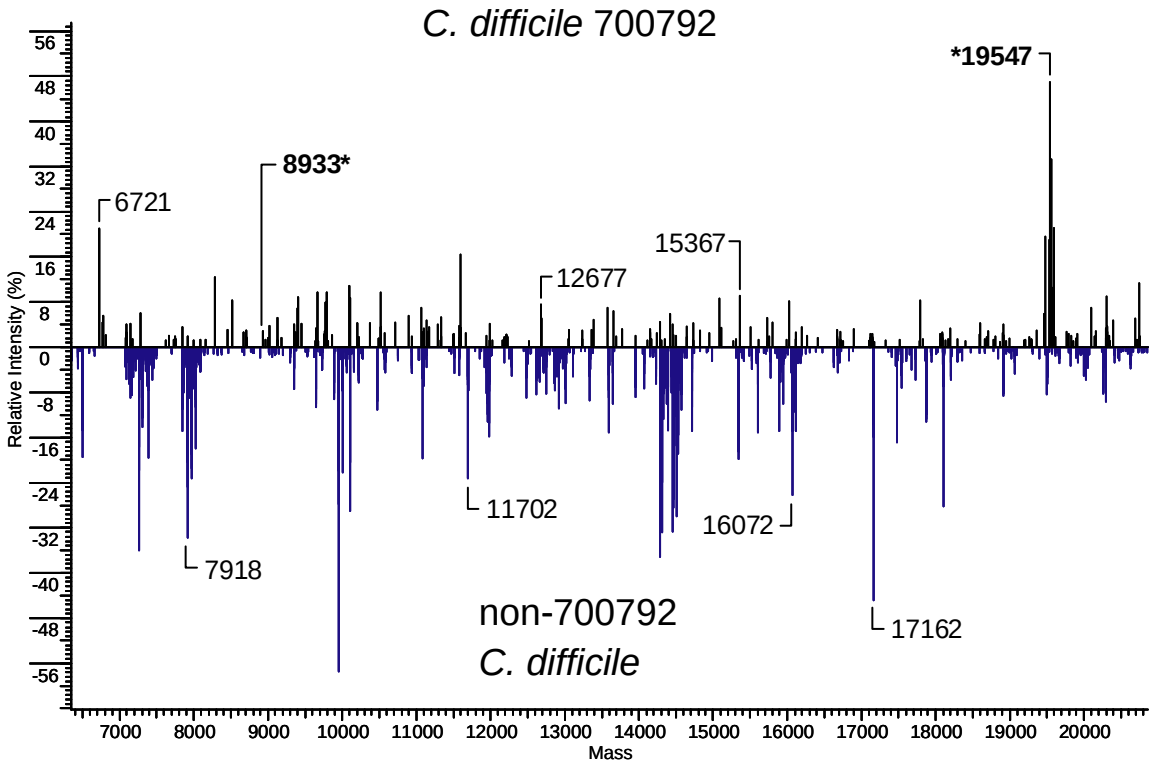


Figure 14. The data from both lysis protocols are combined to display proteins unique to the 700792 strain of *C. difficile* (marked in bold with *). These proteins were observed in all three repeated experiments performed on three different days. The protein weighing 8,933 Da eluted at 16.4 minutes and was obtained using LP2, while the protein weighing 19,547 Da eluted at 40.7 minutes and was obtained using LP1.

5.3.5. Isolates Indistinguishable by PFGE

As mentioned previously, the gold standard for strain typing during outbreak investigations is PFGE. For this reason, comparisons were made between the strain typing capabilities of PFGE and LC/MS. Two sets of the *C. perfringens* isolates studied (06-0835 and 06-0387; and 04-1464, 04-1465, 04-1672 and 05-0070) were indistinguishable by PFGE. Figure 15 depicts the restriction patterns for all of the *C. perfringens* isolates examined in this study, obtained using *Sma*I as the restriction enzyme. Utilizing LC/MS, the strains in both of these sets could be distinguished

incorporating proteins from both lysis protocols (Table 9), indicating the high discriminatory power of this approach. Since more markers were obtained using LP1, this method was more informative for strain typing than LP2. Though beneficial, the reversed phase chromatography step rarely exhibits the specificity needed to distinguish strains. However, a clear difference was observed in the chromatograms of 06-0385 and 06-0387 (Figure 16).

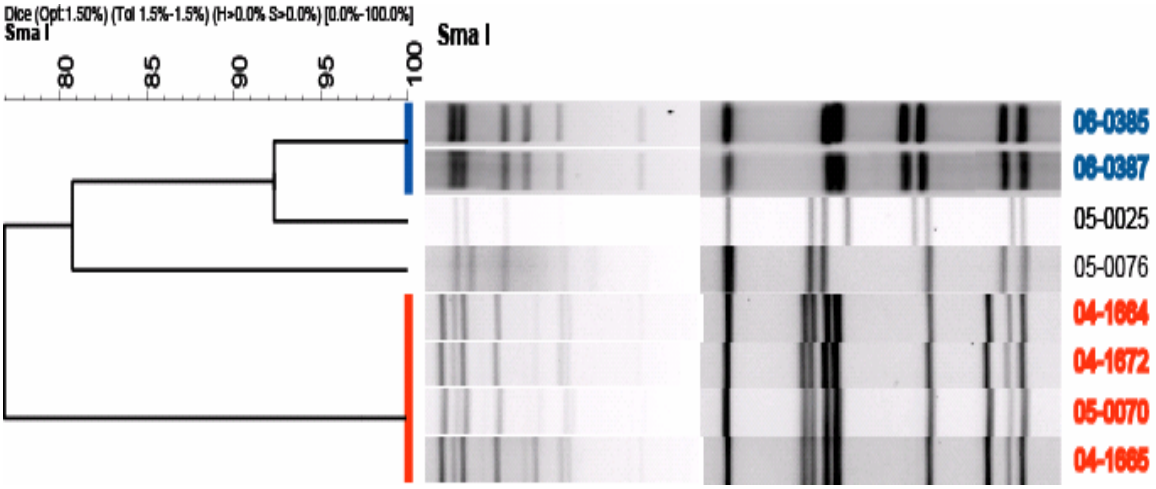


Figure 15. Restriction patterns obtained after PFGE analysis of the eight *C. perfringens* strains employed in this study. *Sma*I was the restriction enzyme used.

Table 9. Proteins found unique to each strain when compared with the other strains that yielded 100% similar restriction patterns. Proteins marked I, were obtained from LP1 and proteins marked II were obtained using LP2. The masses are (± 3 Da) and the retention times listed in parentheses are (± 0.5 min). These proteins were observed in all three repeated experiments performed on three different days.

06-0385	06-0387
9,400 (22.8)^{II} 13,503 (25.7)^{II} 22,160 (34.9)^{I} 32,344 (54.0)^{I} 42,218 (42.8)^{I}	12,334 (28.6)^{II} 34,086 (42.2)^{I} 42,674 (43.4)^{I} 51,374 (37.2)^{I}

04-1664	04-1665	04-1672	05-0070
11,031 (31.9)^{II} 18,849 (43.1)^{I} 52,948 (45.3)^{I}	51,398 (37.5)^{I}	9,136 (18.9)^{II} 9,159 (31.0)^{I}	29,532 (30.7)^{I} 30,706 (38.2)^{I} 41,998 (45.8)^{I}

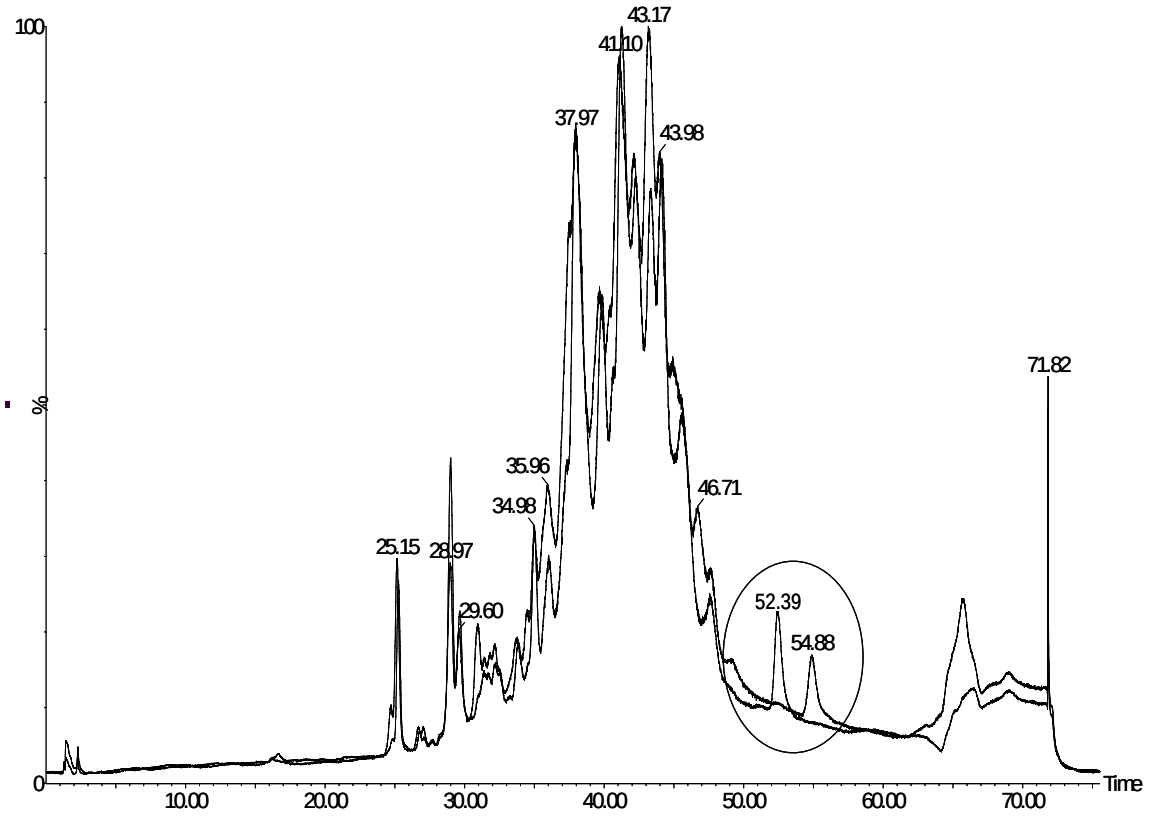


Figure 16. Overlaid chromatograms of two *C. perfringens* strains (06-0385 and 06-0387) that were indistinguishable by PFGE. 06-0385 had protein weighing 32,346 Da which eluted at 54.9 minutes, while 06-0387 had a protein weighing 32,385 which eluted at 52.4 minutes.

In addition to biomarker discovery and speciation, the LC/MS method described here could be used to assist forensic and epidemiological investigations into a terrorist event or outbreak. Being able to distinguish isolates that were indistinguishable by PFGE, the information obtained from this method could not only serve to discriminate between strains, but may also provide insights into phenotypic differences such as the growing conditions or traveling patterns of a single strain. These findings also encourage modification of the current PFGE protocols used in this laboratory to include different or multiple restriction enzymes for added information and specificity. One additional restriction enzyme that could be included is *Sac*II, which has proven useful for the analysis of both *C. perfringens* and *C. difficile*.[101]

5.3.6. Identification of Unknowns

To further challenge the biomarkers that were utilized for speciation, they were applied for the identification of unknowns in a blind study. Both lysis protocols were used during this study and the markers from each protocol correctly identified the species for every unknown tested. Although several months had elapsed in between the time that the known isolates were initially examined and the unknown study was performed, all markers were present within the accepted tolerances of mass accuracy (± 3 Da) and retention time (± 0.5 min) (Table 10). It is important to note that not only was the presence of the biomarkers for each species informative, but also, their absence in the other species provided supporting evidence toward an unknown's identity. In total, ten unknowns which had not been previously analyzed by this method (six *C. perfringens* and four *C. difficile* isolates) were correctly identified indicating that these markers were

both reproducible and applicable for the speciation of unknown *Clostridium* isolates. The time required to collect, process and examine the data to determine the identity of the unknown isolates was approximately two hours per sample post culture.

Table 10. Species specific biomarkers. The masses are (± 3 Da) and the retention times listed in parentheses are (± 0.5 min). These proteins were observed in all three repeated experiments performed on three different days. LP1 = lysis protocol 1, LP2 = lysis protocol 2.

	C. difficile	*C. perfringens*
LP1	9,888 (43.5)	9,092 (28.8)
	13,599 (40.0)	13,509 (39.4)
	44,233 (43.5)	30,298 (44.0)
LP2	7,847 (25.1)	6,926 (13.7)
	9,650 (16.9)	8,133 (26.6)
	11,082 (14.6)	8,965 (25.3)
	13,959 (28.4)	10,324 (24.0)

5.3.7. Potential Utility for Biomarkers

In total, seven biomarkers were found unique to each species (Table 10). Beyond their applicability in LC/MS, they could also assist other established techniques. Real-time PCR and EIA methods are rapid and sensitive approaches for biomarker detection, but neither are capable of biomarker discovery. The protein markers found in this study could be sequenced and then that sequence information be utilized to reverse engineer novel real-time PCR primers.[37, 68, 69] These marker proteins were consistently found during both the known and unknown portion of this study indicating that not only are their gene sequences unique to their respective species, but that these are genes that are

actually expressing the proteins they encode for. Additionally, these markers could be fraction collected as they elute off of the column and have antibodies made for them. These proteins could then aid EIA techniques by increasing the number of analytes that need to be detected in order to be considered a positive result. Since many EIA methods suffer from non-specific binding, having more analytes required for confirmation will likely reduce false positives. One manner in which EIA approaches could detect multiple analytes in a high throughput fashion would be in the form of protein microarrays, which have shown promise as a novel approach for the analysis of bacteria.[70, 102]

5.4. Conclusions

A multifaceted LC/MS method combined with novel data analysis software was examined for its efficacy to characterize and identify pathogenic *Clostridium* species. The capabilities of this method are three-fold: discovering intact protein biomarkers, speciation and strain level typing. Using two newly developed MS amenable lysis protocols, reproducible biomarkers were discovered that enabled speciation and strain level characterization. Discrimination at the species and strain levels may help guide treatment, prophylaxis and the implementation of proper control measures during outbreaks or exposure events. This method is highly specific and was used to distinguish strains that were 100% similar by PFGE analysis. The biomarkers discovered in this study were successfully implemented during the identification of unknowns in a blind study and could be utilized as new targets for detection by either EIA or real-time PCR approaches.

CHAPTER 6 Ultra Performance Liquid Chromatography/Mass Spectrometry of Intact Proteins

6.1. Introduction

6.1.1. Applications of Protein Chromatography

The study of proteins is an integral part of biochemistry. Efforts to understand protein origin, structure and function require efficient means of extraction, purification and characterization – all of which employ some type of separation. For this reason, there exists a correlation between advancements in separation techniques (e.g. liquid chromatography) and advancements in biochemistry.[103] Intact protein separations have utility in numerous applications. Separation and characterization of therapeutic proteins is of increasing concern in the pharmaceutical industry as biopharmaceuticals (e.g. recombinant proteins and monoclonal antibodies) now account for 25% of the new molecular entities entering the market.[104] Manufacturers of foodstuffs need to determine the protein content of their products as well as monitor the quality of protein supplements. Moreover, public health, clinical and forensic laboratories require capabilities to monitor food allergens (e.g. peanut, soy and milk), disease states (hemoglobinopathies) and various protein toxins from sources such as plants (e.g. ricin and abrin), venom (e.g. helothermine) and bacteria (e.g. botulinum toxin).

A rapidly emerging area for intact protein separations is top-down proteomic analysis. This approach has some advantages over the more popular bottom-up method

where proteins are digested and analyzed. The complexity of the peptide mixture makes the bottom up approaches (e.g. shotgun proteomics) often less discerning of protein isoforms which are more readily observed using a top-down approach.[105] Protein isoforms may provide insight into cell regulation and disease states and can be caused by coding polymorphisms, alternative splicing and posttranslational modification.[106] Other advantages of the top-down approach include greater sequence coverage of individual proteins, and that intact molecular weight information is obtained.[105, 107] Knowing the intact mass allows immediate detection of modified proteins during comparative proteomic investigations.

Top-down approaches typically isolate unique proteins either by two-dimensional (2D) gel electrophoresis (GE) or with liquid chromatography/mass spectrometry (LC/MS). Drawbacks to using GE compared to LC/MS include greater difficulty in analyzing highly acidic, basic and hydrophobic proteins, decreased mass resolution and accuracy, poorer reproducibility, and being more labor intensive. Advantages of using reversed phase LC/MS are improved quantitation, ease of automation, samples are enriched on column and isolated protein fractions are amenable to further MS analysis without sample preparation.[99, 108] Reversed phase LC/MS can be considered a 2D approach since distinctions are made according to hydrophobicity in the first dimension and mass in the second.[109] Using split-flow LC/MS, unique proteins can simultaneously be observed and fraction collected into a well plate. Proteins in the collected fractions can then be infused and investigated by mass spectrometry for structural information or digested in-well[110] and then analyzed. Despite the advantages of the top-down approach,

it is less often used due to difficulties during both mass analysis and in the chromatographic separation of intact proteins. Improvements in both areas of analysis are needed to enhance the widespread use of the top-down approach.[105, 107]

6.1.2. Challenges with Intact Protein Chromatography

Relative to small molecules and peptides, the reversed phase chromatography of proteins has been problematic - displaying carryover, multiple peak formation (splitting) and broad, misshapen peaks. These difficulties arise in part due to slow intrapore diffusion times, the presence of unresolved structural microheterogeneity and conformational isomers, and secondary interactions with the stationary phase.[10] Broad peaks have resulted in poorer resolution and therefore longer gradients resulting in increased run times and decreased throughput. One potential method of improving intact protein chromatography is the use of smaller particles.[111] Reducing the particle diameter (e.g. below 2 μm) can afford enhanced resolution and shorter run times. Improved resolution is predicted from the lower plate height minimum in Van Deemter plots of sub-2 μm particles and faster separations are a result of the elongated linear velocity region at that minimum.[8, 9, 112]

6.1.3. Porous vs. Nonporous Particles

Early work using sub-2 μm particles for intact protein separations involved nonporous particles[113] because they are mechanically strong and relatively easy to manufacture. Jorgenson et al. used 1.5 μm nonporous particles and achieved peak widths and run times that were comparable to perfusion chromatography[114] but with increased

sensitivity.[113] However, nonporous particles have decreased surface area compared to porous material. This reduction in surface area results in decreased loading capacity causing mass overload at lower concentrations.[9] When comparing 1 μm nonporous particles to 1 μm particles with 10 nm pores, porous particles possess approximately 100 times more surface area resulting in a 22-fold increase in loading capacity.[115] Increased sample loading capacity is important for preparative applications, for quantifying impurities or degradation products and when trying to detect low abundance proteins as more will need to be injected in order to detect them. Additionally, the decreased surface area of nonporous columns typically results in lower retentivity[112, 116] which may hamper the separation of difficult to retain highly hydrophilic analytes.

6.1.4. Chromatography at Ultra High Pressure

While the above drawbacks of nonporous particles may limit certain applications, some researchers prefer nonporous material citing relatively reduced carryover.[117] To address this issue, Eschelbach and Jorgenson employed an in-house ultra high pressure liquid chromatography (UHPLC) system and columns with 1.5 μm particles and examined the effect of increased pressure on carryover. Using a pressure range from 160 - 1600 bar, it was found that as the pressure was increased, carryover was diminished – an important finding for intact protein chromatography.[41] Other possible advantages such as increased throughput, resolution and sensitivity were not described.

Since the advent of commercial ultra performance liquid chromatography (UPLC) systems, much interest has been generated regarding the prospect of achieving increased resolution and throughput. The use of commercially available columns and systems is

vital for laboratories lacking the expertise to manufacture them in-house. Typically, UPLC refers to applications using porous sub-2 μm particles, at high linear velocities and pressures of 400 – 1000 bar on commercial systems. UHPLC is similar, but is typically utilized at > 1000 bar with custom built systems in academic or research laboratories.[9]

6.1.5. Specific Aims

Many small molecule applications utilizing commercial UPLC systems have been reported[118-123] and have described benefits such as increased throughput, resolution and sensitivity. The aim of this manuscript was to explore whether these same successes could be observed with the much more problematic and difficult to chromatograph intact proteins. Using ten protein standards ranging in mass from 6 - 66 kDa, and mobile phases amenable to electrospray ionization, efforts were made to examine the performance of UPLC relative to high performance liquid chromatography (HPLC) using columns differing only in particle size. The original method employing typical conditions was demonstrated and then parameters such as temperature, organic solvent and particle diameter were optimized. The optimized method was then applied to a more complex cell lysate to determine the overall efficacy of the method.

6.2. Experimental

6.2.1. Materials

HPLC grade solvents (acetonitrile, methanol, formic acid and trifluoroacetic acid) were purchased from Fisher Scientific (Fairlawn, NJ) and 2-propanol was purchased from Honeywell Burdick and Jackson (Morristown, NJ). Organic mobile phases were filtered using 0.2 μm PTFE filters (Pall Corporation, Ann Arbor, MI). The water utilized for

LC/MS analysis was purified in-house to yield organic-free 18.3 MΩ × cm using an E-pure purification system (Barnstead International, Dubuque, IA). The sterile water used during bacteria preparation was autoclaved and purified with a RiOs 5 Water Purification System (Milipore, Billerica, MA).

6.2.2. Protein Standards Preparation

Bovine serum albumin (BSA), horse heart myoglobin, cytochrome C, chymotrypsinogen A and ovalbumin were purchased from MP Biomedicals (Solon, OH). Trypsin inhibitor (soybean), insulin (bovine pancreas), ribonuclease A (RNase A, bovine pancreas), α-lactalbumin (bovine milk) and lysozyme (HEWL) were purchased from Sigma-Aldrich (St. Louis, MO). Individual stock solutions of each protein were prepared at 1 mg/mL in 20 mM NH_4OAc. An equalmolar mixture of the ten proteins was prepared at 1.5 μM in 50% methanol and 0.1% formic acid.

6.2.3. Cell Culture and Lysis

Escherichia coli serotype O157:H7 isolates were obtained in-house. The cells were grown for 24 hours on trypticase soy agar plates containing 5% sheep's blood at 37°C. Cells were then removed from the plate and placed in a test tube containing 1 mL of water until the turbidity reading reached 1.0 using a MicroScan Turbidity Meter (Dade Behring, West Sacramento, CA). A 500 μL aliquot of this suspension was placed in a 1.5 mL protein LoBind tube (Eppendorf, Westbury, NY) and washed three times with 500 μL of sterile water followed by centrifugation (6000 × g at room temperature for 5 minutes) to remove residual media. Finally, the cells were resuspended in 150 μL of the lysis solution (1:1 H_2O: acetonitrile, 0.1% (v/v) trifluoroacetic acid). After chemical

lysis, the sample was again centrifuged (4100 × g for 4 minutes) at room temperature. The supernatant was filtered using a 0.22 μm, 4 mm PVDF low protein binding GV filter (Milipore, Billerica, MA) and 65 μL of the filtered supernatant was transferred to an autosampler vial for analysis.

6.2.4. Chromatography and Mass Spectrometry

Intact proteins were separated by reversed phase chromatography using an Acquity UPLC (Waters, Milford, MA). The HPLC column was the narrowbore X-bridge C18 BEH 300 Å 2.1 × 150 mm, 3.5 μm (Waters). The use of these porous BEH (Ethyl-Bridged Hybrid) particles has been described.[115, 124] For comparison purposes, the UPLC column was identical (e.g. pore size, ligand density, carbon load) except in particle size (1.7 μm). To achieve maximum column life time and to ensure a robust method, manufacturer recommendations were followed by maintaining backpressures near or below 690 bar (10 kpsi) and column temperatures were kept below 70°C.[125] The gradient for HPLC was linear: 5 – 55% B in 60 minutes for a slope of 0.83%B/min. All mobile phases contained 1% (v/v) formic acid. The autosampler was maintained at 15°C. For the 1.5 μM protein standard mixture and cell lysates, injection volumes of 2 μL (3 pmol) and 20 μL respectively were utilized.

The LC was directly interfaced to a Q-TOF Premier (Waters) mass spectrometer utilizing positive ion electrospray ionization. The capillary and cone voltages were 3900 and 40 V respectively. The desolvation gas (900 L/h) was heated to 500°C and the source temperature was 115°C. Ions were monitored over an *m/z* range of 620 - 2450 Da and resolved in single reflectron (V) mode. An acquisition rate of two spectra/sec was

utilized. Other parameters employed in the MS method were optimized for sensitivity and resolution using BSA. Data was collected using MassLynx software version 4.1 (Waters).

6.2.5. LC/MS Data Analysis

LC/MS data was processed using two software packages: Protrawler6 (Bioanalyte, Portland, ME) and MS Manager (Advanced Chemistry Development Laboratories, Ontario, Canada). Protrawler6 provided automated deconvolution of multiply charged ions by first dividing the full-scan data from the chromatogram into time intervals (30 seconds) and summing the data from each interval. Sequential deconvolution events were then performed to obtain neutral masses of the proteins that eluted during each interval. A text file containing neutral masses, intensities and retention times was then created summarizing the results for each chromatogram.[6] The masses and intensities were utilized to create a single mass spectrum representing all of the proteins observed in the chromatogram using MS Manager. MS Manager was also used for spectral mirroring. For ease of comparison, spectral mirroring allows spectra to be mirrored along the abscissa.

6.3. Results and discussion

6.3.1. Original Method

Figure 17 depicts a chromatogram of the ten protein standards using the original method. The use of protein standards is beneficial for three reasons. First, they are inexpensive and easy to obtain; therefore, more convenient for method development purposes than expensive or difficult to procure analytes or lysates. Second, protein

standards have been implemented by several researchers in the past making it easy to compare this work with previous efforts in protein chromatography.[40, 126, 127] Finally, protein standards are useful for quality control standards and in this laboratory, are run at the beginning and end of each day's work to ensure the LC, column and MS are working properly. Examination of Figure 17 shows over 20 minutes of dead time prior to elution of RNase A and 15 minutes after the elution of ovalbumin. It should be noted that the focus was to develop a method applicable to cell lysate analysis and during this type of analysis several proteins will elute during the aforementioned regions.

The original method incorporated an HPLC column with a C18 alkyl chain ligand and 3.5 μm diameter particles that was operated at a moderate column temperature (35°C). This column is consistent with small molecule applications; however, the column utilized in this study had a larger pore diameter (300 Å) to account for the larger size of the analytes. In addition, the flow rate (75 μL/min) was lower than what is typically implemented in small molecule applications for a column having a 2.1 mm internal diameter. Given that the optimal linear velocity is proportional to the diffusion coefficient of the analyte[128], the reduction in flow rate accounts for the lower diffusivity of macromolecules. Acetonitrile, the most commonly used organic modifier for intact protein separations[10], was initially used. A long, shallow gradient (5 - 55% B over 60 minutes) was utilized to maximize resolution. However, despite the long gradient, the resolution was insufficient, particularly for proteins 5 - 8 (BSA - trypsin inhibitor) the most problematic region of the chromatogram.

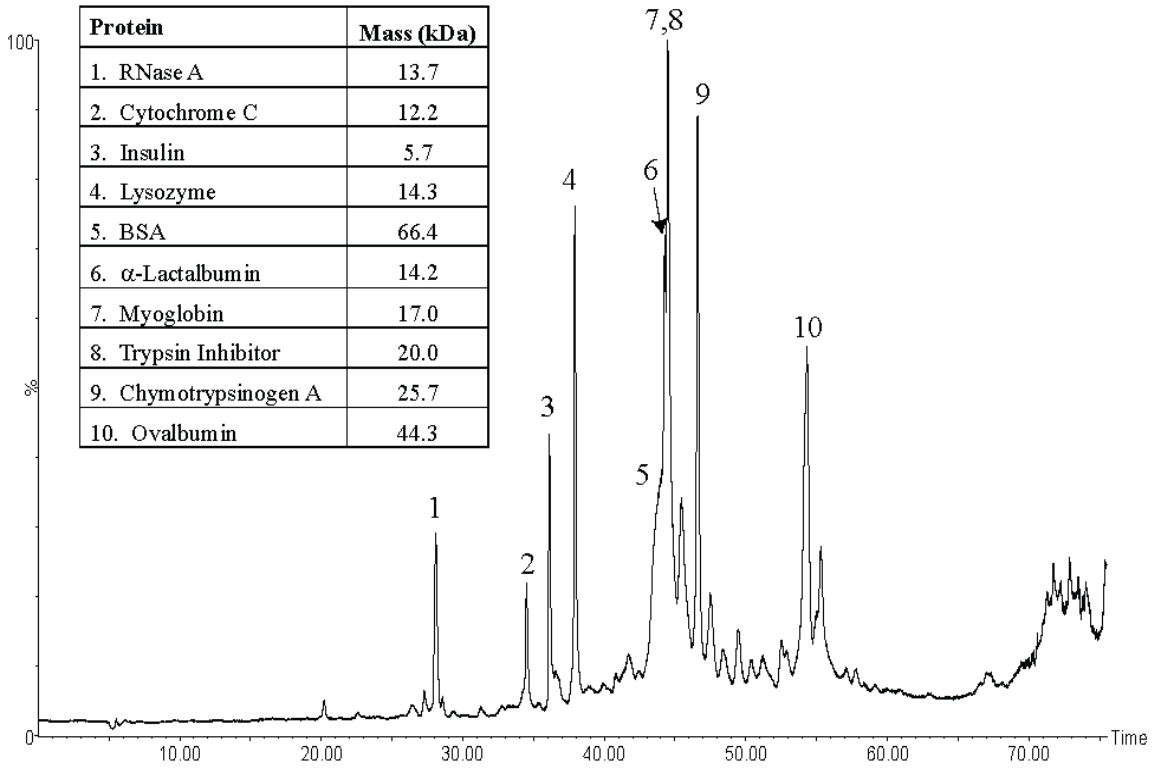


Figure 17. Results obtained using typically employed HPLC conditions (column temperature = 35°C and acetonitrile was the organic modifier used). Poor resolution and distorted peak shape was observed for the ten protein standards used in this study. These ten standards and their masses are provided in the inset table. The numbers assigned to each protein were kept regardless of retention order.

6.3.2. The Effect of Temperature

The first parameter investigated was temperature. The use of elevated column temperature, as in high temperature liquid chromatography (HTLC)[129] has been demonstrated to improve intact protein separations.[130, 131] Increasing the temperature may modify the properties of the column surface and alter protein structure thus affecting analyte retention.[10] Secondly, analyte sorption kinetics increase with temperature which is important for slowly diffusing proteins. Increasing diffusivity both into and out of the pores minimizes band broadening; consequently, column efficiency has been shown to

increase with temperature.[132] Finally, increased temperature reduces mobile phase viscosity causing a concomitant decrease in back pressure. Decreased back pressure allows scaling to higher flow rates resulting in reduced run times in UPLC experiments.

Three temperatures at constant flow rate (75 μL/min) were utilized for comparison (Figure 18). An increase in sorption kinetics was indicated by the slight decrease in retention time for proteins such as RNase A (#1), which eluted at 28.09 minutes at 35°C and 25.97 minutes at 65°C. Selectivity differences were also apparent. As temperature increased, the retention order of two sets of proteins – α-lactalbumin (#6) and myoglobin (#7), and insulin (#3) and lysozyme (#4), were reversed. While lysozyme clearly lost retention as temperature increased, insulin was less susceptible to this change and in fact showed a slight increase in retention (36.12 minutes at 35°C and 36.44 minutes at 65°C). This property of insulin was also observed by Szabelski et al. where the unfolding kinetics of insulin may have increased with temperature, allowing a greater area of contact with the stationary phase.[133] Thus, proteins such as RNase A must either be more resistant to unfolding at these temperatures allowing increased sorption kinetics to dominate, or upon unfolding, more hydrophilic regions are exposed which shield or inhibit previous binding interactions. At 65°C four peaks were observed in the most problematic region of the chromatogram, proteins 5 – 8. For this reason, the optimal temperature was chosen to be 65°C. To maximize column lifetime and ensure a robust method, higher temperatures (> 65°C) were not attempted.

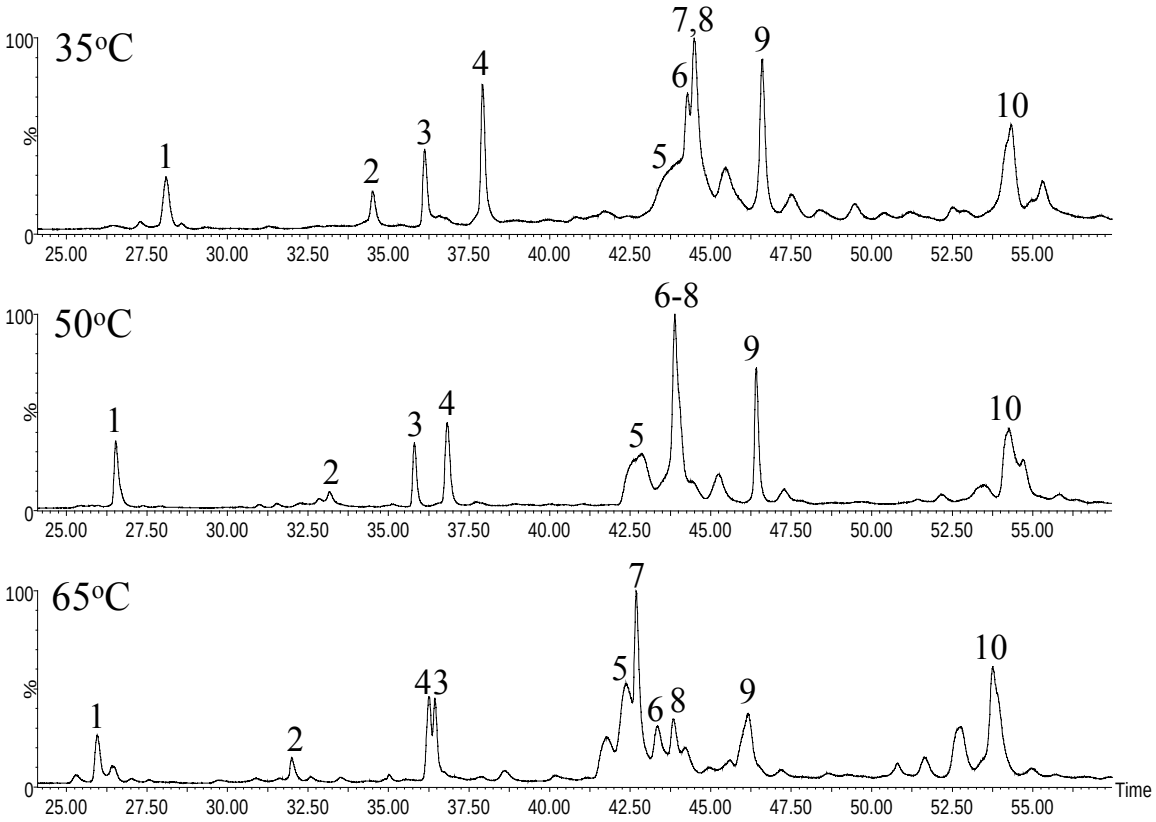


Figure 18. Results from the HPLC temperature study using acetonitrile as the organic modifier. As the temperature was increased from 35°C to 65°C the resolution for proteins 5-8 improved.

6.3.3. The Role of Organic Modifier - HPLC

In addition to temperature, two different organic modifiers, acetonitrile (ACN) and isopropanol (IPA) were examined. These two solvents exhibit different properties which may alter their displacement performance resulting in selectivity differences. Acetonitrile is aprotic, more polar, less viscous and has a lower eluotropic strength than isopropanol (Table 11). ACN and propyl alcohols have different effects on the conformation of proteins which can subsequently cause distinct retention behaviors.[134] IPA is less denaturing than ACN and has shown improved solubility of hydrophobic proteins.[10] Dillon et al., while characterizing antibodies, found using stronger solvents

(e.g. propanol) improved resolution and recovery by reducing secondary column interactions. Additionally, an increased electrospray response was also observed when using propanols relative to ACN.[135, 136]

Table 11. Different solvent properties of acetonitrile and isopropanol. The combination of these properties result in selectivity differences for the analytes in this study.

	ACN	IPA
H^+	Aprotic	Protic
ε_0	37	18
Viscosity (25 °C)	0.34	1.90
Eluotropic strength	3.1	8.3

The stronger eluting power of IPA is evident from Figure 19. Using the same column, gradient, flow rate and the optimal temperature of 65°C, ovalbumin (#10) elutes 14 minutes earlier with IPA, with less peak splitting also being observed. This could result in shorter gradients and less solvent consumption due to a lower %B being required for elution. The resolution was also improved for proteins 5 – 8 (BSA – trypsin inhibitor) with IPA while proteins 2 (cytochrome C) and 3 (insulin) were better resolved with ACN. Often, when attempting to better resolve a pair of proteins, changes will be made to the gradient or perhaps a different column will be used. However, the selectivity differences from temperature and organic solvent may be potentially more useful or convenient and in the case of using a different column, less expensive. To further explore the performance of these two solvents, their comparison was continued throughout the HPLC vs. UPLC comparison.

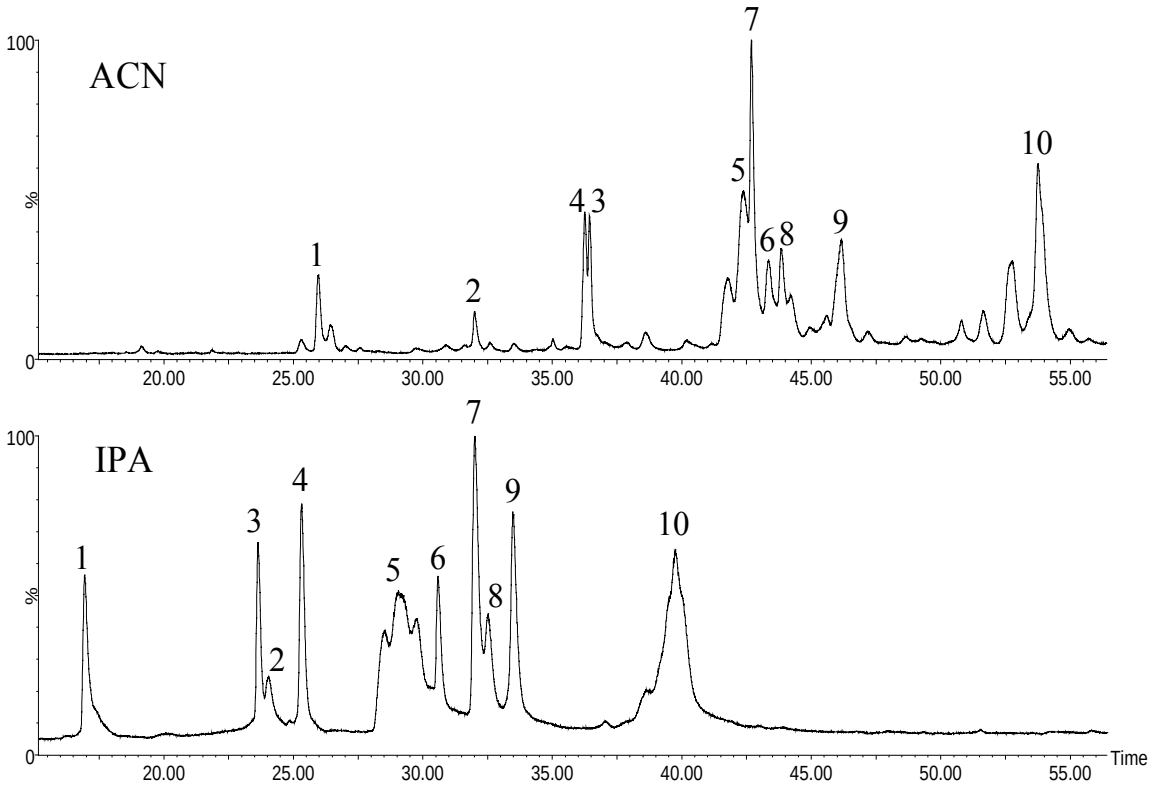


Figure 19. Using the same column, column temperature, flow rate and gradient, the HPLC performance for ACN (top) was compared to IPA (bottom). Due to its stronger eluotropic strength, all ten proteins eluted in less time when IPA was used as the organic modifier.

6.3.4. Ultra Performance Liquid Chromatography

In addition to the potential benefits of increased resolution, throughput and sensitivity which have been observed during small molecule applications, certain aspects of using UPLC relative to HPLC may prove particularly attractive for intact protein separations. As described above, protein carryover is reduced by operating at higher back pressures.[41] Additionally, smaller particles have less resistance to mass transfer which is beneficial for slowly diffusing macromolecules.[132] Furthermore, shorter runtimes mean less time on column for each analyte. As time on column decreases, on-column limited acid hydrolysis is minimized[136] and recovery of hydrophobic proteins is

improved.[10] Shorter runtimes also decrease the amount of time a batch of samples spends in the autosampler prior to analysis. Reducing this time may prevent sample degradation if the autosampler is insufficiently chilled and decrease the extent of unwanted proteolytic digestion if proteases are present in the sample. For the UPLC vs. HPLC comparison, the effect of particle size was examined first, followed by examining the response of the separation to using scaled gradients. These two factors were inspected for both acetonitrile and isopropanol.

Figures 20a and 20b depict the effect of particle size at constant temperature (65^{o}C) and flow rate (75 μL/min) using acetonitrile. Although these columns were identical with the exception of particle size, some selectivity differences were observed with proteins 3 (insulin) and 4 (lysozyme). With the smaller particle size, the elution order was reversed for these proteins and the resolution enhanced. Less resistance to mass transfer is likely demonstrated by the slightly lower retention times and the sharper peak shape for proteins such as chymotrypsinogen A (#9). The same comparison using isopropanol is shown in Figures 21a and b. Here, the decreased resistance to mass transfer is evident again from lowered retention times (e.g. chymotrypsinogen A (#9)) and peak shape enhancement for BSA (#5). Also a selectivity difference occurred for proteins 2 (cytochrome C) and 3 (insulin). However, with IPA, the overall increase in resolution is more clear - particularly between proteins 5 (BSA) and 6 (α-lactalbumin); and 7 (myoglobin) and 8 (trypsin inhibitor). Using extracted ion chromatograms to better measure peak widths, the HPLC resolution between 7 and 8 (myoglobin and trypsin inhibitor) was 0.74 compared to 1.18 for UPLC.

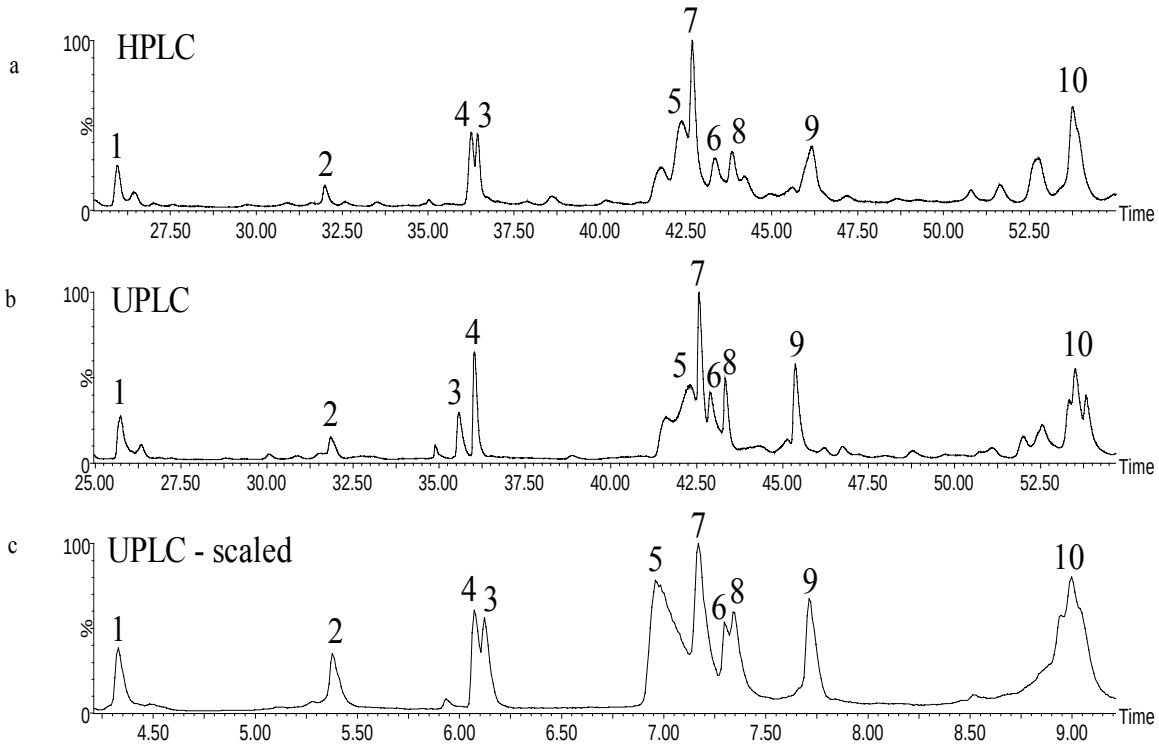


Figure 20a-c. HPLC vs. UPLC comparison using acetonitrile. Figures 20a and 20b demonstrate a modification in selectivity and enhancement in resolution from decreasing the particle size from 3.5 μm (Figure 20a) to 1.7 μm (Figure 20b). All other parameters were kept constant. Despite the six-fold increase in throughput, this increase in resolution was not maintained upon scaling of the gradient from a 60 minute gradient at 75 μL/min to a 10 minute gradient at 450 μL/min (Figure 20c).

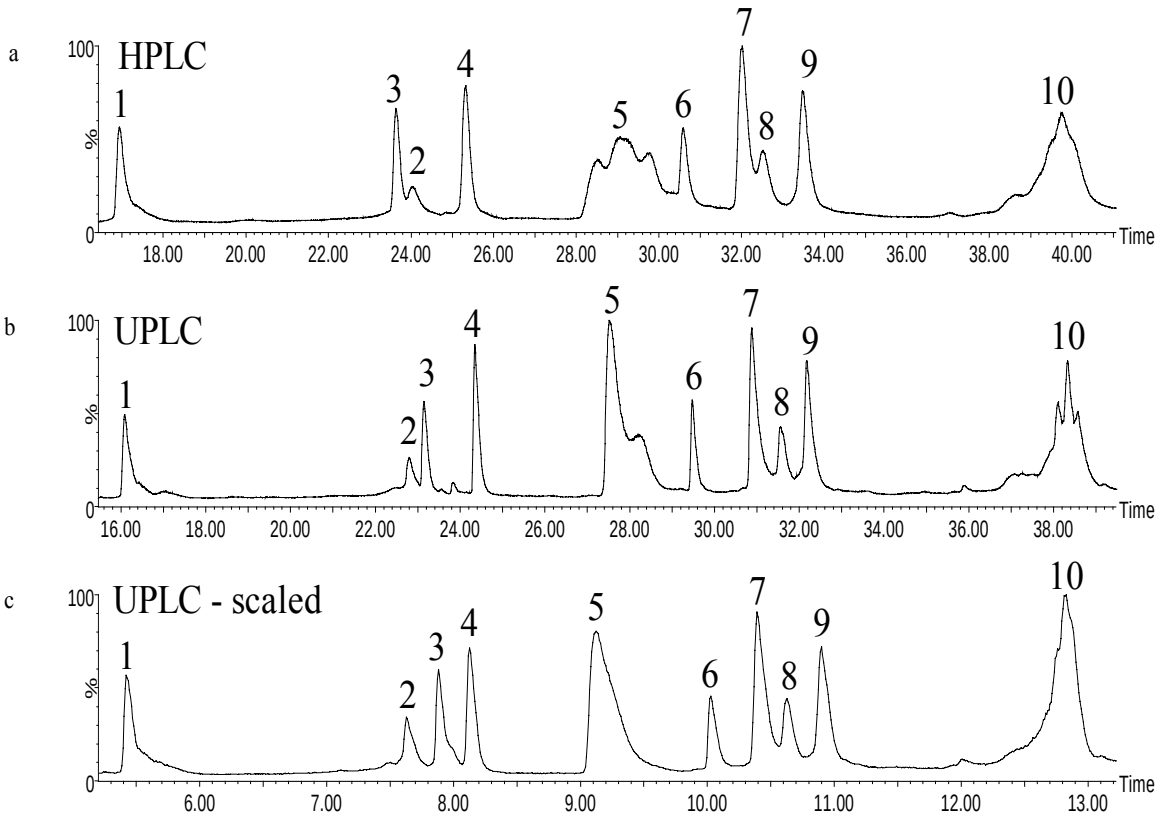


Figure 21a-c. HPLC vs. UPLC comparison using isopropanol. Figures 21a and 21b demonstrate a modification in selectivity and enhancement in resolution from decreasing the particle size from 3.5 μm (Figure 21a) to 1.7 μm (Figure 21b). All other parameters were kept constant. Additionally the peak shape of BSA (#5) was greatly improved when smaller particles were utilized. Upon scaling of the gradient (Figure 21c) from a 60 minute gradient at 75 μL/min to a 20 minute gradient at 225 μL/min, a three-fold increase in throughput was observed with no loss in resolution.

As expected, the decrease in particle size yielded improved resolution. To further explore the benefits of smaller particles, the flow rate was increased and the gradient was scaled accordingly. Figures 20b and 20c depict a six-fold increase in flow rate and decrease in gradient time from 0 – 55 %B in 60 minutes at 75 μL/min to 0 – 55 %B in 10 minutes at 450 μL/min. Scaling the gradient during UPLC separations first involves increasing the flow rate, and then decreasing the time for each gradient segment (e.g. initial separation, column cleaning and re-equilibration) by that same factor. For

example, in Figure 20c, the flow rate was increased six-fold therefore, the time for each segment in the gradient was decreased six-fold. Scaling in this manner holds the number of column volumes per gradient segment constant.

The scaling of the gradient stopped at 450 μL/min for ACN because it approached the limit of the mass spectrometer to effectively desolvate. Other problems associated with using higher flow rates may include greater solvent consumption and more dilute fractions during fraction collection. While the elution time for ovalbumin (#10) was 1/6 of its original value, the resolution was not maintained during the scale up process (proteins 3 and 4, insulin and lysozyme respectively). This is in contrast to IPA in which the resolution was maintained and even increased for proteins 2 (cytochrome C) and 3 (insulin) (Figure 21b and 21c). In addition, the elution time of ovalbumin (#10) was approximately 1/3 of its original value. The gradient could only be scaled by 1/3 because IPA is much more viscous than ACN. Near the end of the gradient at 225 μL/min, the back pressure began to exceed 690 bar, the manufacturer's recommended limit for maximizing column life time, with maximum pressures being approximately 745 bar (10,800 psi).

6.3.5. The Role of Organic Modifier - UPLC

A final comparison of the two solvents using the scaled UPLC gradients is shown in Figure 22. In terms of the elution time of ovalbumin (#10), using ACN provides increased throughput. However IPA, which elutes ovalbumin four minutes later than with ACN exhibits a pronounced increase in resolution. As mentioned previously, these standards have been commonly used by other researchers, and some have found the

separation of cytochrome C and insulin[137] and BSA and myoglobin[41] to be challenging. However, with the IPA method described here, cytochrome C (#2) and insulin (#3) are separated and not only are BSA (#5) and myoglobin (#7) baseline resolved but α-lactalbumin (#6) is baseline resolved in between them. Increased resolution allows for reduced ion suppression, cleaner mass spectra, purer fractions to be collected and simplifies chromatographic comparison and peak area determination. For these reasons, the enhanced resolution with IPA was deemed more valuable than the improvement in throughput observed with ACN.

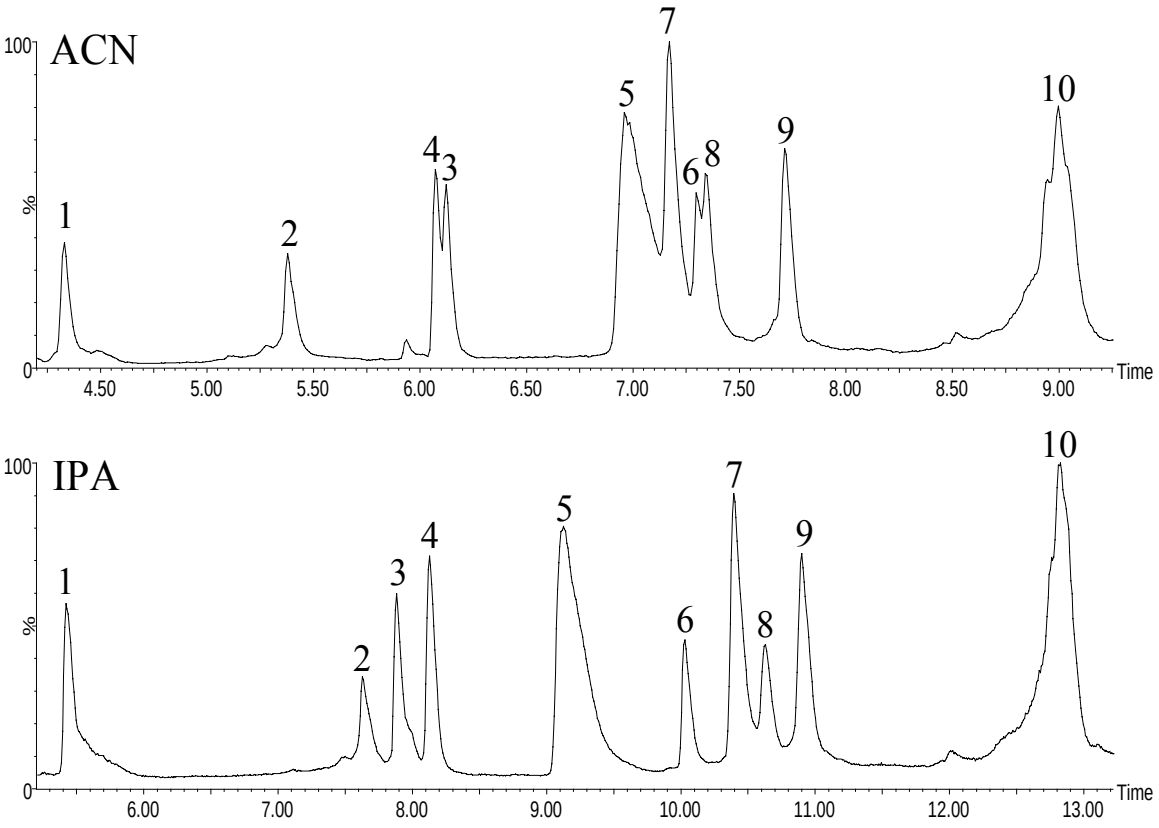


Figure 22. UPLC solvent comparison using the scaled gradients for each solvent but the same UPLC column and column temperature of 65°C. While the elution time of ovalbumin (#10) is four minutes later, the resolution with IPA (bottom) is significantly improved relative to ACN (top).

6.3.6. Original vs. Optimized Method

Figure 23 compares the original and optimized methods for the protein standards. The original conditions were typical, utilizing a moderate column temperature of 35°C, acetonitrile as the organic solvent, 3.5 μm particles and a shallow gradient of 5 – 55 %B in 60 minutes at a flow rate of 75 μL/min. After systematic optimization of the method, enhanced resolution was achieved and the elution time of ovalbumin (#10) was 1/4 of its original value. In fact, the entire method was completed prior to the elution of the first compound in the original method. Additionally, a significant decrease in peak width was observed. The baseline peak width of chymotrypsinogen A (#9) was approximately 45 seconds in the original method compared to 15 seconds after optimization. In addition to the 2/3 reduction of the peak width, the gradient time for the original method (60 minutes) was also reduced by 2/3 compared to the optimized method (20 minutes). Likewise, the time required for column cleaning and re-equilibration was equally abridged from fifteen minutes to five minutes for a total runtime reduction from approximately 75 minutes to 25 minutes. Moreover, since IPA is a much stronger solvent than ACN, the gradient and therefore the runtime could be further reduced. Any protein eluting at 55 %B with ACN would require a much smaller %B when using IPA.

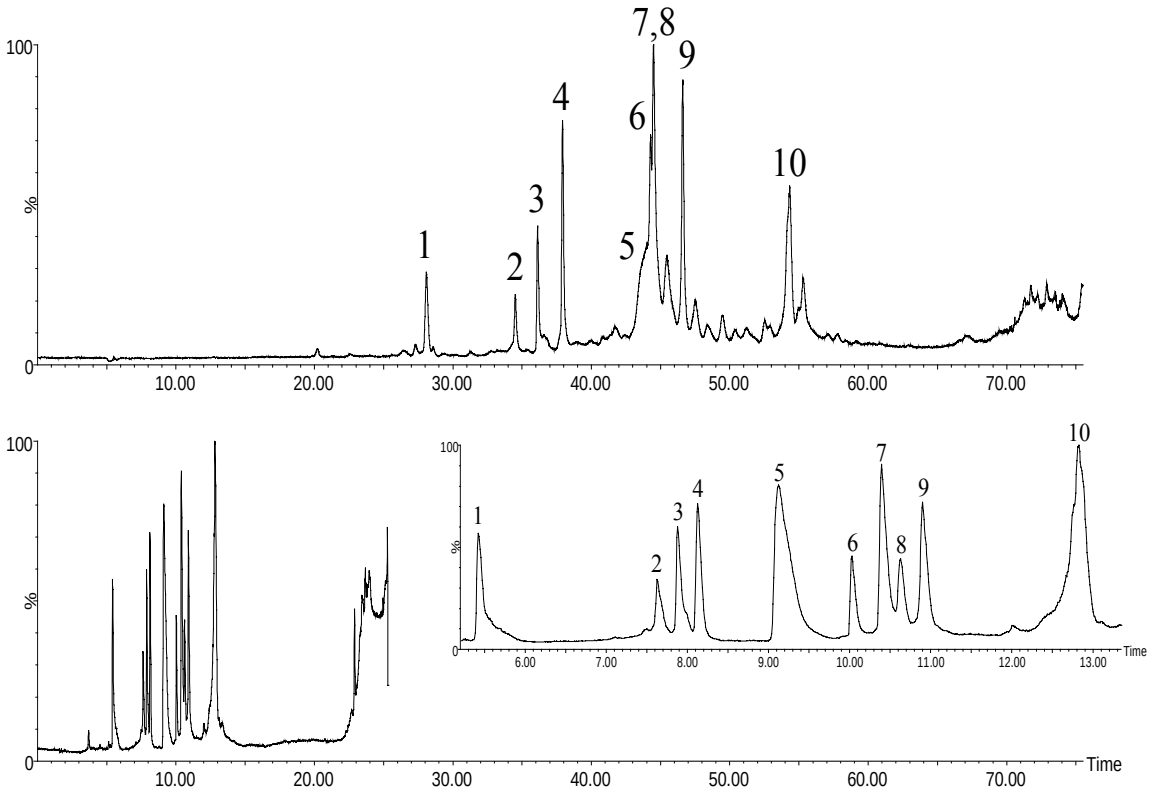


Figure 23. Method optimization before and after. The optimized method (column temperature = 65°C, IPA as organic modifier, 1.7 μm particle size and 20 min gradient) is completed prior to the elution of the first compound in the original method (column temperature = 35°C, ACN as organic modifier, 3.5 μm particle size and 60 min gradient). The inset more clearly reveals the resolution advantage afforded by the optimized method.

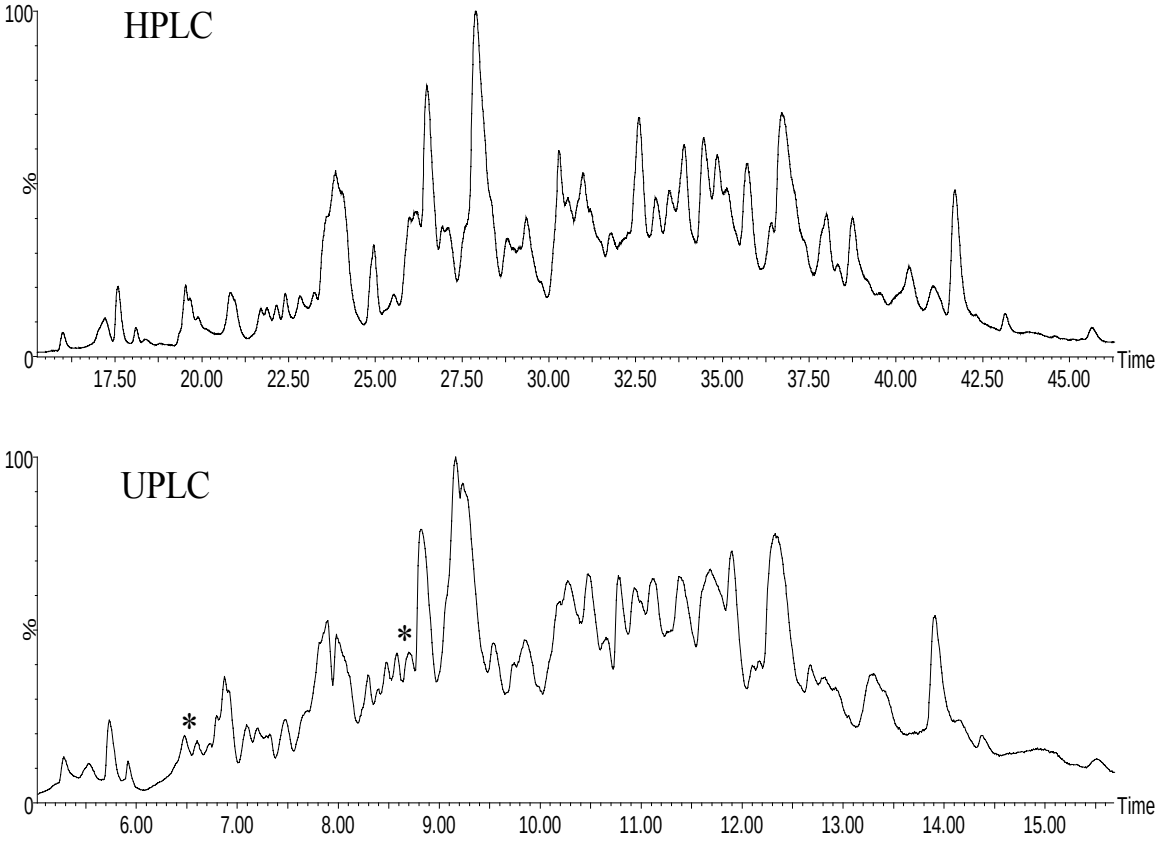
6.3.7. Sensitivity Enhancement with UPLC

Since the peak widths were reduced when the gradient was scaled, it was likely that there would be a concomitant sensitivity increase. The change in sensitivity was calculated by taking the individual peak heights from the optimized method and dividing them by the corresponding peak height in the original method. However, only a small increase in sensitivity was observed. One possible explanation was that as the flow rate increased, the desolvation efficiency of the electrospray source decreased resulting in lowered MS signal. To test this, both the 60 minute and 20 minute gradients were

compared using the UPLC column and IPA as the organic modifier, but at the same flow rate of 225 µL/min for both gradients. During this comparison, the average increase in sensitivity was two-fold when using the scaled gradient. However, when the signal intensities from the 60 minute gradient at 225 µL/min and 75 µL/min were compared, the intensity was elevated with the lower flow rate. Thus, while the scaled gradient yielded narrower peaks and therefore increased sensitivity, this increase in sensitivity was compromised by poorer desolvation at higher flow rates. Consequently, an increase in overall sensitivity of (1.4X) was observed from the HPLC vs. UPLC comparison.

6.3.8. Application of the Optimized Method

To explore the benefits of the optimized method on a more complex sample, an *E. coli* O157:H7 cell lysate was analyzed. Using the optimal temperature and organic solvent, a comparison was made between HPLC and UPLC to determine what impact the increased resolution would have on mass spectral data quality. The results of this comparison are shown in Figure 24. The enhanced chromatographic resolution which was obvious when only ten proteins were analyzed is difficult to discern due to the complexity of the lysate containing > 100 proteins. Two instances of improved resolution with UPLC are denoted by asterisks in Figure 24a. Despite careful optimization of the method, significant co-elution was evident. The representative mass spectra for each technique were very similar (Figure 24b) and even upon close inspection, no significant differences were detected. Both spectra exhibited a biomarker unique to the O157:H7 serotype of *E. coli* weighing 18,996 Da[138], which eluted at 38.8 minutes with HPLC and 13.3 minutes with UPLC.



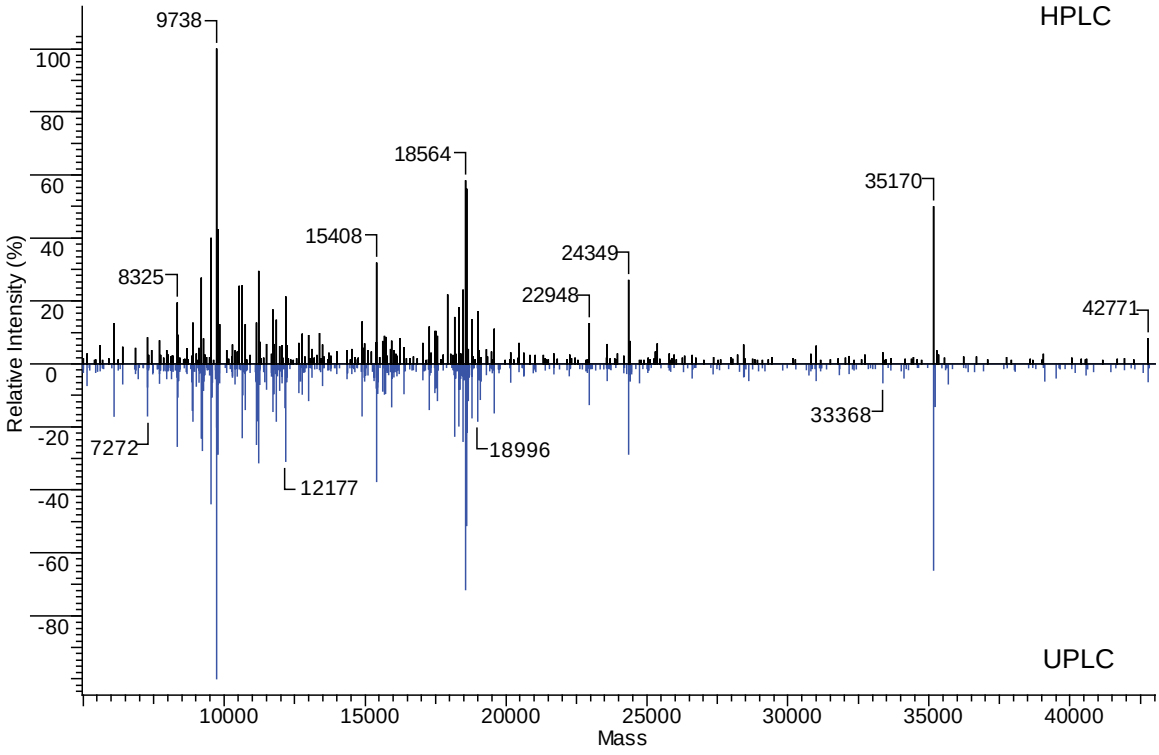


Figure 24. A region of the chromatogram (Figure 24a -top) is used to compare the chromatographic results between the two lysate analyses. Other proteins eluted during the times not shown. * denotes instances of improved resolution with UPLC. Figure 24b (bottom). Mass spectral comparison of the lysate data. No significant differences were observed between the two methods however, the data collection and analysis time with UPLC was 1/3 that of HPLC.

Due to the overwhelming complexity of the cell lysate, the increase in resolution afforded by UPLC was not enough to allow more proteins to be observed, as both mass spectra contained approximately 150 proteins. To this end, further efforts to reduce the complexity of the lysate could be made. For instance, differential solubilization[139] or sub-cellular proteomics[140] extraction techniques could be applied. This would yield multiple extracts based on differential protein solubility or different components of the cell being specifically lysed and extracted. These multiple extracts could then be separately analyzed using the optimized method. Additionally, UPLC could be

implemented to enhance 2D separations where in the first or second dimension an orthogonal separation such as capillary isoelectric focusing[141], chromatofocusing[142] or size exclusion chromatography[143] was employed. Relative to 1D separation, 2D separations afford much higher peak capacities, and the extension of this technology to *E. coli* lysates has been demonstrated.[144]

While more proteins were not observed using UPLC, the run time was reduced by 2/3, at no cost in data quality. In addition, not only was the data collection time reduced, but the data analysis time was also reduced using the UPLC method. For the 75 minute HPLC chromatogram, the time required to perform automated charge state deconvolution with Protrawler6 was approximately 30 minutes. This is in contrast to the 25 minute UPLC chromatogram which required only 10 minutes per chromatogram. The step of charge state deconvolution is crucial when using electrospray ionization to analyze intact proteins from complex cell lysate samples and streamlines the data interpretation process. Reducing the deconvolution time minimized the gap between data collection and interpretation allowed data directed decisions to be made in less time.

6.4. Conclusions

Separation is an essential aspect of protein biochemistry and numerous applications for intact protein chromatography exist. Beginning with typically employed conditions, the parameters of temperature, organic modifier and particle size were optimized. Using a combination of high temperatures, strong solvents and small particles at very high pressure yielded optimal results, which included enhanced resolution, sensitivity and a three-fold increase in throughput. Increased throughput saves time and

money in industry applications and during forensic or public health analyses allows faster reporting of results to investigating authorities and hospitals. The optimized method was applied to a bacterial cell lysate and a reduction in both data collection and analysis time was observed with no compromise in the quality of the data. The utilization of porous particles, as in UPLC, provides a useful alternative to nonporous columns for researchers looking for greater sample loading capacities and larger retention factors.

CHAPTER 7 Conclusions and Future Work

The overall goal of this project was the development of a novel method to detect and characterize microbial pathogens. Of particular interest was developing a method that is complementary and at times advantageous compared to commonly employed assays. When used in conjunction with existing methods, the combined approach would yield increased confidence prior to reporting results to investigating authorities. Initial investigations involved the analysis of the Gram negative enteric pathogens *Escherichia coli* and *Shigella* species. Next, the method was extended to other types of bacteria (Gram positive and endospore forming) during the analysis of *Clostridium* species. Both projects discovered intact protein biomarkers for identification which were evaluated for their reproducibility and subsequently their utility during blind studies. The method was then optimized for throughput utilizing ultra performance liquid chromatography.

For the *E. coli* and *Shigella* species work, a simple chemical lysis method, analysis by LC/MS, followed by automated charge state deconvolution and spectral subtraction was employed to discover biomarkers. Reproducible intact protein markers that displayed unique retention times, masses and or relative intensities allowed the speciation of *Shigella*. This is significant because real-time PCR primers for this task are commercially unavailable. Serotype level distinctions were also observed. Three serotypes of enterohaemorrhagic *E. coli*, which typically display identical symptoms, were investigated. Because of the similar symptoms they produce, they can't be

distinguished by symptoms alone; therefore a molecular approach must be employed. Markers were discovered that easily distinguished *E. coli* O157:H7 from O126:H11 and O111:NM. Once this distinction was made, a protein weighing 11,779 Da which eluted at 26.2 minutes was utilized to distinguish the two non-O157:H7 serotypes from each other.

Beyond serotypic discrimination, the specificity of the method was displayed when two distinct isolates of *E. coli* O157:H7 were distinguished by a single protein. This level of specificity is required during epidemiological investigations of food or waterborne outbreaks and during forensic traceback efforts to pinpoint the source of exposure. Given that thousands of people are infected with *E. coli* O157:H7 each year in the United States, assays with strain level specificity are beneficial for establishing strain relatedness - needed to indicate whether a series of infections were random or connected. The makers discovered during this investigation were then employed to correctly identify thirteen unknown isolates during a blind study.

Currently, identification of unknowns by this method is achieved through association with previously examined (known) pathogens, and is therefore limited to the types of pathogens listed in Table 2. However, this does not limit the strain typing ability of the method as typically in outbreak situations where PFGE is relied upon, general information such as genus, species and serotype have already been determined before attempting to establish strain relatedness. Future work specific to the *E. coli* and *Shigella* project could comprise of extending the list of pathogens investigated to other species of *Shigella* and *Escherichia* and additional serotypes of *E. coli* known to cause shigellosis.

Examples include *S. dysenteriae*, *E. vulneris*[145] and *E. coli* O91:H21. Extension to these pathogens should be straightforward due to the successful proof of concept work described in Chapter 4. Initial expansions of the method should be focused on areas where established techniques are lacking e.g., species for which no commercial real-time PCR primers exist, serotypes with no commercial antibodies and strains not previously typed by PFGE.

In an effort to develop a method applicable to all types of bacteria, endospore forming Gram positive *Clostridium* species were investigated. Thirteen isolates of two species (five *C. difficile* and eight *C. perfringens*), for which no commercial real-time PCR primers exist, were used. Both of these pathogens are known causes of antibiotic associated diarrhea (AAD). Initially the lysis method utilized for Gram negative bacteria was implemented and found incapable of yielding sufficient protein recovery and lysis of *Clostridia*. Therefore, two lysis methods were developed that enabled sufficient lysis and protein extraction. Interestingly, each lysis method extracted two sets of proteins differing in both mass and retention time. Due to the structural integrity of endospores, neither method was able to lyse them, thus they were removed by filtration with a 0.22 μm PVDF filter designed to minimize protein loss during filtration.

After their efficacy for Gram positive bacteria was demonstrated, both lysis methods were then applied to the Gram negative *E. coli* and *Shigella* samples and yielded sufficient lysis and protein recovery. Therefore, the above lysis methods are equally suitable for Gram positive or Gram negative bacteria thus precluding the need for a Gram stain prior to the analysis of unknown bacteria. While the Gram stain is not a lengthy

process (approximately ten minutes) and yields useful information, when rapid results are needed, the elimination of any unnecessary step may be beneficial.

Not only did both lysis methods yield sufficient protein recovery, proteins from both lysis methods enabled the speciation of *Clostridium*. Their speciation allows for diagnosis of which species of *Clostridium* is the causative agent for AAD in hospital settings, and can direct future treatment and pretreatment as distinct infection control measures for each species have been recommended. As with the Gram negative bacteria, the markers found during this project were further challenged for their reproducibility and were then successfully implemented for the speciation of ten unknown isolates during a blind study.

A clear display of the specificity of the LC/MS approach described here occurred during the analysis of two different sets of *C. perfringens* isolates, which were indistinguishable by PFGE, the gold standard for strain differentiation. In particular, one set of two isolates, accession numbers: 06-0385 and 06-0387, were distinguished not only by differences observed in the mass spectrum, but also by differences in the chromatogram. These results indicated that LC/MS could yield further insight in to the history of an exposure than current PFGE protocols allow. For this reason, these results serve as basis for the recommendation that current PFGE protocols for the analysis of *C. perfringens* be amended to include different or additional restriction enzymes which may give more informative restriction patterns.

Specific to the *Clostridium* project, future work could include the analysis highly virulent, drug resistant strains of *C. difficile* that have been classified as 'superbugs'.

Since strain level discrimination of *C. difficile* was demonstrated in Chapter 5, the extension of the method to these strains of *C. difficile* would likely be successful. Markers that could distinguish drug resistant strains could direct the administration of antibiotics appropriate for treatment. Furthermore, if distinctions are observed that implicate highly virulent strains, their detection could serve to expedite remediation processes such as disinfecting hospitals and quarantining infected patients etc.

Since both projects described in Chapters 4 and 5 had the same goal and other than the lysis and endospore filtration procedures, utilized the same methods, they have many common applications and implications. The work that has been performed is two fold: one, discovering biomarkers that can be used for the identification of unknowns at the species and serotype level; two, molecular subtyping of strains that may be useful for outbreak or source identification studies. The first area uses LC/MS to complement and at times outperform immunoassays and real-time PCR. These two techniques are targeted approaches often used for identification, but not strain level typing. It would be impractical to develop primers or antibodies for every strain of *E. coli* O157:H7 for example. For this reason, a screening approach that does not require primers or antibodies such as PFGE is used for strain typing. LC/MS however, can accomplish target detection and strain differentiation in a single analysis. Due to the two fold nature of this method; its implications will be discussed separately.

As mentioned previously, the LC/MS approach is complementary to existing techniques by providing proteome information and is advantageous for bacteria for which primers or antibodies are commercially unavailable. However, in addition to being well

established, compared to LC/MS, EIA and real-time PCR methods are less expensive to use and maintain and more amenable to multiplexing for high throughput applications making them ideal for biomarker detection. On the other hand, neither of these techniques have the capabilities for biomarker discovery. The LC/MS approach here is ideal for biomarker discovery having the ability to detect hundreds of proteins in a single analysis and can detect subtle differences between closely related isolates. Once discovered, the biomarkers mentioned in Chapters 4 and 5 could be purified and used for antibody formation to enhance serological approaches or be sequenced and used to reverse engineer novel real-time PCR primers. By monitoring intact proteins, information useful for both primer design and antibody formation are collected simultaneously. In addition to antibody formation, another way to assist clinical immunological assays using these markers would be to have them sequenced, then grown recombinantly and used in assays to detect host antibodies – indicating exposure.

Of the biomarkers discovered in this work, the best candidates for antibody formation or primer synthesis would be ones that are not isoforms of proteins common to other species or serotypes and rather, are completely unique. For example, as described in Chapter 4, one protein useful for the speciation of *Shigella* was a marker found only in *S. flexneri* weighing 7,287 Da which eluted at 27.9 minutes. In all of the other *E. coli* and *S. sonnei* isolates examined however, a protein of similar mass (7,273 Da) and the same retention time was observed. Because of the similar mass and identical retention behavior, these proteins are likely isoforms of one another. The mass difference of 14 Da could be a PTM (methylation) or an amino acid substitution (e.g. I for V) indirectly

caused by a SNP. Since EIA techniques already suffer from low specificity yielding a high false positive rate, they would not likely be able to discern these two proteins. Likewise, it has been demonstrated that using the same primers, real-time PCR will amplify two strands of DNA that differ by one base, as is the case when a SNP has occurred. This will be most problematic if the SNP occurs between the primer locations. Once more this demonstrates the specificity of the LC/MS approach, which unlike real-time PCR or immunoassays, can readily distinguish most protein isoforms.

An example of a protein that has no clearly observed isoforms in the isolates studied and would therefore be a best initial candidate for primer or antibody formation is the 18,996 Da *E. coli* O157:H7 marker which eluted at 43.3 minutes. Developing new primers will allow novel determinations by real-time PCR such as discriminating *S. flexneri* from *S. sonnei* or allow more confident identifications of bacteria for which only one or few primers exist. Likewise, if more qualifiers for identification are added, the lack of specificity of EIA methods can be reduced. For example, a serological approach for *E. coli* O157:H7 may involve diagnosis for only the O and H antigens. To help alleviate the high false positive rate, an antibody could be made for the 18,996 Da protein and the EIA assay be modified so that a positive result would be reported only if all three antigens were detected.

In addition to assisting the well established techniques (e.g. real-time PCR) the biomarkers discovered in this study could also be used to enhance the less well known, but promising technique of protein microarrays.[146-149] Protein microarrays involve the immobilization labeled proteins (e.g. antibodies, enzymes) to a chip. Samples can then

be applied to the chip and reactions detected by fluorescence or radioactivity etc. Recent efforts have also reported the combination MS and microarray technology.[150, 151] By being in array format, these devices are high throughput and can be utilized for to quantify and detect hundreds or thousands of specific protein-protein or other protein-ligand interactions simultaneously. Performing an EIA experiment using a chip-based microarray would be an ideal platform for detecting multiple antigens for the same bacteria as described above for *E. coli* O157:H7. Extension of this technology to the diagnosis of bacteria has been demonstrated.[70, 102]

As for strain typing, the second area of this work, the method described here has shown strain level typing capabilities which at times exceed those of PFGE. Additionally, LC/MS is easily automated, less labor intensive, provides complementary information by examining the proteome instead of the genome, and for small sample sizes, has higher throughput. Like PFGE data, the data from this approach could be uploaded to a database such as PulseNet, which monitors outbreaks by analyzing PFGE data from public health laboratories across the nation.

To better encourage other public health laboratories to consider LC/MS technology, efforts could be made to reduce the initial cost of the instrument. The method as described here operates in full-scan mode in order to detect as many proteins as possible and since currently no sequencing is performed, MS/MS capabilities are not required. For this reason, a single TOF instrument could be used rather than a QTOF. In addition to being less expensive, single TOF instruments are more sensitive because they

lack the two additional quadrupole mass analyzers and therefore would suffer less loss from transmission within the quadrupole region.[5]

Another issue that should be addressed before extending this approach to other public health laboratories is intra and inter-lab variability. Intra-lab variability may include differences in the samples run at the beginning of a batch vs. those run at the end as well as day-to-day variability. These variations may result from lab temperature fluctuation which could affect mass accuracy, degradation in the cleanliness of the column and source which could affect retention behavior and intensity, sample degradation with time and minor variations in mobile phase composition which would also alter retention time. Examples of inter-laboratory variability would mainly stem from differences in LC or MS instrumentation such as discrepancies in the tubing size used to connect the LC and MS causing changes in dead volume or in column oven heating efficiencies altering retention times, or in the performance of the mass spectrometer used which would affect intensity and mass accuracy. Differences in growing or lysis procedures may also result in variability therefore to minimize variability, partner labs should use methods and instrumentation as similar to this one as possible, i.e. an SOP should be formulated and followed as closely as possible.

Besides using as similar as possible methods and instrumentation across laboratories, another way to address intra and inter-laboratory variability would be to add a standard protein to the lysate prior to analysis. Being a standard, this protein's mass and retention time would be well characterized before use. To address retention time shifts, the biomarker retention times would not only be reported as ± 0.5 min, but also as

on a system where on average myoglobin eluted at 32.4 minutes e.g. If another laboratory uses a system with more dead volume than the one described here and myoglobin elutes at 32.9 minutes, 0.5 minutes could be subtracted from all of the results in that lab prior to searching for the biomarkers observed in this laboratory. This standard could also be utilized as an internal calibrant to correct for mass shifts. Finally, since the same concentration would be added to all lysates, the intensities of the proteins in the lysates could be held relative to this standard and these relative intensities be used to better ascertain quantitative differences between lysates. This is in contrast to using absolute intensities which may vary based on source cleanliness etc.

Future work for strain typing analysis would involve applying the LC/MS method to numerous isolates from a known outbreak to see what if any new information is provided from this technique. Samples from both exposed individuals and from the organisms collected from the actual foodstuffs responsible and other related samples such as the soil where the produce was grown etc., should be tested. This experiment would use LC/MS to compare the differences and similarities amongst strains known to be common and or unrelated by both epidemiological investigation and PFGE data. The isolates studied should also come with detailed case information to help answer the questions: "If a pattern is different, did it come from a different state, a different route of exposure etc.? Was the patient administered antibiotics prior to sample collection?" Case history would be implemented to compare and classify data sets and to examine how well proteomic data correlate to known case history and demographic/epidemiological information.

Much could be learned from the analysis of superbugs. Beyond finding markers for identification, unique proteins could be sequenced and if the protein and corresponding gene function is known, this could yield insight into the mechanism of resistance. If the marker protein function is unknown, some understanding can be ascertained by determining the sequence homology with known proteins in homology databases. Better understanding the mechanism of a bacterium's aptitude for developing resistance (drug, heat, or otherwise) could be used to predict future resistances by common pathogens and may assist with: antibiotic drug design and delivery, the development of novel disinfectant and infection control measures or possibly learning how to reverse the resistance process.

After successful implementation of the method for the characterization and identification of bacteria described in Chapters 4 and 5, the method was optimized to improve throughput. Chapter 6 describes the optimization of the separation of ten protein standards using UPLC and the application of that optimized method to cell lysate analysis. Temperature, organic modifier and particle size were adjusted for optimal resolution, speed and sensitivity. The optimized method incorporated a combination of high temperature, a strong solvent and small particles at very high pressure. This optimization resulted in a slight increase in sensitivity and a 2/3 reduction in runtime from 75 minutes to 25 minutes. Furthermore, the reduction in the size of the data file resulted in a concomitant decrease in data analysis time by Prowtrawler6. For this reason, the total post culture and lysis time required for data collection and analysis

decreased by 2/3 from 120 minutes (two hours) to 40 minutes. This increase in throughput allows data directed decisions to be made in less time.

In addition to the expected sensitivity increase gained by narrowing the protein peak width to 1/3 of its original value, utilizing UPLC yielded an unforeseen increase in sensitivity. Protrawler6 software has a data file size limit of 2.1 GB. If a file contains more than that amount of information, Protrawler6 will stop processing after 2.1 GB and ignore any remaining information. During tuning of the mass spectrometer, the optimal detector voltage was found to be 1900 V. However, since the data file size is proportional to the number of ions counted and the time spent counting them, the 75 minute chromatograms required a detector voltage of 1750 V or less to remain under 2.1 GB. But after optimization of the method with UPLC, a 25 minute chromatogram could be obtained using a detector voltage of 1900 V and the ensuing data file have a size of 2.1 GB or less. Thus, incorporating UPLC not only increases the throughput three fold, it also allows analysis at the optimized detector voltage which will increase sensitivity and likely enhance the observation of lower abundance proteins.

Relative to small molecule analysis, intact protein chromatography is challenging. For this reason, there was some initial skepticism from within this group and even from more than one scientist from the UPLC vendor as to whether the advantages typically observed with the UPLC of small molecules could also be observed with intact proteins. Chapter 6 reports the first successful separation of intact proteins using commercial UPLC equipment and demonstrates the advantages typically observed for small molecules (increased throughput, improved resolution etc.). The potential applications of

this work are broad and include: biopharmaceuticals, protein supplements, food allergens, protein toxins and various top-down proteomic investigations. One particularly interesting application that could be pursued is coupling the most recent advances in chromatography (e.g. UPLC) with the most recent advances in top-down proteomics (top-down sequencing on a chromatographic time scale).[152-154] Combined, these two advancements could provide sequence analysis on a complex mixture of proteins in approximately 30 minutes post lysis – an unheard of feat.

In addition to the throughput enhancement using UPLC, modifications could be made to the HPLC gradient to shorten the runtime. In the *E. coli* O157:H7 chromatogram in Figure 1, it appears that much time is wasted between 45 and 58 minutes. However, proteins extracted using LP1 did elute during this time, with the most well retained biomarker eluting at 54.9 minutes (Figure 16). The tail end of this peak occurs at approximately 56.5 minutes and the dead volume as seen by the small non retained peaks at the beginning of the chromatogram in Figures 1 and 16 requires ~ 1.5 minutes. Therefore, the peak which finished eluting at 56.5 minutes in the chromatogram, actually left the column at 55 minutes. Allowing an extra minute for retention time shifts, the gradient which currently stops at 60 minutes (Table 3) need only run to 56 minutes. Since the current gradient runs from 5-55%B in 60 min (0.83 %B/min) subtracting four minutes (3.32 %B) would make the new gradient from 5-51.68%B in 56 minutes while keeping the same slope (0.83 %B/min). Maintaining a constant slope allows the runtime to be reduced without affecting previous separation results.

With the last biomarker eluting at 55 minutes, the step in Table 3 of 55-95%B from 60-65 minutes could be removed - saving another 5 minutes. The new gradient would begin pumping 95%B at 56.01 minutes. The remaining 10.5 minutes in Table 3 would remain unchanged allowing the total runtime to be 66.5 minutes, a 9 minute reduction from the gradient in Table 3. This new gradient could then be reduced by 2/3 using UPLC to approximately 22.2 minutes.

Although applied only to bacterial cell lysates in this work, the approach described here using UPLC/MS combined with automated charge state deconvolution and spectral subtraction would be well suited to aid any comparative proteomics analysis. Examples include: the comparison of diseased cells vs. non diseased, treated diseased cells vs. non-treated diseased cells, as well as plant and wildlife proteomic analysis e.g., before and after environmental stresses or exposure to pollutants etc.

Using LC/MS (a technique commonly found in analytical chemistry laboratories but rarely found in diagnostic microbiology laboratories), a complementary approach with certain advantages over typical microbiological methods has been developed. This study used as a proof of concept ten isolates of *Shigella* and *Escherichia* and thirteen isolates of *Clostridium* including the species most often implicated in human disease to examine the efficacy of the LC/MS approach to characterize microorganisms. This approach has demonstrated greater specificity than obtainable using current real-time PCR protocols, allowing for distinctions at the strain level, and is automated and less labor intensive than PFGE, the gold standard for subtyping. No primers, antibodies or proteomic database searches were required for this study. Using a combination of

automated charge state deconvolution and spectral subtraction, reproducible intact protein biomarkers were observed at the species and sub-species (serotype, strain) levels, and were successfully implemented for the identification of unknown pathogens.

This method could be used to enhance public health response (treatment, remediation, etc.) by guiding the data directed decision making of hospital staff, public health officials and investigating authorities. Ultimately, this method would be ideal as a complementary technique adding another dimension to the polyphasic approach of bacterial identification. This enhanced polyphasic approach would then lead to more confident results which are required to cease production and distribution during an outbreak, or for the prosecution of suspected terrorists. Additionally, this LC/MS method could be expanded to monitor biomarkers for other foodborne pathogens or biowarfare agents, and implemented to screen food items before they enter the market to prevent possible outbreaks from occurring. In addition to its complementary nature, the biomarkers discovered using this approach could be utilized to enhance current and future methods in the field of diagnostic microbiology.

References

References

1. Foodborne Illness. http://www.cdc.gov/ncidod/dbmd/diseaseinfo/foodborneinfections_g.htm (Accessed 04/22/07).
2. Food Safety and Foodborne Illness. http://www.who.int/mediacentre/factsheets/fs237/en/ (Accessed 04/22/07).
3. Rotz, L.D.; Khan, A.S.; Lillibridge, S.R.; Ostroff, S.M.; Hughes, J.M. *Emerg. Infect. Dis*. **2002**, *8*, 225-230.
4. Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, M.G.; Whitehouse, C.M. *Science* **1989**, *246*, 64-71.
5. Chernushevich, I.V.; Loboda, A.V.; Thomson, B.A. *J. Mass Spectrom*. **2001**, *36*, 849-865.
6. Williams, T.L.; Leopold, P.; Musser, S. *Anal. Chem.* **2002**, *74*, 5807-5813.
7. Harrison, S.T.L. *Biotech. Adv*. **1991**, *9*, 217-240.
8. Mazzeo, J.R.; Neue, U.D.; Kele, M.; Plumb, R.S. *Anal. Chem*. **2005**, 460A-467A.
9. Nguyen, D.T.T.; Guillarme, D.; Rudaz, S.; Veuthey, J.L. *J. Sep Sci.* **2006**, *29*, 1836-1848.
10. Schluter, H. In *Protein Liquid Chromatography*; Kastner, M. Ed.; Elsevier: Amsterdam; New York, 2000; Chapter 3.
11. Bacteriology, The History Channel Website. http://www.history.com/encyclopedia.do?articleId=202006 (Accessed 01/30/08).
12. Todar, K. Growth of Bacterial Populations, Todar's Online Textbook of Bacteriology. http://www.textbookofbacteriology.net/growth.html (Accessed 01/29/08).
13. Sutherland, J.B.; Rafii, F. In *Identification of Microorganisms by Mass Spectrometry*; Wilkins, C.L.; Lay, J.O., Eds.; John Wiley & Sons, Inc.: Hoboken, 2006; Chapter 1.
14. Feng, P. Bacteriological Analytical Manual, 8th Edition, Final Revision: 2001-Jan-25. http://www.cfsan.fda.gov/~ebam/author (Accessed 03/02/07).
15. Forbes, B.A.; Sahm, D.F.; Weissfeld, A.S. *Bailey & Scott's Diagnostic Microbiology*, Eleventh Ed.; Mosby: St. Louis, 2002; Chapter 13.
16. Houpikian, P.; Raoult, D. *Emerg. Infect. Dis*. **2002**, *8*, 122-131.
17. Saiki, R.K.; Gelfand, D.H.; Stoffel, S.; Scharf, S.J.; Higuchi, R.; Horn, G.T.; Mullis, K.B. Erlich, H.A. *Science* **1988**, *239*, 487-491.
18. Lekowska-Kochaniak, A.; Czajkowska, D.; Popowski, J. *Acta Microbiol. Polon*. **2002**, *51*, 327-337.

19. Higuchi, R.; Fockler, C. Dollinger, G.; Watson, R. *Biotechnology* **1993**, *11*, 1026-1030.
20. Fox, A. In *Identification of Microorganisms by Mass Spectrometry*; Wilkins, C.L.; Lay, J.O., Eds.; John Wiley & Sons, Inc.: Hoboken, 2006; Chapter 2.
21. Schwartz, D.C.; Saffran, W.; Welsh, J.; Haas, R.; Goldenberg, M.; Cantor, C.R. *Cold Spring Harbor Symp. Quant. Biol*. **1983**, *47*, 189-195.
22. Forbes, B.A.; Sahm, D.F.; Weissfeld, A.S. *Bailey & Scott's Diagnostic Microbiology*, Eleventh Ed.; Mosby: St. Louis, 2002; Chapter 12.
23. Goering, R.V. In *Molecular Microbiology: Diagnostic Principles and Practice*; Persing, D.H.; Tenover, F.C.; Versalovic, J.; Tang, Y.; Unger, E.R.; Relman, D.A.; White, T.J., Eds.; ASM Press: Washington, D.C., 2004; Chapter 15.
24. Swaminathan, B.; Barett, T.J.; Hunter, S.B.; Tauxe R.V. *Emerg. Infect. Dis*. **2001**, *7*, 382-389.
25. Monday, S.R.; Whittam T.S.; Feng, P.C. *J. Infect. Dis*. **2001**, *184*, 918-921.
26. Tominaga, A.; Mahmoud, M.A.; Mukaihara, T.; Enomoto, M. *Mol. Microbiol*. **1994**, *12*, 277-285.
27. Monday, S.R.; Minnich, S.A.; Feng, P.C. *J. Bacteriol*. **2004**, *186*, 2319-2327.
28. Fenselau, C. In *Mass Spectrometry for the Characterization of Microorganisms*; Fenselau, C., Ed.; ACS Symposium Series 541; American Chemical Society: Washington, DC, 1994. Chapter 1.
29. Drucker, D.B. In *Mass Spectrometry for the Characterization of Microorganisms*; Fenselau, C., Ed.; Chapter 3. ACS Symposium Series 541; American Chemical Society: Washington, DC, 1994.
30. Song, Y.; Talaty, N.; Tao, W.A.; Pan, Z.; Cooks, R.G. *Chem. Commun*. **2007**, 61-63.
31. Johnson, Y.A.; Nagpal, M.; Krahmer, M.T.; Fox, K.F.; Fox, A. *J. Microbiol. Meth*. **2000**, *40*, 241-254.
32. Walsh, C.T. In *Posttranslational Modifications of Proteins: Expanding Nature's Inventory*; Roberts and Company Publishers: Colorado, 2006; Preface.
33. Fenselau, C.; Pribil, P. In *Identification of Microorganisms by Mass Spectrometry*; Wilkins, C.L.; Lay, J.O., Eds.; John Wiley & Sons, Inc.: Hoboken, 2006; Chapter 12.
34. Williams, T.L.; Callahan, J.H.; Monday, S.R.; Feng, P.C.H.; Musser, S.M. *Anal. Chem.* **2004**, *76*, 1002-1007.
35. Dworzanski, J.P.; Syder, A.P. *Expert Rev. Proteomics* **2005**, *2*, 863-878.
36. Demirev, P.A.; Feldman, A.B.; Lin, J.S. *Johns Hopkins APL Tech. Dig*. **2004**, *25*, 27-37.
37. Williams, T.L.; Monday, S.R.; Edelson-Mammel, S.; Buchanan, R.; Musser, S.M. *Proteomics* **2005**, *5*, 4161-4169.
38. Petit, L.; Gibert, M.; Popoff, M.R. *Trends Microbiol*. **1999**, *7*, 104-110.
39. Schaechter, M.; Ingraham, J.L.; Neidhardt, F. *Microbe*; ASM Press: Washington, DC 2006; Chapter 14.

40. Simpson, C.F. In *High Performance Liquid Chromatography: Principles and Methods in Biotechnology*; Katz, E.D., Ed.; John Wiley & Sons Ltd: West Sussex, 1996; Chapter 1.
41. Eschelbach, J.W.; Jorgenson, J.W. *Anal. Chem*. **2006**, *78*, 1697-1706.
42. Nguyen, S.; Fenn, J.B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1111-1117.
43. Willoughby, R.; Sheehan, E.; Mitrovich, S. *A Global View of LC/MS: How to Solve Your Most Challenging Analytical Problems*. First ed. Global View Publishing: Pittsburg, 1998; Chapter 2.
44. The Mass Spectrometer: Instrument Architectures and Main Characteristics. http://www.waters.com/watersdivision/ContentD.asp?ref=EGOO%2D66MNYR 2007 Waters Corporation. (Accessed 02/28/07).
45. Mann, M; Whitehouse, C.; Fenn, J.B. *Anal. Chem.* **1989**, *61*, 1702-1708.
46. Reinhold, B.B.; Reinhold, V.N. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 207-215.
47. Labowsky, M.; Whitehouse, C.; Fenn, J.B. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 71-84.
48. Hagen, J.J.; Monnig, C.A. *Anal. Chem.* **1994**, *66*, 1877-1883.
49. Ferrige, A.G.; Seddon, M.J.; Jarvis, S.A. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 374-379.
50. Ferrige, A.G.; Seddon, M.J.; Green, B.N.; Jarvis, S.A.; Skilling, J. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 707-711.
51. Khardori, N. In, *Bioterrorism Preparedness*; Khardori, N., Ed.; WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2006. Chapter 1.
52. Smith, J.L.; Fratamico, P.M. In *Foodborne Pathogens Microbiology and Molecular Biology*; Fratamico, P.M.; Bhunia, A.K.; Smith, J.L., Eds.; Caister Academic Press: Wymondham, 2005; Chapter 18.
53. Nataro J.P.; Kaper, J.B. *Clin. Microbiol. Rev.* **1998**, *11*(1), 142-201.
54. *Escherichia coli* O157:H7. Centers for Disease Control and Prevention. http://www.cdc.gov/ncidod/dbmd/diseaseinfo/escherichiacoli_g.htm (Accessed 07/27/07).
55. Congressional testimony by Lonnie J. King, D.V.M. U.S. Department of Health and Human Services. http://www.hhs.gov/asl/testify/t061115.html (Accessed 07/27/07).
56. Shigellosis. Centers for Disease Control and Prevention. http://www.cdc.gov/ncidod/dbmd/diseaseinfo/shigellosis_t.htm (Accessed 07/27/07).
57. Kotloff, K.L; Winickoff, J.P.; Ivanoff, B.; Clemens, J.D.; Swerdlow, D.L.; Sansonetti, P.J.; Adak, G.K.; Levine, M.M. *Bull. World Health Organ.* **1999**, *77* (8), 651-666.
58. CDC Select Agent Program. http://www.cdc.gov/od/sap/docs/salist.pdf (Accessed 07/27/07).
59. Kolavic, S.A.; Kimura, A.; Simons, S.L.; Slutsker, L.; Barth, S.; Haley, C.E. *JAMA* **1997**, *278*, 396-398.

60. Carus, S.W. *Bioterrorism and Biocrimes: The Illicit Use of Biological Agents in the 20th Century*. Center for Counterproliferation Research, National Defense Univ., Washington, DC, 2002.
61. Ake, J.A.; Jelacic, S.; Ciol, M.A.; Watkins, S.L.; Murray, K.F.; Christie, D.L.; Klein, E.J.; Tarr, P.I. *Pediatrics* **2005**, *115*, e673-680.
62. Centers for Disease Control and Prevention (CDC). *MMWR Morb. Mortal. Wkly. Rep*. **2006**, *55* (38), 1042-1045.
63. Busse, H.J.; Denner, E.B.M.; Lubitz, W. *J. Biotechnol*. **1996**, *47*, 3-38.
64. Sabat, G.; Rose, P.; Hickey, W.J.; Harkin, J.M. *Appl. Environ. Microbiol.* **2000**, *66*, 844-849.
65. Pupo, G.M.; Lan, R.; Reeves, P.R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10567-10572.
66. Schwartz, D.C.; Saffran, W.; Welsh, J.; Haas, R.; Goldenberg, M.; Cantor, C.R. *Cold Spring Harbor Symp. Quant. Biol*. **1983**, *47*, 189-195.
67. Swaminathan, B.; Barett, T.J.; Hunter, S.B.; Tauxe R.V. *Emerg. Infect. Dis*. **2001**, *7* (3), 382-389.
68. Williams, T.L.; Musser, S.M.; Nordstrom, J.L.; DePaola, A.; Monday, S.R. *J. Clin. Microbiol.* **2004**, *42*, 1657-1665.
69. Williams, T.L.; Monday, S.R.; Feng, P.C.H.; Musser, S.M. *J. Biomol. Tech.* **2005**, *16*, 134-142.
70. Mezzasoma, L.; Bacarese-Hamilton, T.; Di Cristina, M; Rossi, R.; Bistoni, F.; Crisanti, A. *Clin. Chem.* **2002**, *48* (1), 121-130.
71. Todar, K. The Pathogenic Clostridia, from Todar's Online Textbook of Bacteriology, 2006. http://www.textbookofbacteriology.net/clostridia.html (Accessed 03/04/08).
72. Stackebrandt, E.; Rainey, F.A. In The Clostridia: Molecular Biology and Pathogenesis; Rood, J.I.; McClane, B.A.; Songer, J.G., Titball, R.W., Eds.; Academic Press: San Diego, 1997; Chapter 1.
73. Johnson, E.A. In The Clostridia: Molecular Biology and Pathogenesis; Rood, J.I.; McClane, B.A.; Songer, J.G., Titball, R.W., Eds.; Academic Press: San Diego, 1997; Chapter 3.
74. Borriello, S.P.; Carman, R.J. In Clostridia in Gastrointestinal Disease; Borriello, S.P., Ed.; CRC Press: Boca Raton, 1985; Chapter 10.
75. European Commission, Directorate-General for Research. In Clostridia in Medical, Veterinary and Food Microbiology: Diagnosis and Typing; Mainil, J.; Duchesnes, C.; Granum, P.E., Minozzi, M.G.; Peck, M.; Pelkonen, S.; Popoff, M.; Stackebrandt, E.; Titball, R., Eds.; European Communities: Luxembourg, 2006; pp 7.
76. Asha, N.J.; Tompkins, D.; Wilcox, M.H. *J. Clin. Microbiol.* **2006**, *44*, 2785-2791.
77. Schroeder, M.S. *Am. Family Phy*. **2005**, *71*, 921-928.
78. Barbut, F.; Petit, J.C. *Clin. Microbiol. Infect*. **2001**, *7*, 405-410.
79. Rodriguez-Palacios, A.; Staempfli, H.R.; Duffield, T.; Weese, J.S. *Emerg. Infect. Dis*. **2007**, *13*, 485-487.

80. Broda, D.M.; DeLacy, K.M.; Bell, R.G.; Braggins, T.J.; Cook, R.L. *Int. J. Food Microbiology* **1996**, *29*, 335-352.
81. Rodriguez-Palacios, A.; Staempfli, H.R.; Duffield, T.; Peregrine, P.S.; Trotz-Williams, L.A.; Arroyo, L.G.; Brazier, J.S.; Weese, J.S. *Emerg. Infect. Dis.* **2006**, *12*, 1730-1736.
82. Ackerman, G.; Thomalla, S.; Ackerman, F.; Schaumann, R.; Rodloff, A.C.; Ruf, B.R. *J. Med. Microbiol*. **2005**, *54*, 149-153.
83. Asha, N.J.; Wilcox, M.H. *J. Med. Microbiol*. **2002**, *51*, 891-894.
84. Novak, J.S.; Peck, M.W.; Juneja, V.K.; Johnson, E.A. In Foodborne Pathogens: Microbiology and Molecular Biology; Fratamico, P.M.; Bhunia, A.K.; Smith, J.L., Eds.; Caister Academic Press: Wymondham, 2005; Chapter 19.
85. Juneja, V.K.; Novak, J.S.; Huang, L.; Eblen, B.S. *Food Control* **2003**, *14*, 163-168.
86. Davis, C.J. *Emerg. Infect. Dis*. **1999**, *5*, 509-512.
87. Garret, L. *Foreign Affairs* **1996**, *75* (1), 66-79.
88. Krt, B. *FEMS Immunol. Med. Micrbiol.* **1999**, *24*, 293-297.
89. Pelleschi, M.E. *Crit. Care Nurse* **2008**, *28*, 27-35.
90. Bruggemann, H. *Curr. Opin. Microbiol*. **2005**, *8*, 601-605.
91. Andreesen, J.R.; Bahl, H.; Gottschalk, G. In Clostridia; Minton, N.P.; Clarke, D.J., Eds.; Plenum Publishing Corporation: New York, 1989; Chapter 2.
92. Graham, R.L.J.; Graham, C.; McMullan, G. *Microb. Cell. Fact*. **2007**, *6*, 26.
93. Anhalt, J.P.; Fenselau, C. *Anal. Chem*. **1975**, *47*, 219-225.
94. Fox, A. *J. Clin. Microbiol.* **2006**, *44*, 2677-2680.
95. Black, G.E.; Snyder, A.P.; Heroux, K.S. *J. Micriobiol. Meth*. **1997**, *28*, 187-199.
96. Simpson, R.C.; Fenselau, C.C.; Hardy, M.R.: Townsend, R.R.; Lee, Y.C.; Cotter, R.J. *Anal. Chem*. **1990**, *62*, 248-252.
97. Hurst, G.B.; Doktycz, M.J.; Vass, A.A.; Buchanan, M.V. *Rapid Commun. Mass Spectrom*. **1996**, *10*, 377-382.
98. Greis, K.D.; Zhou, S.; Siehnel, R.; Klanke, C.; Curnow, A.; Howard, J.; Layh-Schmitt, G. *Antimicrob. Agents Chemother*. **2005**, *49*, 3428-3434.
99. Wall, D.B.; Parus, S.J.; Lubman, D.M. *J. Chromatogr. B* **2001**, *763*, 139-148.
100. Wren, B.W. *Future Microbiol.* **2006**, *1* (3), 243-245.
101. Tenover, F.C.; Arbeit, R.D.; Goering, R.V.; Mickelsen, P.A.; Murray, B.E.; Persing, D.H.; Swaminathan, B. *J. Clin. Microbiol*. **1995**, *33*, 2233-2239.
102. Steller, S.; Angenendt, P.; Cahill, D.J.; Heuberger, S.; Lehrach, H.; Kreutzberger, *J. Proteomics*, **2005**, *5*, 2048-2055.
103. Regnier, F. E. *Science* **1983**, *222*, 245-252.
104. Walsh, G. *Trends Biotechnol*. **2005**, *23*, 553-558.
105. Yates, J.R. *Annu. Rev. Biophys. Biomol. Struct*. **2004**, *33*, 297-316.
106. Roth, M.J.; Forbes, A.J.; Boyne, M.T.; Kim, Y.; Robinson, D.E.; Kelleher, N.L. *Mol. Cell Proteomics* **2005**, *4*, 1002-1008.
107. Kelleher, N.L. *Anal. Chem*. **2004**, 197A- 203A.
108. Issaq, H.J.; Conrads, T.P.; Janini, G.M.; Veenstra, T.D. *Electrophoresis* **2002**, *23*, 3048-3061.

109. Zolla, L.; Rinalducci, S.; Timperio, A.M.; Huber, C.G. *Eur. J. Mass Spectrom.* **2004**, *10*, 321-333.
110. Dauly, C.; Perlman, D.H.; Costello, C.E.; McComb, M.E. *J. Proteome Res.* **2006**, *7*, 1688-1700.
111. Martin, A.J.P.; Synge, R.L.M. *Biochem. J.* **1941**, *35*, 1358-1368.
112. Swartz, M.E. *J. Liq. Chromatogr. Rel. Technol.* **2005**, *28*, 1253-1263.
113. MacNair, J.E.; Patel, K.D.; Jorgenson, J.W.; Moseley, M.A., III. *Rapid Commun. Mass Spectrom*. **1997**, *11*, 1279-1285.
114. Garcia, M.C.; Marina, M.L.; Torre, M. *J. Chromatogr. A* **2000**, *880*, 169-187.
115. Mellors, J.S.; Jorgenson, J.W. *Anal. Chem*. **2004**, *76*, 5441-5450.
116. Wu, N.; Liu, Y.; Lee, M.L. *J. Chromatogr*. A **2006**, *1131*, 142-150.
117. Lubman, D.M.; Kachman, M.T.; Wang, H.; Gong, S.; Yan, F.; Hamler, R.L.; O'Neil, K.A.; Zhu, K.; Buchanan, N.S.; Barder, T.J. *J. Chromatogr. B* **2002**, *782*, 183-196.
118. Rainville, P.D.; Stumpf C.L.; Shockcor J.P.; Plumb R.S.; Nicholson J.K. *J. Proteome Res*. **2007**, *6*, 552-558.
119. Kaufmann, A.; Butcher, P.; Maden, K.; Widmer, M. *Anal. Chim. Acta* **2007**, *586*, 13-21.
120. Zhang, Y.; Jiao, J.; Cai, Z.; Zhang, Y.; Ren, Y. *J. Chromatogr. A* **2007**, *1142*, 194-198.
121. Salquebre, G.; Bresson, M.; Villain, M.; Cirimelle, V.; Kintz, P. *J. Anal. Tox*. **2007**, *31*, 114-118.
122. Bendahal, L.; Hansen, S.H.; Gammelgaard, B.; Sturup, S.; Nielsen, C. *J. Pharm. Biomed. Anal.* **2006**, *40*, 648-652.
123. O'Connor, D.; Mortishire-Smith, R.; Morrison, D.; Davies, A.; Dominguez, M. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 851-857.
124. Wyndham, K.D.; O'Gara, J.E.; Walter, T.H.; Glose, K.H.; Lawrence, N.L.; Alden, B.A.; Izzo, G.S.; Hulda, C.J.; Iraneta, P.C. *Anal. Chem.* **2003**, *75*, 6781-6788.
125. Acquity UPLC BEH Column - Care and Use Instructions. http://www.waters.com/WatersDivision/pdfs/715001371CU.pdf (Accessed 09/18/07).
126. Paliwal, S.K.; De Frutos, M.; Regnier, F.E. In, *Methods in Enzymology*, *Vol. 270*; Karger, B.L.; Hancock, W.S. Eds.; Academic Press Inc.: San Diego, 1996; Chapter 6.
127. Benedek, K. In, *High Performance Liquid Chromatography: Principles and Methods in Biotechnology*; Katz, E.D. Ed.; John Wiley & Sons Ltd.: Chichester, 1996; Chapter 9.
128. Wren, S.A.C.; Tchelitcheff, P. *J. Chromatogr. A* **2006**, *1119*, 140-146.
129. Greibrokk, T.; Andersen, T. *J. Chromatogr. A* **2003**, *1000*, 743-755.
130. Wang, Y.; Balgley, B.M.; Rudnick, P.A.; Lee, C.S. *J. Chromatogr. A* **2005**, *1073*, 35-41.
131. Yang, Y.; Ma, L.; Carr, P.W. *J. Chromatogr. A* **2005**, *1079*, 213-220.
132. Chen, H.; Horvath, Cs. *J. Chromatogr. A*. **1995**, *705*, 3-20.

133. Szabelski, P.; Cavazzini, A.; Kaczmarski, K.; Liu, X.; Van Horn, J.; Guiochon, G. *J. Chromatogr. A* **2002**, *950*, 41-53.
134. Oroszlan, P.; Wicar, S.; Teshima, G.; Wu, S.L.; Hancock, W.S.; Karger, B.L. *Anal. Chem.* **1992**, *64*, 1623-1631.
135. Dillon, T.M.; Bondarenko, P.V.; Rehder, D.S.; Pipes, G.D.; Kleemann, G.R.; Speed Ricci, M. *J. Chromatogr. A* **2006**, *1120*, 112-120.
136. Rehder, D.S.; Dillon, T.M.; Pipes, G.D.; Bondarenko, P.V. *J. Chromatogr. A* **2006**, *1102*, 164-175.
137. Neville, W.A. In *Methods in Molecular Biology, vol. 244: Protein Purification Protocols: Second Edition*; Cutler, P. Ed.; Humana Press Inc.: Totowa, 2004; Chapter 28.
138. Everley, R.A.; Mott, T.M.; Wyatt, S.A.; Leopold, P.; Croley T.R. In, *Proceedings of the 55th ASMS Conference,* Indianapolis, 2007; MPT-322, 561 (CD).
139. Molloy, M.P.; Herbert, B.R.; Walsh, B.J.; Tyler, M.I.; Traini, M.; Sanchez, J.C.; Hochstrasser, D.F.; Williams, K.L.; Gooley, A.A. *Electrophoresis*, **1998**, *19*, 837-844.
140. Dreger, M. *Mass Spectrom. Rev*. **2003**, *22*, 27-56.
141. Zhou, F.; Hanson, T.E.; Johnston, M.V. *Anal. Chem.* **2007**, *79*, 7145-7153.
142. Park, K.H.; LiPuma, J.J.; Lubman, D.M. *Anal. Chim. Acta* **2007**, *592*, 91-100.
143. Simpson, D.C.; Ahn, S.; Pasa-Tolic, L.; Bogdanov, B.; Mottaz, H.M.; Vilkov, A.N.; Anderson, G.A.; Lipton, M.S.; Smith, R.D. *Electrophoresis* **2006**, *23*, 2722-2733.
144. Zheng, S.; Schneider, K.A.; Barder, T.J.; Lubman, D.M. *Biotechniques* **2003**, *35*, 1202-1212.
145. Chaudhury A.; Nath, G.; Tikoo, A.; Sanyal, S.C. *J. Diarrhoeal Dis. Res.* **1999**, *17* (2), 85-87.
146. Poetz, O.; Schwenk, J.M.; Kramer, S.; Stoll, D.; Templin, M.F.; Joos, T.O. *Mech. Age Dev.* **2005**, *126*, 161-170.
147. Templin, M.F.; Stoll, D.; Schwenk, J.M.; Poetz, O.; Kramer, S.; Joos, T.O. *Proteomics* **2003**, *3*, 2155-2166.
148. Kricka, L.J.; Master, S.R.; Joos, T.O.; Fortina, P. *Ann. Clin. Biochem*. **2006**, *43*, 457-467.
149. Cretich, M.; Damin, F.; Pirri, G.; Chiari, M. *Biomolec. Eng*. **2006**, *23*, 77-88.
150. Evans-Nguyen, K.M.; Tao, S.C.; Zhu, H.; Cotter, R.J. *Anal. Chem*. **2008**, *80*, 1448-1458.
151. Pal, M.; Moffa, A.; Sreekumar, A.; Ethier, S.P.; Barder, T.J.; Chinnaiyan, A.; Lubman, D.M. *Anal. Chem*. **2006**, *78*, 702-710.
152. Parks, B.A.; Jiang, L.; Thomas, P.M.; Wenger, C.D.; Roth, M.J.; Boyne, M.T.; Burke, P.V.; Kwast, K.E.; Kelleher, N.L. *Anal. Chem.* **2007**, *79*, 7984-7991.
153. Bunger, M.K.; Cargile, B.J.; Ngunjiri, A.; Bundy, J.L.; Stephenson, J.L. *Anal. Chem.* **2008**, *80*, 1459-1467.
154. Chi, A.; Bai, D.L.; Geer, L.Y.; Shabanowitz, J.; Hunt, D.F. *Int. J. Mass Spectrom*. **2007**, *259*, 197-20

Appendix I Measurement of Aflatoxin and Aflatoxin Metabolites in Urine by Liquid Chromatography Tandem Mass Spectrometry

In addition to the previously described work on bacteria, the toxins produced from the fungi *Aspergillus flavus*, known as aflatoxins, were also investigated. The subsequent material was reproduced from the *Journal of Analytical Toxicology* (*J. Anal. Toxicol*. **2007**, *31* (3), 150-156) by permission of Preston Publications, a division of Preston Industries, Inc. The figure, table and reference numbers apply only to this appendix.

Abstract

Automated immunoaffinity solid phase extraction followed by liquid chromatography/tandem mass spectrometry (LC/MS/MS) and chemical analogue internal standardization was employed to detect and quantify the aflatoxins AFB_1, AFB_2, AFG_1, AFG_2 and the metabolites AFM_1 and AFP_1 in urine. The dynamic range of the method is nearly three orders of magnitude with limits of detection in the low femtogram on column range. The method has been validated over a 12 day period by eight analysts. This method is suitable for agricultural, forensic and public health laboratories during an accidental outbreak or a chemical terrorism event where a rapid and accurate diagnosis of aflatoxicosis is needed.

Introduction

Due to the ever increasing threat of terrorist attacks around the globe and more specifically the threat of a chemical terrorism attack, analytical chemistry laboratories that would aid in forensic investigations and public health domains must be prepared to provide quality laboratory results quickly and efficiently. In such an event, the number of victims could be large and the type of warfare agent may not be immediately obvious. To this end, analytical methods that can provide rapid and sensitive confirmation and quantitation of the agent are vital in determining which agent was used, each individual's degree of exposure and the extent of the population that was exposed.[1]

Of the many toxins that could be used in an attack, those previously weaponized are of particular interest. One example is aflatoxins, which were weaponized by the Iraq government during the first Gulf War. According to a United Nations Special Commission (UNSCOM) report, the Salman Pak weapons facility in Iraq had produced 2,200 liters of aflatoxins loaded in 122 mm rockets, 400 pound bombs and SCUD missiles.[2]

Aflatoxins are secondary metabolites of the fungi *Aspergillus flavus* from which their name (A. fla.) is derived. The predominant aflatoxins AFB_1, AFB_2, AFG_1, and AFG_2 are designated B and G due to their blue and green fluorescent color observed under UV illumination. The most studied and most hepatotoxic aflatoxin is AFB_1 (LD_{50} 1.16 mg/kg in rat)[3] for which the World Health Organization (WHO) suggests there is no safe dose.[4] Moreover, these compounds are known to be mutagenic, and teratogenic. Clinical symptoms of aflatoxin exposure include abdominal pain, rash and

gastrointestinal bleeding.[5, 6] The commonality of these symptoms with those seen in other illnesses prevent them from providing unambiguous identification of their cause, which further emphasizes the need for an analytical method which provides more definitive information and enables a conclusive diagnosis.

The work presented here builds upon previous investigations of aflatoxins in urine[7, 8] in an effort to improve both the sensitivity and dynamic range of those methods. To improve upon the speed of previous methods, an automated immunoaffinity solid phase extraction method has been developed in conjunction with LC/MS/MS analysis to take advantage of the sensitivity, specificity, and ease of quantitation the technique provides. Figure 1 shows the structures for the four parent aflatoxins of interest and the metabolites of AFB_1 that were chosen for this study.

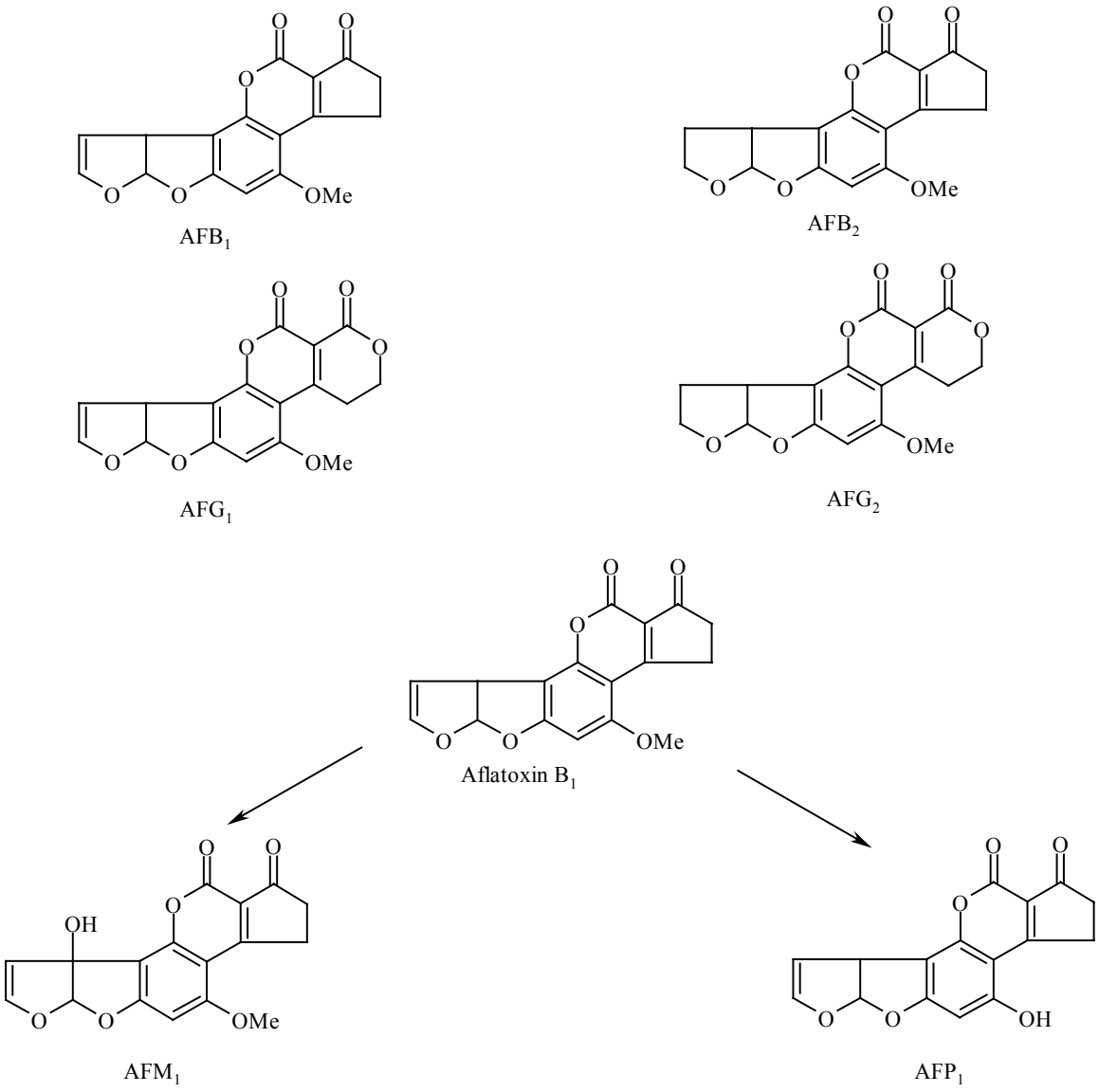


Figure 1. Structures of the four parent aflatoxins and the two metabolites of AFB_1 monitored in this study.

Experimental / Apparatus / Methods

Chemicals and Materials

HPLC grade methanol, acetonitrile and formic acid were purchased from Fisher Scientific (Fairlawn, NJ). Deionized water was purified in house to yield organic-free 18.3 MΩ·cm water using an E-pure purification system (Barnstead International Dubuque, IA). Aflatoxin reference standards (AFB_1, AFB_2, AFG_1, AFG_2, AFM_1 and AFP_1) were purchased from Sigma-Aldrich Chemical Company (St. Louis, MO).

Standard Preparation and Characterization

Aflatoxins and the internal standard AFB_2 were dissolved in acetonitrile and diluted in 85:15 MeOH:H_2O (v/v) to a final concentration of 1 ng/μL. Eight calibration standards, low and high quality controls (QC's), and a urine blank containing internal standard were stored in 50 mL polypropylene conical tubes at 4°C. Calibration standards were prepared in 1 mL of pooled human urine spiked with 25 μL of the stock solution of AFB_2. To all urine samples (unknowns, QC's, blanks, and standards) an equal amount of 85:15 MeOH:H_2O was added. For example, the highest concentration standard was made by mixing (per 1 mL of urine) 250 μL of the 1 ng/μL aflatoxin mixture, and 25 μL of the 1 ng/μL internal standard for a total volume of 1.275 mL. To the other samples the same amount of urine and internal standard were used but with varying amounts of aflatoxin standard and 85:15 MeOH:H_2O while maintaining a total volume of 1.275 mL. The final concentrations of the urine standards were: 0.392, 0.784, 3.92, 7.84, 19.6, 58.8, 118, and 196 ng/mL for the calibration standards, 1.96 and 157 ng/mL for the QC low and high respectively and the internal standard concentration in all samples was 19.6 ng/mL.

Extraction

Urine (1 mL) and water (1 mL) were added to glass tubes (10 x 75 mm), and empty tubes of the same size (for elution) were inserted into the Gilson 215 Liquid Handler (Middleton, WI) for automated extraction. Custom-made immunoaffinity columns, (3 mL barrel, 400ng aflatoxin equivalents binding capacity, Vicam, Watertown, MA) were used for the extraction. The buffer was discarded followed by rinsing the column twice with water, leaving a small amount of water on top of the resin. The columns were conditioned with water (2 x 2 mL). After the diluted urine sample was loaded, the column was washed with water (2 x 2 mL), followed by an air push (3 sec) using the Gilson solenoid valve. The analytes were eluted from the columns with 85:15 MeOH:H_2O with 1% v/v formic acid (2 x 0.5 mL), followed by an air push (30 sec). The extract was then transferred to an autosampler vial for LC/MS/MS analysis.

LC/MS/MS

Chromatography was performed with an Agilent 1100 HPLC (Wilmington, DE) equipped with a 3 μm 2.0 x 150 mm phenyl-hexyl column (Phenomenex, Torrance, CA) at 50ºC. 1 μL injections of the extract were made on the HPLC using a mobile phase consisting of H_2O (0.1% formic acid) (Solvent A) and acetonitrile (0.1% formic acid) (Solvent B). The mobile phase gradient is given in Table 1a.

Samples were analyzed by positive ion electrospray/tandem mass spectrometry operating in multiple reaction monitoring (MRM) mode on an API 4000 LC/MS/MS system (Applied Biosystems, Foster City, CA). The mass spectrometer settings are listed

in Table 1b. Individual compound specific parameters (i.e. declustering potentials, entrance potentials, and collision cell exit potentials) were optimized for each analyte.

Tables 1(a) and (b). Instrument parameters for LC (a) and MS (b).

Time (min)	Rate (mL/min)	%B
0	0.325	30
10	0.325	30
10.1	0.325	95
11.5	0.325	95
11.51	0.325	30
15.5	0.325	30

Analyte	Precursor	CE	Product
AFP_1	299	33.0	271
AFM_1	329	33.0	273
AFG_2	331	35.0	313
AFB_2	315	37.0	287
AFG_1	329	39.0	243
AFB_1	313	33.7	285

Data Analysis and Recovery

The product ion abundances of the analyte and internal standard were used to calculate analyte/internal standard ratios for quantitation. Linear regression analysis with "1/x" weighting was used for curve fitting. All data processing was performed automatically using Analyst 1.4 software (Applied Biosystems, Foster City, CA). Extraction efficiencies were calculated as a percentage of the ratio of extracted analyte peak area and the non-extracted standard peak area. Four extractions over a period of two days were performed for this study all at a concentration of 7.84 ng/mL. Four injections of non-extracted standard at an equal concentration were also made. The peak

areas for the extracted and non-extracted samples were averaged and this average was used to calculate the ratio for each analyte.

Animal Study

Urine samples were obtained from two male F344 rats (173-6 g body weight). AFB_1 (91 μg/kg body weight) or the vehicle (DMSO) were administered by intraperitoneal injection (150 μL) on two consecutive days and rats were housed in metabolic cages. Urine was collected for approximately 18 hours after the second dose and stored at -20°C. Urine aliquots (1 mL) were treated with 250 μL of 85:15 MeOH:H_2O and one mL of this mixture was then extracted. The animal study was conducted in accordance with John Hopkins University's Animal Care and Use Committee requirements which comply with the NRC's Guide for the Care and Use of Laboratory Animals.

Results and Discussion

The parent aflatoxins AFB_1, AFB_2, AFG_1 and AFG_2 and also AFM_1 and AFP_1, two phase I oxidative metabolites of AFB_1,[9] were selected for analysis. Both AFP_1 and AFM_1 are excreted in human urine[10] and studies have demonstrated that AFM_1 is the most abundant AFB_1 metabolite found in the urine of rats and humans.[11] Furthermore, research involving AFM_1 has shown that urinary AFM_1 levels reflect exposure in humans.[12] Metabolites of AFB_2, AFG_1 and AFG_2 were not included in this study because these three compounds are generally not observed in the absence of AFB_1 and AFB_1 is the most occurrent aflatoxin found in food.[13]

Urine (vs. blood or serum) was chosen as the sample matrix primarily due to its ease of collection. If victims are in reasonably stable condition, they can provide a urine sample without the assistance of a medical professional. Minimizing the workload of hospital staff is critical in a scenario such as a terrorist attack where hospitals will likely be overwhelmed.

AFB_2 was chosen as a chemical analogue internal standard because it behaves similarly to the other aflatoxins both in the immunoaffinity column (IAC) and in the analytical method. In addition, AFB_2 is chromatographically well-resolved from the other analytes and is less toxic than AFB_1 and AFG_1. Chemical analogue internal standardization was employed for two reasons. One, isotopically labeled standards are only available for AFB_2 and are prohibitively expensive (≈ $2000 for 8.3 μg AFB_2 ^3H). Two, the upper limit of linearity, ULOL, is determined by occupying the available binding sites of antibody in the column. The use of isotopically labeled internal standards for each aflatoxin would result in more aflatoxin being added to the immunoaffinity column which would lower the ULOL and therefore decrease the dynamic range of the method.

If AFB_2 were in a real world sample, a T-test could be performed to determine if the internal standard peak areas in unknown samples are significantly higher than the mean AFB_2 peak area in the calibration standards. If this were the case, then since all of the compounds of interest are equally suitable as chemical analogue internal standards, an aflatoxin not present in the sample would be added to a separate aliquot of unknown and then re-extracted. A second alternative would be to measure the concentration of AFB_2

by the method of standard additions. Both of these methods however would be difficult if the concentration of AFB_2 was at or near the limit of detection so a third alternative, analyzing each unknown in duplicate (one with internal standard, one without), could be employed.

Immunoaffinity extraction of aflatoxins in various matrices has been reviewed[14] and involves non-covalent binding of the toxins to monoclonal antibodies in aqueous environment followed by release upon denaturing of the antibodies using high organic content solvents. After a comparison with C18 solid phase extraction in our laboratory, the immunoaffinity method was chosen due to its increased recovery, selectivity and cleanliness of extracts. Since the extraction columns were originally manufactured for food analysis, the antibody used is designed to target only the parent aflatoxins; however, due to the structural similarities between the parent and metabolites, the metabolites are efficiently extracted as well. For this reason, the columns can readily be used for clinical samples.

Automated extraction was utilized as means to increase sample throughput. The extraction takes 1.67 minutes/sample corresponding to 36 samples/hour. The efficiency of the automated extraction was measured at a concentration of 7.84 ng/mL in urine. The extraction efficiencies for each of the five compounds are shown in Table 2. The results range from 80 – 93% and are in agreement with manufacturer specifications which are defined for food matrices.

The best chromatographic resolution was obtained using a phenyl-hexyl column. Figure 2 depicts the separation of all six aflatoxins used in this study extracted from

human urine at a concentration of 19.6 ng/mL. The MS/MS fragmentations of the aflatoxins are shown in Figure 3. It can be seen that both B aflatoxins and the AFB_1 metabolites fragment by losing the carbonyl on the cylcopentanone group and that AFM_1 further breaks down by losing the two adjacent CH_2 groups on the ring. The G aflatoxins differ by containing a lactone group in place of the pentanone ring structure. These compounds fragment by the loss of water and the further loss of the outer carbonyl group and adjacent portion of the ring, as well as the loss of H_2. The fragmentation patterns suggest that the site of protonation for each aflatoxin is the inner carbonyl group. These fragmentation patterns are based upon interpretation of the MS/MS spectra only as isotopically labeled experiments were not performed for reasons stated previously.

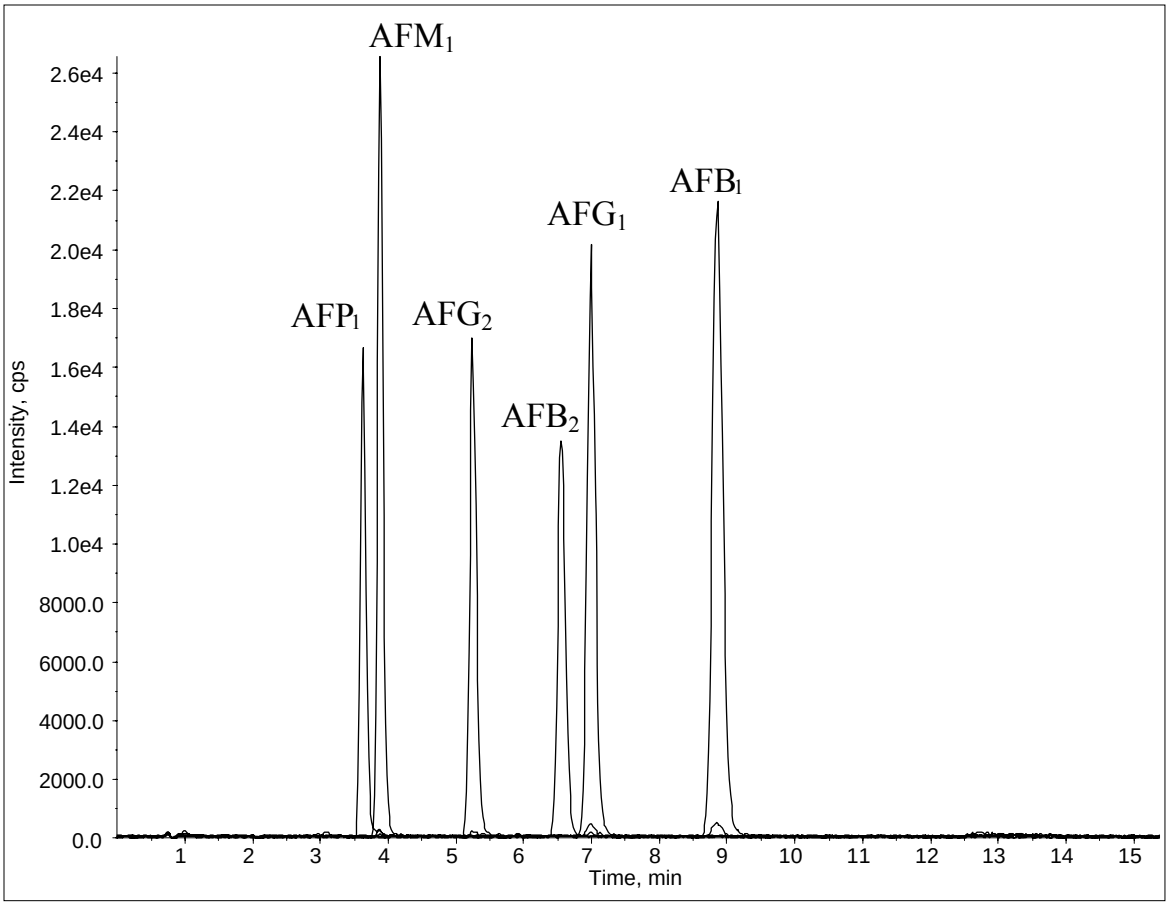


Figure 2. MRM extracted ion chromatogram of a 19.6 ng/mL urine extract using the transitions shown in Table 1b.

Figure 3a

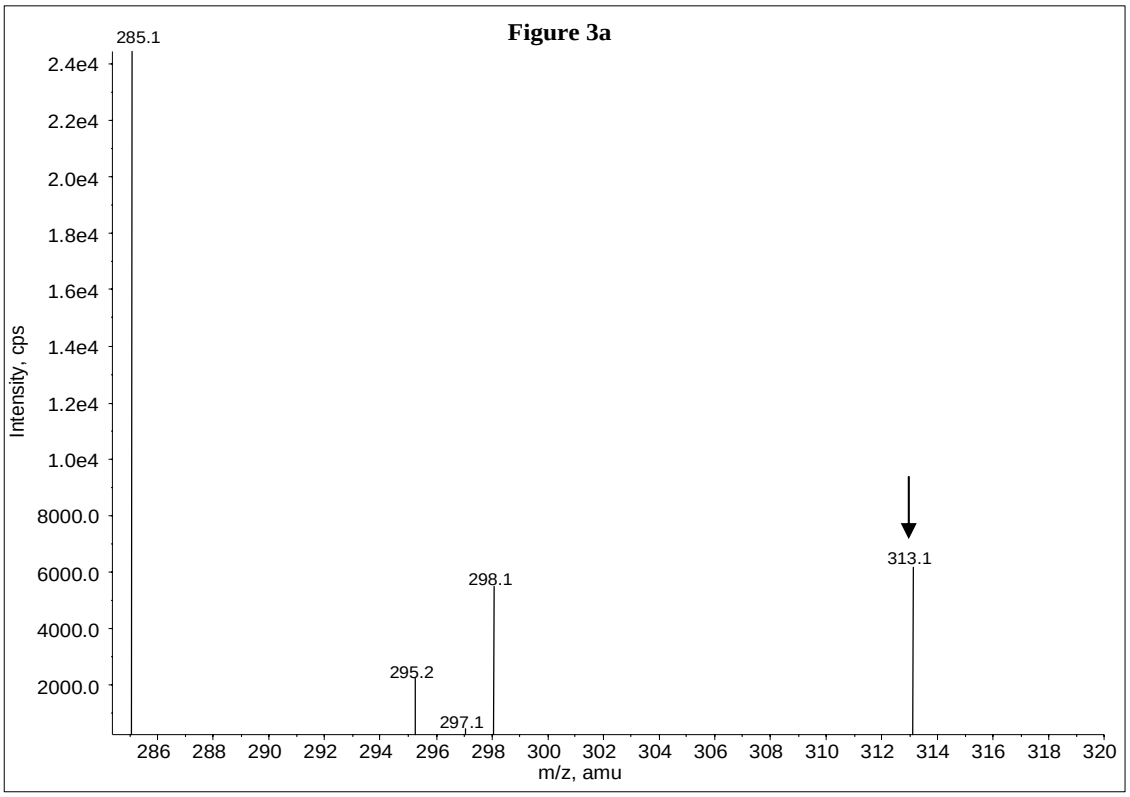


Figure 3b

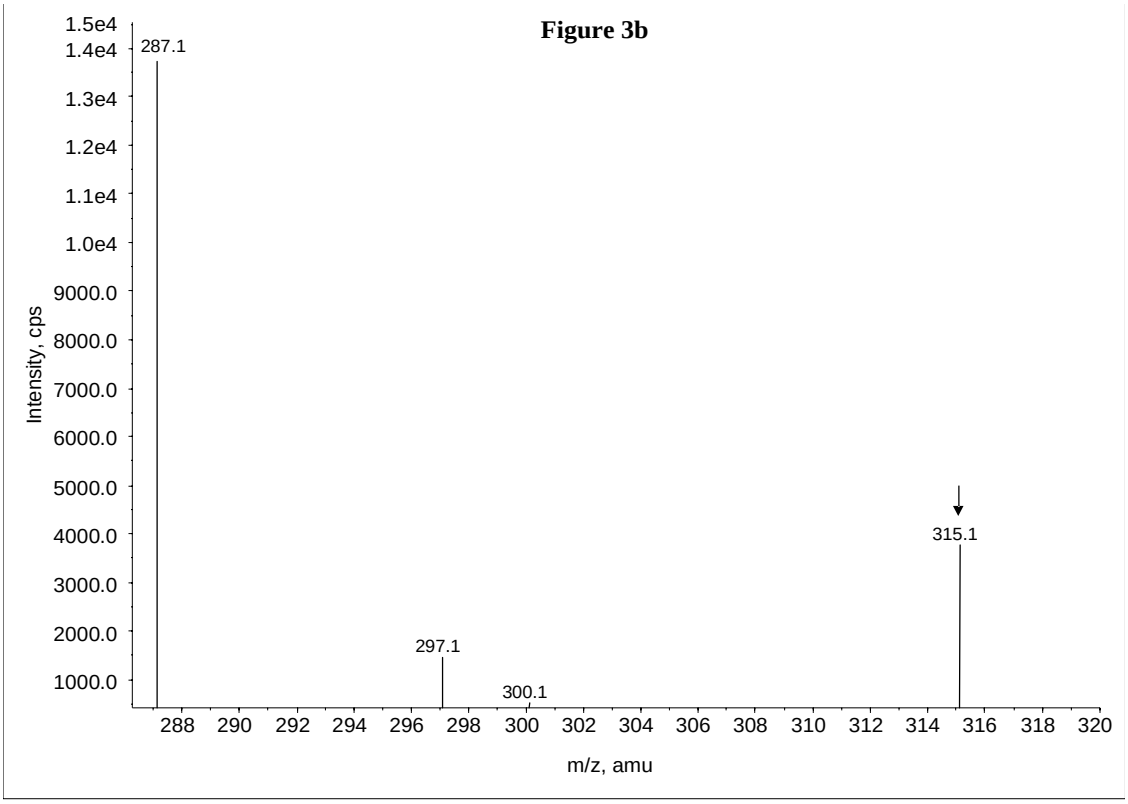


Figure 3c

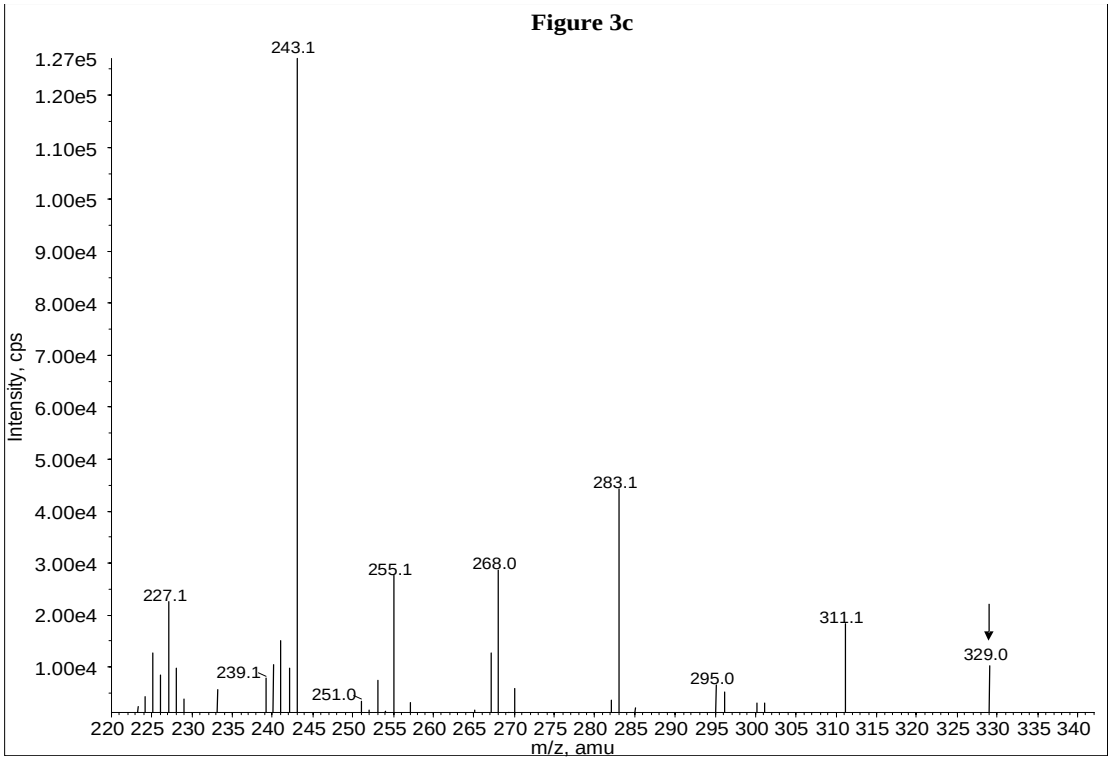


Figure 3d

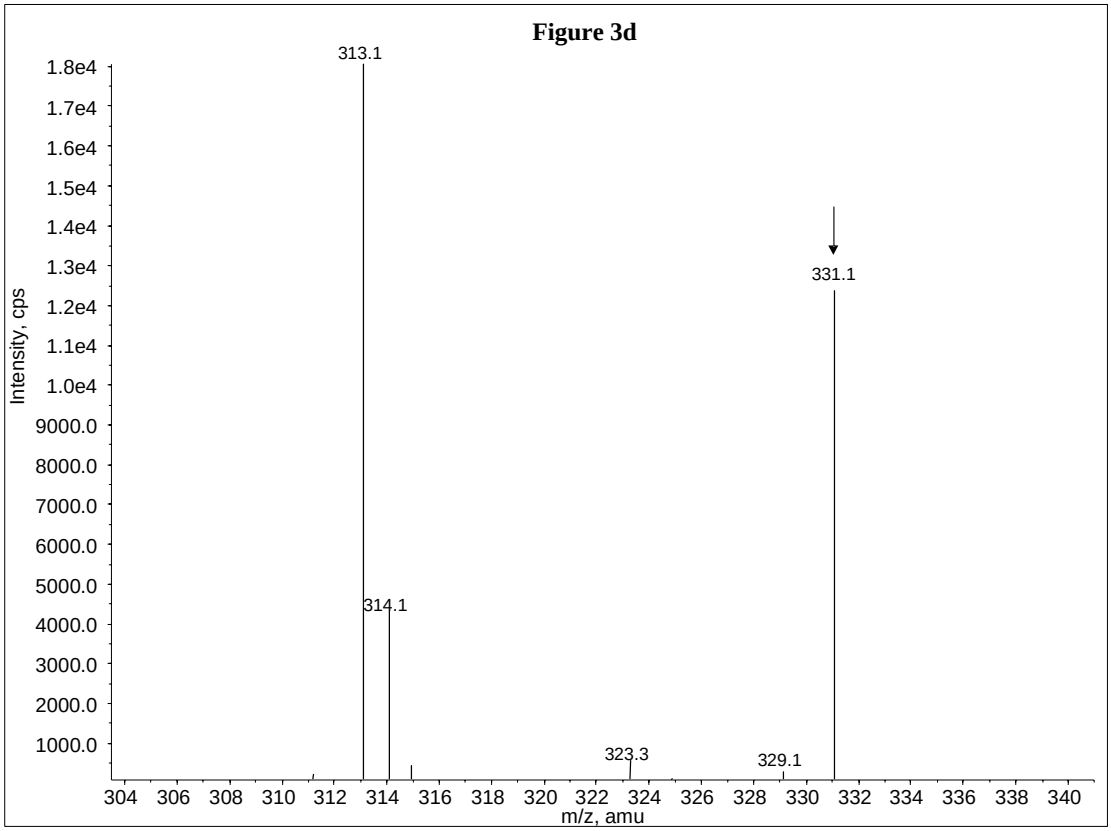
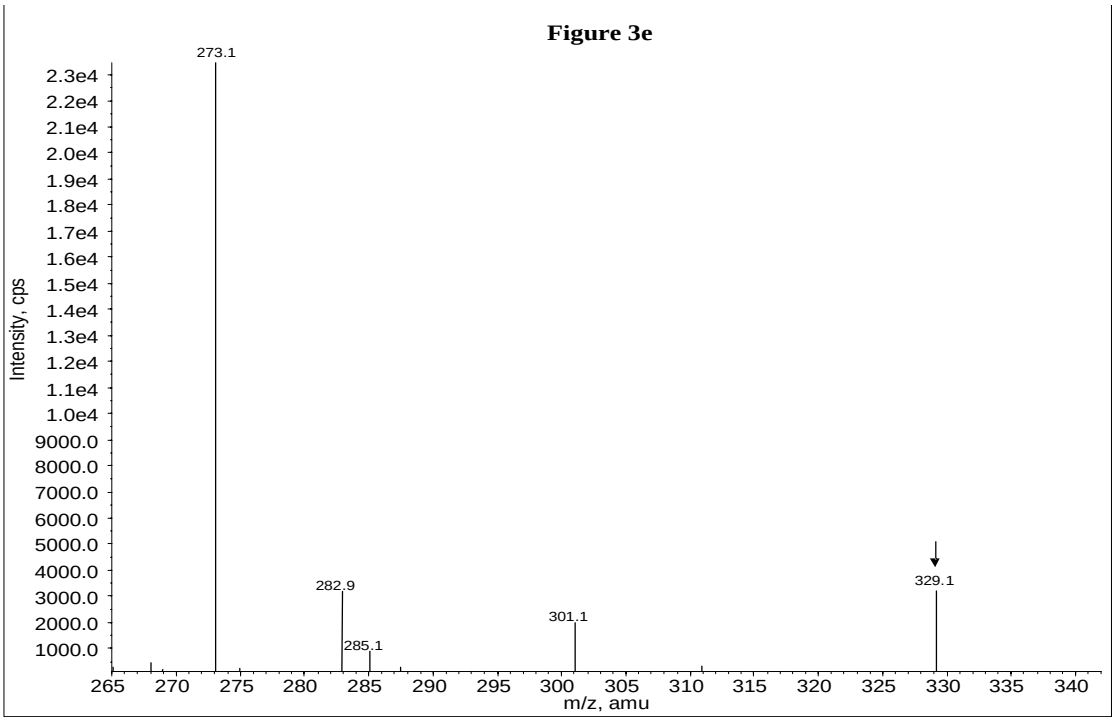


Figure 3e



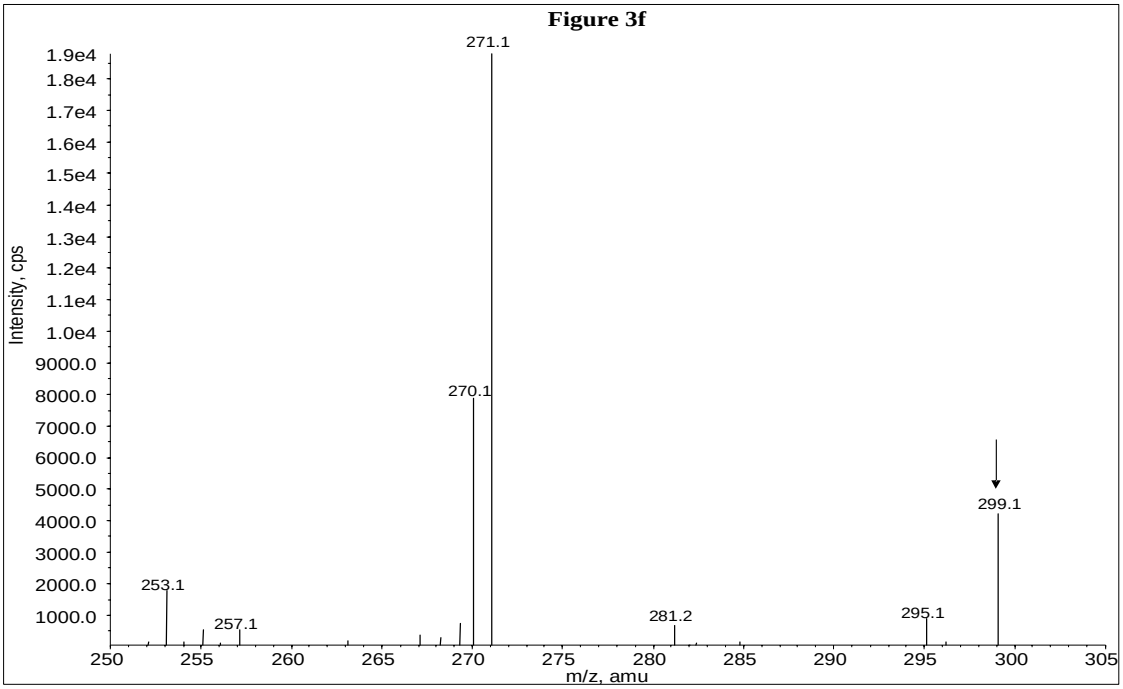


Figure 3. MS/MS fragmentation spectra for each aflatoxin with arrows indicating the precursor ion. (a) AFB_1, (b) AFB_2, (c) AFG_1, (d) AFG_2, (e) AFM_1 and (f) AFP_1.

The instrument limit of detection, LOD (S/N = 3/1) for each compound is given in Table 3. LOD values are reported in femtograms (fg) on column to avoid ambiguity inherent in using units of concentration such as parts per billion (ppb). This ambiguity makes it difficult to determine a relative LOD in papers where the pertinent information to calculate the LOD on column is not included.[15, 16] By providing the LOD as fg on column, the values reported in this manuscript are independent of the volume of urine extracted and/or injection volume and reflect the amount of sample that is detected by this method. Table 3 shows the LOD for each analyte in terms of grams on column and in moles. An LOD of 100 fg is equivalent to a 1 μL injection of non-extracted standard at 100 fg/μL. These values meet the goal of developing a sensitive method as these

detection limits are more than 10 times lower than those previously reported[17, 18] for aflatoxins in urine. The limit of quantitation, LOQ (S/N = 10/1), was calculated in urine using the same procedure stated above. The LOQ was determined to be 392 fg on column for all analytes.

Table 3. Limits of detection for each analyte defined as a S/N of 3/1.

Analyte	Femtograms on Column	Attomoles
AFP_1	50	168
AFM_1	100	305
AFG_2	100	302
AFG_1	100	305
AFB_1	50	160

A wide dynamic range is needed in the analysis of chemical warfare agents for two reasons. One, since there are no studies in the literature showing the range of concentrations in a victim exposed to weaponized aflatoxin, having a dynamic range of nearly three orders of magnitude increases the likelihood that this method will cover relevant concentrations. Second, the exposure of victims after an event will not be uniform due to their different proximities to the attack epicenter. To achieve this goal, custom-made extraction columns with a lager bed size were employed. The dynamic range of the method was from the LOQ at 0.392 pg – 196 pg on column for each analyte.

Procedures that require time consuming concentration steps which may allow for the detection of lower aflatoxin levels were avoided to increase assay throughput (i.e. one mL of urine is extracted into one mL of MeOH:H_2O with 1% v/v formic acid). Figure 4 shows a calibration curve for all five toxins in human urine. The dynamic range presented here meets the aforementioned goal of covering a wider range than those previously reported.[8,17]

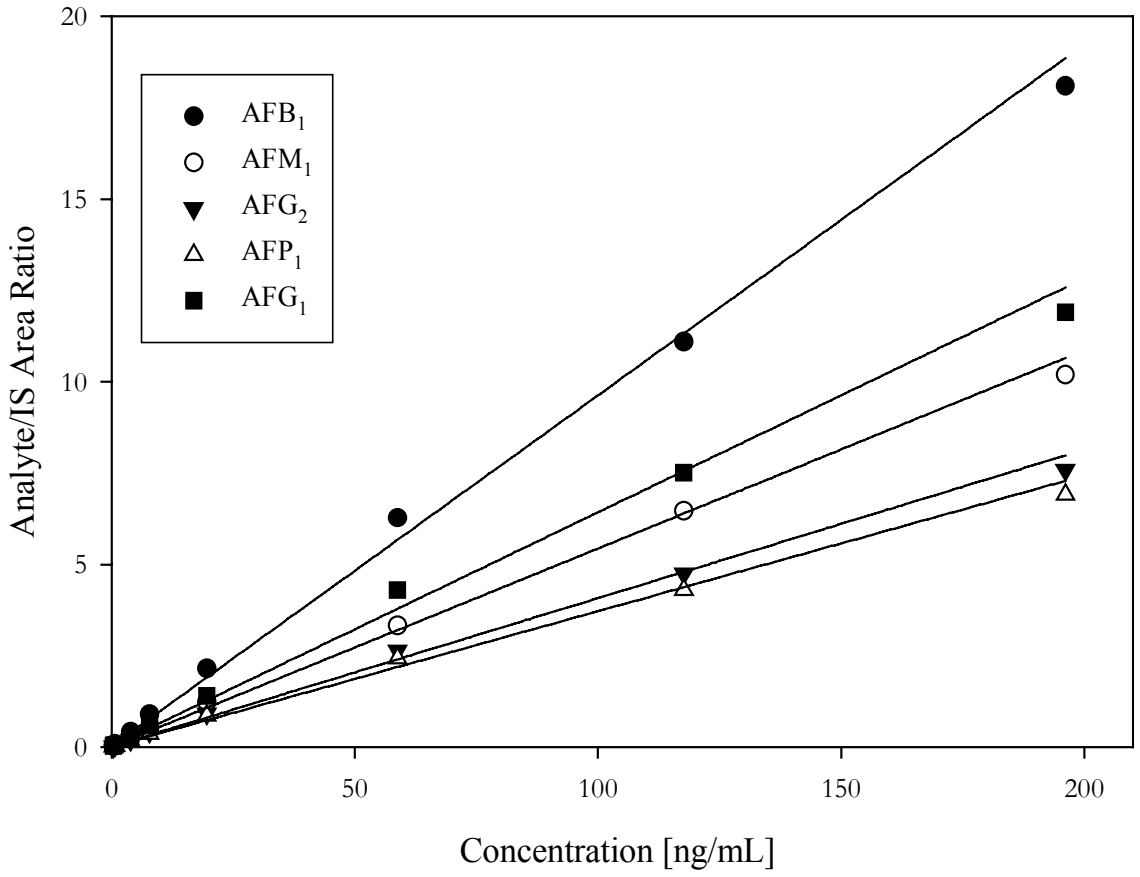


Figure 4. An eight point calibration plot from 0.392 ng/mL to 196 ng/mL in urine using linear regression with "1/x" weighting.

As seen in Figure 4, a slight deviation from linearity begins to occur at the highest point. This is due to the large aflatoxin/antibody ratio at this concentration. This was examined by analyzing non-extracted standards which showed increased linearity at this same concentration and at even higher concentrations (data not presented) which verified that our ULOL is dictated by the binding capacity of the extraction column and not by the instrument. At concentrations above the ULOL, the number of binding sites becomes a limiting factor and some unbound aflatoxin is removed during the wash step of the extraction. This is an important consideration when determining the ULOL of an extraction method using immunoaffinity columns.

The method was validated by analyzing a calibration curve, two quality controls (low and high) and a urine blank spiked with internal standard. This experiment was repeated twenty times over a period of 12 days with no more than two sets being analyzed in a single day. Eight analysts conducted the experiments during the 12 day period. Linear regression with "1/x" weighting was used for each analyte to account for heteroscedasticity in the data. The calibration standards were analyzed in a random order and the curves for each aflatoxin had an average correlation coefficient ≥ 0.995. The quality of the method is represented by quality control low and high plots of AFG_1 (Figure 5). Similar results were obtained for the four other toxins. The results of the accuracy and precision of the method for each analyte are shown in Table 4. The range of the percent accuracy of the means for all five analytes is from 97.0 – 105.6% and the highest %RSD found was 6.67%. In all but one case the mean is $\leq$ one standard

deviation away from the true value. No significant contributions from carryover were seen in the blank samples.

Figure 5a

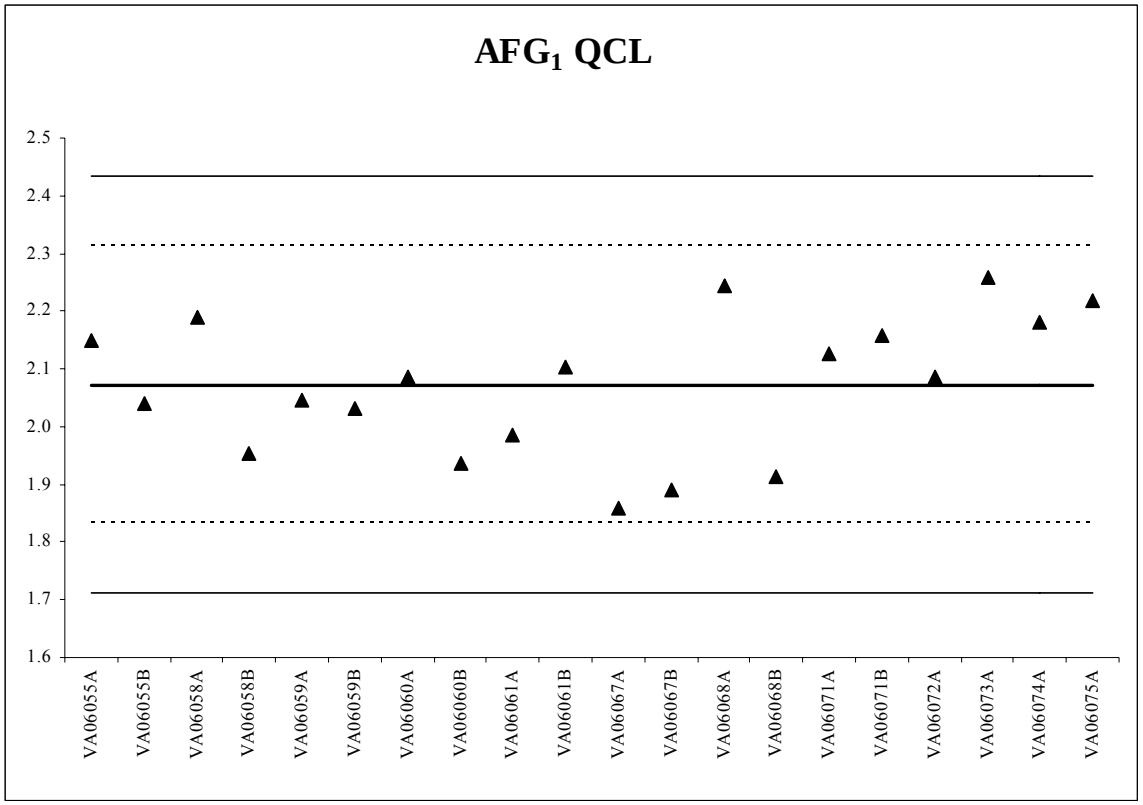


Figure 5b

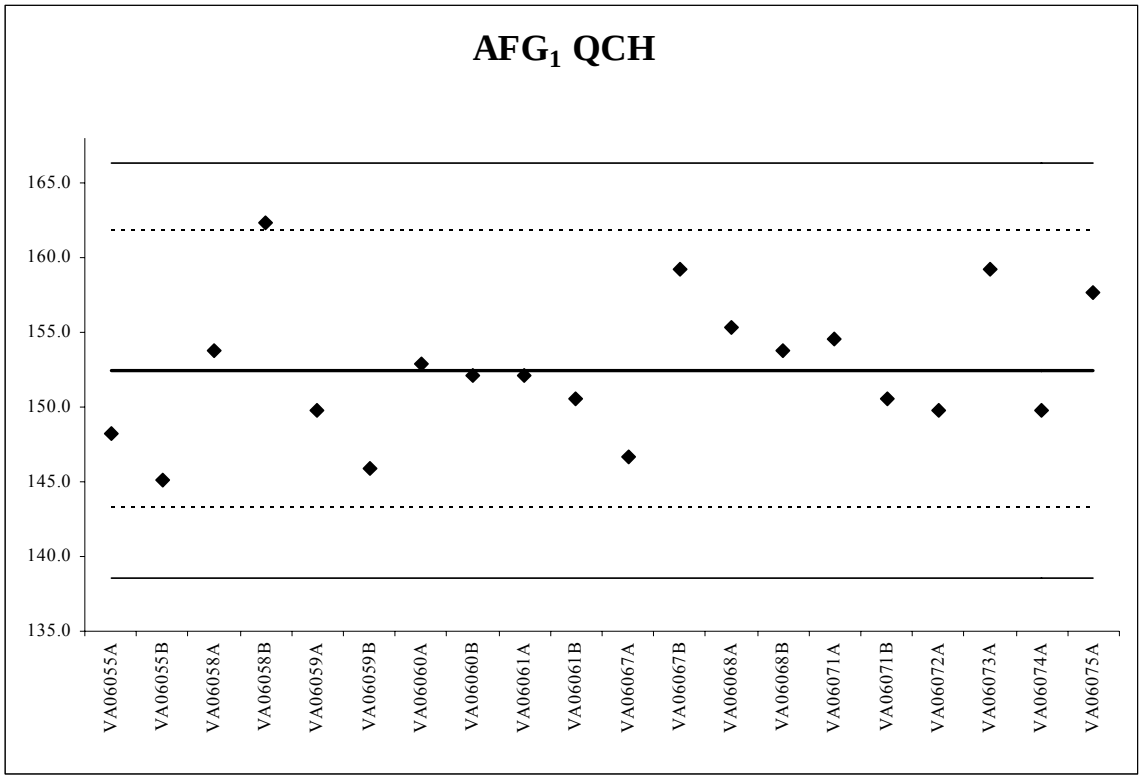


Figure 5. Quality control plots for AFG_1, (a) QC Low at 1.96 ng/mL and (b) QC High at 157 ng/mL. The central line represents the mean, the dashed line is 2 standard deviations and the outer line is 3 standard deviations from the mean.

Table 4. Validation results depicting the precision and accuracy of the method.

Analyte	Concentration (ng/mL)	Mean (ng/mL)	% Accuracy	Lower 95% limit (ng/mL)	Upper 95% limit (ng/mL)	% RSD
AFB_1	1.96	2.01	103	1.95	2.08	6.67
AFB_1	157	153	97.5	149	157	5.18
AFM_1	1.96	2.01	102	1.95	2.07	6.59
AFM_1	157	152	97.0	148	156	5.58
AFG_1	1.96	2.07	106	2.02	2.13	5.68
AFG_1	157	152	97.0	150	155	3.04
AFP_1	1.96	2.03	104	1.97	2.09	6.30
AFP_1	157	151	96.5	147	155	5.16
AFG_2	1.96	2.06	105	2.02	2.11	4.79
AFG_2	157	146	93.5	144	150	5.07

After validation, the method was further tested by analyzing the urine of an AFB_1 exposed rat. The dose administered to the rat (91 μg/kg body weight) is well below the LD_{50} for rats and corresponds to a dose of several milligrams for a human adult. A dose of 2-6 mg/day was observed during an outbreak of aflatoxicosis in western India,[19] and similar amounts were consumed in a recent outbreak in Kenya.[20] It was expected that roughly equal amounts of the two metabolites AFM_1 and AFP_1 and a small amount of unmetabolized AFB_1 may be excreted. No aflatoxins other than the internal standard were detected in the negative control rat urine sample. AFB_1, AFM_1 and AFP_1 were respectively detected in urine from the AFB_1 dosed rat at 1.38, 48.8 and 41.4 ng/mL. Four mL of urine was collected indicating a total excretion of 5.52, 195.2 and 165.6 ng during the 18 hour period. The creatinine levels of the two urines were not measured,

thus no comparison of the metabolite concentrations with literature values were made. However, relative to AFP_1, the amount of AFM_1 was more abundant which is consistent with previous studies of AFB_1 exposure in rats.[11, 21] Figure 6 shows the chromatograms for the negative control and AFB_1 dosed rat urine samples.

Figure 6a

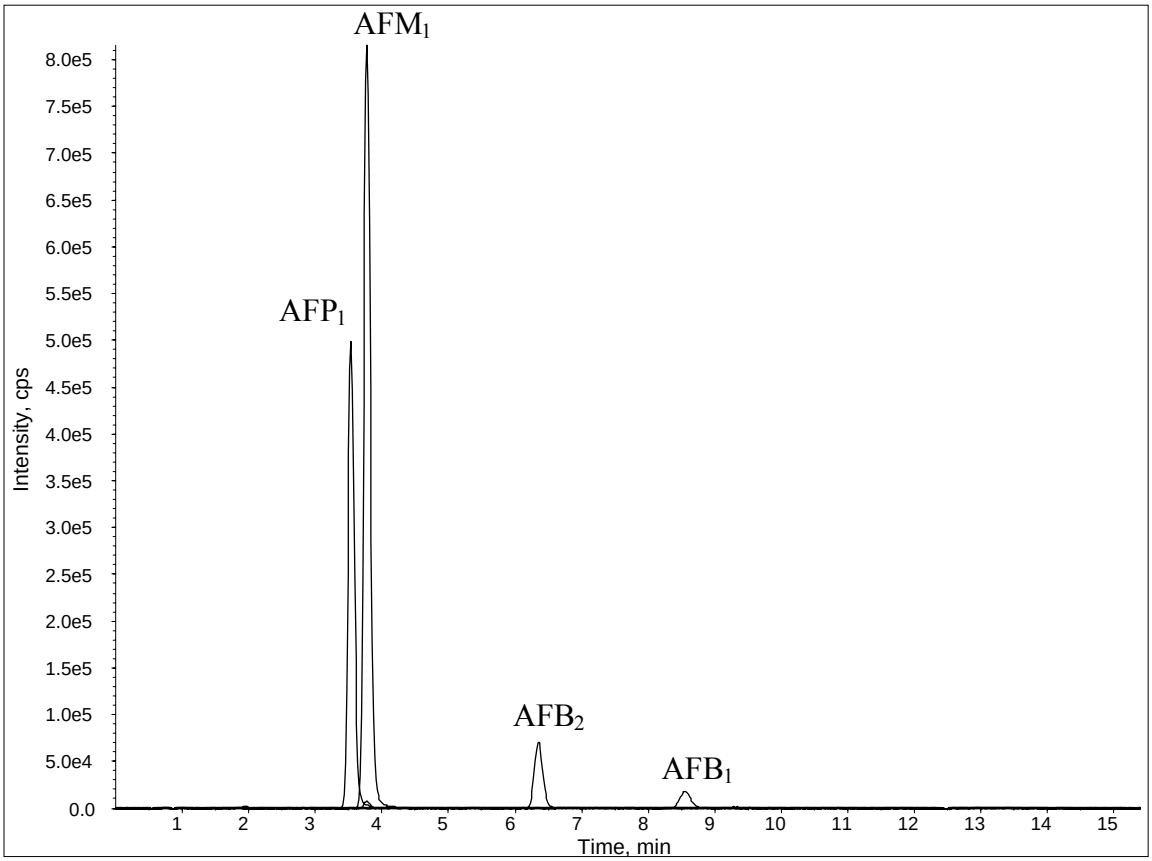


Figure 6b

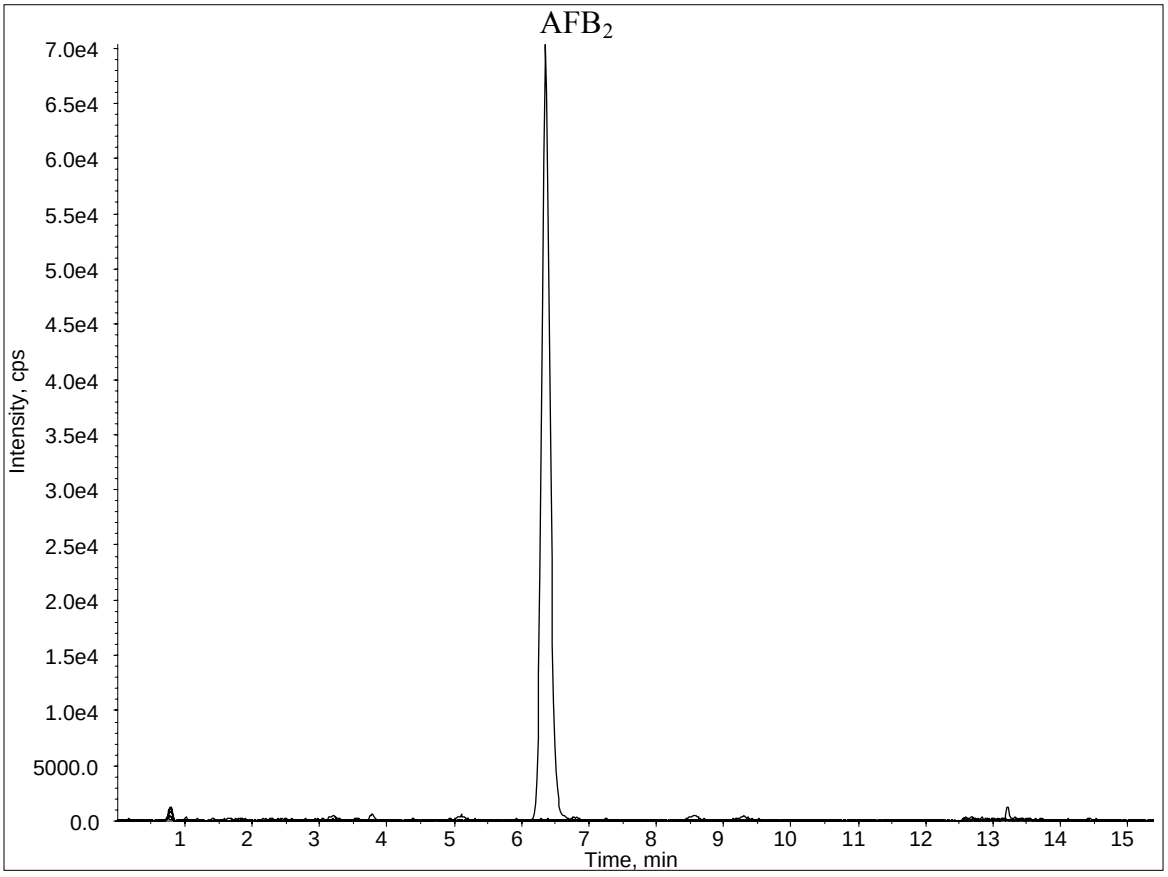


Figure 6. Chromatograms from the animal study, (a) dosed rat and (b) control rat.

It was also expected that other metabolites of aflatoxin would be present in the dosed rat urine. Precursor and product ion masses of other AFB_1 metabolites[18] were used to build a separate MRM method to detect the presence of these compounds. Metabolites such as AFB-diol and AFQ_1 were found. The presence of AFQ_1 was further confirmed by matching the retention time with a chromatogram obtained previously when AFQ_1 was commercially available (data not presented). This method is suitable for monitoring

other metabolites on a qualitative level, but quantitative analysis is difficult since AFM_1 and AFP_1 are the only commercially available AFB_1 metabolite standards.

Conclusion

The LC/MS/MS method described uses less urine and has lower limits of detection than previously reported methods. Other advantages of this approach are the specificity of both immunoaffinity extraction and tandem mass spectrometry, and no requirement for derivitization. The dynamic range of 0.392 – 196 pg on column combined with an automated extraction yield rapid results over a wide range of exposure. This method is well suited to aid forensic and public health laboratories during the investigation of a terrorist attack by providing confirmation of military/civilian exposure to weaponized aflatoxin. In addition, the method presented here can be utilized to diagnose aflatoxicosis[22] caused by consumption of a food supply contaminated intentionally by the hands of terrorists or accidentally during an outbreak.[19, 23, 24]

References

1. Barr, J.R. *J. Anal. Toxicol.* **2004**, *28*, 305.
2. Zilinskas, R.A. *J. Am. Med. Assoc.* **1997**, *278*, 418–424.
3. Wogan, G.N.; Edwards, G.S.; Newberne, P.M. *Cancer Res.* **1971**, *31*, 1936-1942.
4. Anklam, E.; Stroka, J.; Boenke, A. *Food Control* **2002**, *13*, 173-183.
5. Cullen, J.M.; Newberne, P.M. In *The Toxicology of Aflatoxins: Human Health, Veterinary, and Agricultural Significance*; Eaton, D.L.; Groopman, J.D. Eds. Academic Press: San Diego, 1994, pp 3-26.
6. Willis, R.M.; Mulvihill, J.J.; Hoofnagle, J.H. *Lancet* **1980**, *315*, 1198-1199.
7. Kussak, A.; Andersson, B.; Andersson, K.; Nilsson, C. *Chemosphere* **1998**, *36*, 1841-1848.
8. Vazquez, B.I.; Fente, C.A.; Franco, C.M.; Cepeda, A.; Mahuzier, G.; Prognon, P. *Anal. Commun.* **1999**, *36*, 5-7.
9. Eaton, D.L.; Ramsdell, H.S.; Neal, G.E. In *The Toxicology of Aflatoxins: Human Health, Veterinary, and Agricultural Significance*; Eaton, D.L.; Groopman, J.D. Eds. Academic Press: San Diego, 1994, pp 45-72.

10. Groopman, J.D.; Donahue, P.R.; Zhu, J.Q.; Chen, J.S.; Wogan, G.N. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6492-6496.
11. Hsieh, D.P.H.; Wong, J.J. In *The Toxicology of Aflatoxins: Human Health, Veterinary, and Agricultural Significance*; Eaton, D.L.; Groopman, J.D. Eds. Academic Press: San Diego, 1994, pp 73-88.
12. Zhu, J.Q.; Zhang, L.S.; Hu X.; Xiao, Y.; Chen, J.S.; Xu, Y.C.; Fremy, J.; Chu, F.S. *Cancer Res.* **1987**, *47*, 1848-1852.
13. Ritchie, J.C. In *Molecules of Death*; Waring, R.H., Steventon, G.B.; Mitchell, S.C. Eds. Imperial College Press: London, 2002, pp 1-18.
14. Scott, P.M.; Trucksess, M.W. *J. Assoc. Off. Anal. Chem.* **1997**, *80*, 941-949.
15. Wang, J.S.; Shen, X.; He, X.; Zhu, Y.; Zhang, B.; Wang, J.; Quian, G.S.; Kuang, S.; Zarba, A.; Egner, P.A.; Jacobson, L.P.; Munoz, A.; Helzlsouer, K.J; Groopman, J.D.; Kensler, T.W. *J. Natl. Cancer Inst.* **1999**, *91*, 347-354.
16. Mykkanen, H.; Zhu, H.; Salminen, E.; Juvonen, RO.; Ling, W.; Ma, J.; Polychronaki, N.; Kemilainen, H.; Mykkanen, O.; Salminen, S.; El-Nezami, H. *Int. J. Cancer* **2005**, *115*, 879-884.
17. Kussak, A.; Nilsson, C.; Andersson, B.; Langridge J. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1234-1237.
18. Walton, M.; Egner, P.; Scholl, P.F.; Walker, J.; Kensler, T.W.; Groopman, J.D.; *Chem. Res. Toxicol.* **2001**, *14*, 919-926.
19. Krishnamachari, K.A.V.R.; Nagarajan, V.; Bhat, R.V.; Tilak, T.B.G. *Lancet* **1975**, *305*,1061-1063.
20. Azziz-Baumgartner, E.; Lindblade, K.; Gieseker, K.; Rogers, H.S.; Kieszak, S.; Njapau, H.; Schleicher, R.; McCoy, L.F.; Misore, A.; DeCock, K.; Rubin, C.; Slutsker, L.; the Aflatoxin Investigative Group. *Environ. Health Perspect.* **2005**, *113*, 1779-1783.
21. Scholl, P.; Musser, S.M.; Kensler, T.W.; Groopman J.D. *Carcinogenesis* **1996**, *17*, 1385-1388.
22. Williams, J.H.; Phillips, T.D.; Jolly, P.E.; Stiles, J.K.; Jolly, C.M.; Aggarwal, D. *Am. J. Clin. Nutr.* **2004**, *80*, 1106-1122.
23. Nyikal, J.; Misore, A.; Nzioka, C. Njuguna, C.; Muchiri, E.; Njau, J.; Maingi, S.; Njoroge, J.; Mutiso, J.; Onteri, J.; Langat, A.; Kilei, I.K.; Nyamongo, J.; Ogana, G.; Muture, B.; Lewis, L.; Luber, G.; Rogers, H.; Backer, L.; Rubin, C. *MMWR Weekly* **2004**, *53*, 790-793.
24. Serck-Hanssen, A. *Arch. Environ. Health* **1970**, *20*, 729-731.

VITA

Robert Anthony Everley was born on November 9, 1980 in Hartford, Kentucky. He graduated from Ohio County High School in Hartford in 1999, and then attended Eastern Kentucky University where he obtained his Bachelor of Science degree in forensic science in 2003. That same year, he began graduate studies in the Department of Chemistry at Virginia Commonwealth University. Robert has co-authored six publications and has presented his research findings at numerous scientific meetings.

Publications:

Robert A. Everley, Frederic L. Ciner, Dongliang Zhan, Peter Scholl, John Groopman and Timothy R. Croley, "Measurement of Aflatoxin and Aflatoxin Metabolites in Urine by Liquid Chromatography Tandem Mass Spectrometry," *J. Anal. Toxicol.* 2007, 31: 150-156.

Robert A. Everley, Tiffany M. Mott, Shane A. Wyatt, Denise M. Toney, and Timothy R. Croley, "Liquid Chromatography/Mass Spectrometry Characterization of *Escherichia coli* and *Shigella* Species," Submitted to *JASMS*.

Robert A. Everley and Timothy R. Croley, "Ultra Performance Liquid Chromatography/Mass Spectrometry of Intact Proteins," *J. Chrom. A*, in press.

Tiffany M. Mott, Robert A. Everley, Shane A. Wyatt, Denise M. Toney, and Timothy R. Croley, "Evaluation of MALDI-TOF/MS and LC-QTOF/MS Methods for the Identification of Enteric Bacteria," Submitted to *Appl. Environ. Microbiol*.

Michelle T. Sheldon, Robert A. Everley, Frederic L. Ciner, Ellen Pace, Jack Henion, Robert Mistrik and Timothy R. Croley, "Chip-based LC/MS^n with Precursor Ion Fingerprinting for Unknown Compound Identification," In preparation.

Robert A. Everley, Tiffany M. Mott, Denise M. Toney, and Timothy R. Croley, "Characterization of *Clostridium* Species Utilizing Liquid Chromatography/Mass Spectrometry of Intact Proteins," In preparation.

Presentations:

Robert A. Everley, Frederic L. Ciner, Dongliang Zhan and Timothy R. Croley "Quantitative Analysis of Aflatoxin and Aflatoxin Metabolites using LC/MS/MS…The Past, Present and Future." Chemical Terrorism Laboratory Response Network Meeting, November 15 – 17, 2005, Atlanta, GA.

Robert A. Everley, Emily Hopkins, Dee Pettit and Timothy R. Croley, "LC/MS Analysis of Potential Bioterrorism Agents," American Society for Mass Spectrometry Fall Workshop: Characterization of Microorganisms by Mass Spectrometry, San Diego, California December 8 – 9, 2005.

Robert A. Everley, Frederic L. Ciner, Dongliang Zhan and Timothy R. Croley "Determination of Aflatoxin and Aflatoxin Metabolites in Urine using Automated Immunoaffinity SPE and LC/MS/MS," Chemical Terrorism Laboratory Response Network Meeting, April 18 – 20, 2006, Richmond, VA.

Robert A. Everley, Frederic L. Ciner, Dongliang Zhan and Timothy R. Croley, "High-Throughput Quantitative Analysis of Aflatoxin Exposure by LC/MS/MS," American Society for Mass Spectrometry and Allied Topics, Seattle, Washington May 28 – June 1, 2006.

Jack Henion, Ellen Pace, Amie Prince, Gary Schultz, Robert A. Everley, Ricky Ciner, Michelle Sheldon, Tim Croley and Robert Mistrik, "Possible Strategy for ID of Unknowns: Metabolite ID Employing LC/MS, Fraction Collection and Infusion MS^3 with Modern Bioinformatics" and, "Qualitative Analysis: LC-MS/MS with Simultaneous LC Fraction Collection Maximizing MS Information Of Complex Samples With Nanoelectrospray Ionization," FACSS, Lake Buena Vista, Florida September 24-28, 2006.

Tiffany M. Mott, Robert A. Everley, Shane A. Wyatt, Denise M. Toney and Timothy R. Croley, "Evaluation of MALDI-TOF/MS and LC/MS Methods for the Identification of Enteric Bacteria," American Society for Microbiology Biodefense and Emerging Diseases Research Meeting, Washington, DC February 27- March 2, 2007.

Tiffany M. Mott, Robert A. Everley, Shane A. Wyatt, Denise M. Toney and Timothy R. Croley, "Evaluation of MALDI-TOF/MS and LC/MS Methods for the Identification of *Clostridium* species," American Society for Microbiology, Toronto ON May 21 – 25, 2007.

Frederic L. Ciner, Michelle T. Sheldon, Robert A. Everley, Gary Schulz, Ellen Pace, Jack Henion, Robert Mistrik and Timothy R. Croley, "Automated Identification of Unknown Compounds Utilizing Nanoelectrospray LC/MS Coupled with Infusion MS^n Employing the Concept of Precursor Ion Fingerprinting," American Society for Mass Spectrometry and Allied Topics, Indianapolis, Indiana June 3 – June 7, 2007.

Robert A. Everley, Tiffany M. Mott, Shane A. Wyatt, Peter Leopold, and Timothy R. Croley, "Biomarkers for Bioweapons: An Automated LC/MS Approach," American Society for Mass Spectrometry and Allied Topics, Indianapolis, Indiana June 3 – June 7, 2007.

Robert A. Everley and Timothy R. Croley, "Ultra Performance Liquid Chromatography/Mass Spectrometry (UPLC/MS) of Intact Proteins," 24th Montreux Symposium on LC/MS & Related Techniques, Hilton Head Island, South Carolina October 10-12, 2007.

Robert A. Everley, Tiffany M. Mott, Denise M. Toney, and Timothy R. Croley, "Characterization of *Clostridium* Species Utilizing Liquid Chromatography/Mass Spectrometry of Intact Proteins," American Society for Mass Spectrometry and Allied Topics, Denver, Colorado June 1 – June 5, 2008 (abstract accepted).