MALDI-TOFMS BASED PROTEIN PROFILING AS A DIAGNOSTIC TOOL FOR THE ANALYSIS OF *Bacillus* SPORES AND CELLS

By

DANIELLE NICOLE DICKINSON

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

UNIVERSITY OF FLORIDA

2004

For my inspiration.
My grandfather, the smartest man I have ever known. You are my light and my strength, and are always in my heart. Shine on.

And for my biggest fan.
My mother, for everything she has sacrificed and for making me into who I am. I can't thank you enough.

"I am not bound to win, but I am bound to be true. I am not bound to succeed, but I am bound to live up to what light I have"—Abraham Lincoln

ACKNOWLEDGMENTS

It is difficult to know where to begin to thank all of the people who made my graduate career worthwhile and helped to make my dissertation project possible. I ask them to forgive me ahead of time for the length and breadth of this list but I want an opportunity to thank everyone.

I must start first with the three whom I consider my cochairs, regardless of whether UF recognizes them as such: Dr. James D. Winefordner, Dr. David H. Powell, and Dr. Kasthuri Venkateswaran.

I thank Dr. Winefordner for his faith and courage in taking me under his wing and allowing me to go "willy-nilly" on a project that was somewhat beyond the scope of his repertoire of work. His enthusiasm and love for science I will forever try to imitate. I thank Dr. Powell, aka the leader of the mass spectrometry orphanage. His patience, love, kindness, and guidance will never be forgotten, and neither will "dark matter." I thank him for allowing me the privilege to work in his facility and with his instrumentation. I don't know if he realizes what a true friend and mentor he is for us in both science and life. And I do apologize for the ambulance ride I took him on. I don't think many advisors have had to endure a ride like that! In thanking Venkat, I don't even know where to begin! This project came as far as it did thanks to him. I don't know if he even realizes its impact in the mass spectrometry arena and without his microbial guidance and spores it would have never happened. I thank him for having the patience to turn an

analytical chemist into a microbiologist, for never giving up on me, and for the constant supply of wit and banter.

I must pause to thank the funding sources, scarce as they seemed sometimes! I thank the NASA Graduate Student Research Program for 3 years of support on this project; and the Planetary Protection and Biotechnology Group at the Jet Propulsion Laboratory for all the funding they provided.

I owe much gratitude the people in the Biotechnology and Planetary Protection Group at the Jet Propulsion Laboratory and the contacts I have made through them. I thank Karen Buxbaum for having the faith to bring me out again after a rather interesting first round. I thank Wayne, Roger, Cecilia, Shirley, and Gayan for their help and guidance the first summer I was there. I thank Mike Kempf for providing some of the first spore samples we worked on and for his help in learning the ropes of making spores myself. Most especially I thank Myron LaDuc, a guy with some of the most interesting personality quirks I have ever encountered. Turning an analytical chemist into a microbiologist was not an easy task, but he did it patiently, as long as we could listen to Dave and not my "hillbilly" music. I am honored to call him a colleague and friend. I thank Dr. Adam Driks at Loyola University for the spore coat mutants of *B. subtilis* and for being such a great wealth of knowledge on spore coats and sporulation procedures. My last trip to California would also not have been possible without a free place to camp out for 8 weeks. I thank 'Nay and Alicia for opening their home to me and for taking me out on the town. It was surreal and crazy to live in "Hollywood" for a time. I especially thank them for not being too scared when "the geek" was wandering around the house muttering things to herself!

At Kennedy Space Center (KSC), I also send thanks to Drs. Ray Wheeler, Jay Garland, Mike Roberts, and Langfang Levine, who were always willing to listen, give advice, and get me passes onto KSC where I got to run around like a kid in a candy store (oh I don't think they were supposed to know that).

Back on the UF front, there are also a lot of great people and resources that deserve an abundance of praise. First and foremost I thank Kenny, aka the Bruker Reflex II. He might be old, but he works like a charm when he's in a good mood, and the picture of Kevin in the crown is present. Long live the BK crown! I thank the electronics and machine shop guys, who are incredible. In the end, they are the only reason Kenny stayed in a good mood for any length of time. I apologize to Steve for still being unable to translate the schematics. Big thanks go to all the support staff (Bev, Maribel, Beth, Lisa, Gracie, Jill, Jim, Joe, Darrius, and Matt) in the department for their help and encouragement along the way. They are the backbone that keeps this place going. I cannot forget my three favorites, Jeanne, Lori, and Romaine. On many days that I don't know what I would have done without their help and laughter. They have kept me going and tried their best to keep me on the straight, narrow, and focused path (and yes I know I am not good at it). I thank them tremendously for it all.

On the technical side, I had a great deal of assistance at UF as well. Dr. Jodie Johnson, Dr. William Haskins, and Regina Wolper helped with the HPLC analysis. Igor Gornushkin developed the linear correlation software and library searching algorithms that were used. Scott and Stan in the protein core helped me get through my lengthy database searches. Dr. Denslow offered extremely helpful discussions on proteins and allowed me to use the instrumentation and database search tools in the ICBR. I thank

Danielle Anderson and Qian Li for taking some SEM images of the spores. Stephanie, Greg, and Nicole offered helpful discussion on microbiology and gel electrophoresis. Dr. Rasche, initially took on the task of turning me into a microbiologist and provided me with an incubator I could use in my lab for cell growth. Her faith in me and constant encouragement have were a source of inspiration on many a frustrating day. She was so willing to give me time and assistance, both in the lab and out, and I can't thank her enough for what it has meant to me and the respect I have for her because of it.

Last but certainly not least, I thank all of my family for putting up with me while I came on this journey. Most especially I thank my husband Owen, for he by far has dealt more with my graduate school blues than anyone else. I thank him for being there and for just being himself and for not running for the hills, despite my temperament during this tenure. I thank Owen's family, Martin and Lonna. I am so blessed to have them as a second set of parents. I also thank PaJ, Dick, and the Sallys, for their endless encouragement and praise; it has helped me press on, each step along the way. And I thank little Ms. Lindsey for being a source of spiritual inspiration for me whether she knows it or not.

I also thank all my family who has taught me the perfected art of putting the fun in dysfunction. I thank Mom, Jim, Deb, and Mike for believing in me and for making the journey all the more interesting, even while having no idea where I was heading. I love them more than they know or that I have had time to show over the last few years. My Mom is the strongest, most stubborn person alive and I am proud to be just like her. Jim, for his wisdom and guidance over the years I am deeply indebted. Mike, keep up the good work bro, you will get here one day too. To the greatest cook in the US, my sister

Debbie, I am so proud of you, cooking is much more difficult than chemistry and you a doctor of Pi. E. in that area! And to the rest of my family, Grandma and Grandpa, June, Judy, Don, Cindy, and Heather, I thank them for their love and support. I also have to step back a few years and thank Dr. Pat Saulson, Dr. Brad Herbert, and Dr. Karen Sentell for steering me down this path. They are the greatest of mentors.

Graduate school was made more pleasant by the presence of many good friends and colleagues. I thank the members of the Winefordner group, past and present, for your friendship. I thank my fellow orphans in the mass spec laboratory and Lydia and Jodie. I will miss the entertainment and excitement of working in the orphanage with them; it has been such a treat! I remember starting in the MS lab a lost soul. Then one at a time I got all these sisters and a brother to work with. I hope I have inspired them as much as they have inspired me. No more spore talks or blue cupcakes! Gabby, Chad, Kelly, Danielle, Tracy, Kristen, Romaine, Lori, Lani, and Violeta all made my bad days good, and my good days better, for that I can't thank them enough.

If there is anyone I have arbitrally left off, I thank them from the bottom of my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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

MALDI-TOFMS BASED PROTEIN PROFILING AS A DIAGNOSTIC TOOL FOR THE ANALYSIS OF *Bacillus* SPORES AND CELLS

By

Danielle Nicole Dickinson

August 2004

Chair: James D. Winefordner
Major Department: Chemistry

This research focuses on the development of Matrix-Assisted Laser Desorption/Ionization Time-of -Flight Mass Spectrometry (MALDI-TOFMS)-based protein profiling as a rapid diagnostic tool to detect and discriminate microbial species. MALDI-TOFMS is well suited for this task because of its rapid analysis time (<1 minute), low sample requirement, sensitivity, reproducibility, and resolving power. Analysis of whole bacterial cells and spores with this technique has given rise to unique "protein fingerprints" that can be used for identification at the species and strain level. Identification can be accomplished by using statistical algorithms to find the best match in a database containing fingerprints from previously analyzed bacterial species. The diversity found within bacterial species and the effects of environmental conditions on protein profiles from identical strains have proven to be a challenge for the statistical analysis of the spectra. To this end, we have sought an understanding of the variability in the protein profiles among strains of the same species and have evaluated the factors

affecting the expression and extraction of the proteins used as biomarkers. Systematic evaluation of these factors is crucial for bringing this technology into fruition as a viable diagnostic tool for microbial analysis.

We have demonstrated the versatility and efficacy of MALDI-TOFMS protein profiling for bacterial identification by examining over 50 different spore strains of *Bacillus*, the most diverse study of the genus reported to date. A one-step sample treatment and MALDI-TOFMS preparation was designed to obtain spectra rapidly with a wide range of protein biomarkers, including several higher molecular weight (10-25 kDa) protein species not reported in other MALDI spore preparations. Linear correlation analysis, hierarchal cluster analysis, and spectral visualization were used to identify and catalog all Bacillus spores evaluated. To validate the use of MALDI-TOFMS protein profiling for species and strain differentiation, result of the protein profiling were compared with 16S rDNA sequences and DNA:DNA hybridization for their bacterial systematics and molecular phylogenetic affiliations. The effect of strain variation and environmental conditions (such as age, storage conditions, and exposure to radiation and sterilization) were examined to facilitate identification of invariant and omnipresent biomarkers in the spectra. The biomarkers needed for species delineation were targeted for further proteomic identification.

CHAPTER 1
INTRODUCTION

This research focuses on the development of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOFMS) protein profiling as a rapid diagnostic tool for identification of bacteria and bacterial spores. Analysis of whole bacterial cells and spores with this technique gives rise to unique "protein fingerprints" that can be used for the identification at the species and strain level. Identification can be accomplished by using statistical algorithms to find the best match in a database containing fingerprints from previously analyzed bacterial species. The algorithms have been tested with both a large diversity of bacterial species and strains and by the effect of environmental conditions on the resulting spectra of identical strains. To this end, we have sought an understanding of the variability in the protein profiles among strains of the same species; evaluated the factors affecting the expression and extraction of the proteins used as biomarkers; and identified omnipresent genus-, species- and strain-specific protein biomarkers. The systematic evaluation of these factors is crucial for bringing this technology into fruition as a viable diagnostic tool for microbial analysis.

Our interest in investigating this technology is two-fold. The primary interest stems from the fact the *Bacillus* spores are the major source of contamination found in spacecraft assembly facility (SAF) clean rooms. The planetary protection requirements of space missions destined to contact the surface of other planets require technologies for validating decontamination processes and archiving the bioburden of flight hardware and

facilities. These technologies must be sensitive, accurate, rapid, and cost-effective, and must be able to provide an "organic signature" of the organism that could allow scientists to distinguish it as "forward contamination" in the search for extraterrestrial life. The MALDI-TOFMS methodology developed here provides one possible answer to these challenges, and could record in the form of a protein fingerprint, the microbial diversity associated with space missions. Our more general interest is in rapid, sensitive, and selective microbial detection and identification at the species and strain level, which is a necessity for the differentiation of viable pathogenic and nonpathogenic microbial species. The development of technologies that accomplish this level of distinction would have a significant impact in the areas of occupational and health care, homeland defense, and environmental monitoring.

The Genus *Bacillus*

The genus *Bacillus* is one of the largest and most ubiquitous genera of bacteria containing 65 valid species, with new species continually being described.[1] The type species of *Bacillus*, *Vibrio subtilis* was first described by Ehrenberg in 1835 and was renamed *Bacillus subtilis* in 1872 by Cohn. The genus has become the graveyard for all aerobic or facultatively anaerobic, spore-forming, rod-shaped bacteria. Taxonomic characterization and systematics of *Bacillus* have been an area of great debate for over a century. The genus has been classified into six RNA groups based on 16S rDNA sequence similarity, spore morphology, spore position in the mother cell, and the presence or absence of mother cell swelling during sporulation.[2] The genus has gained notoriety with taxonomists for its extreme phenotypic diversity and heterogeneity. As a result, this is one of the most animated areas in systematic bacteriology studies.

Most *Bacillus* species are regularly encountered and cultivated from soil samples, their primary habitat, from which they can contaminate anything. *Bacillus* species are particularly important in the medical, veterinary, military, and industrial fields. They are probably most noted for their negative effects, which include food spoilage, clean room contamination, biodeterioration, and causing various infections and foodborne illnesses in humans and many animals. The most infamous member is *B. anthracis*, the bioterrorism agent that causes anthrax. Although notorious for the negative effect they can have on human health in particular, *Bacillus* species possess redeeming qualities. They are rich sources of extracellular enzymes (such as proteases and amylases); of peptide antibiotics such as bacitracin; and of insecticides such as the widely used toxins from the species such as *B. thuringiensis* and *B. popilliae*.[2]

The most distinguishing characteristic of the genus is the ability to produce a resistant endospore. The spore is formed within the mother cell in response to nutrient deprivation and can be oval, spherical, or cylindrical. Spores are highly resistant to agents such as heat and radiation, and cannot be easily destroyed even by harsh chemical treatment, disinfectants, or desiccation.[1,3] The increased resistance of spores, although not completely understood, has been partly explained by the impermeability of the spore coat, dehydration of the core, and the protective proteins that bind to the DNA. In the metabolically inert spore form, these bacteria can remain dormant for hundreds of thousands of years. Within the spore the essential macromolecules (and a variety of other substances) are stored until conditions become favorable for survival; at which point they are triggered to return to an active vegetative state. The resiliency they exhibit enables

the genus to be ubiquitous in the environment, a common source of contamination, sterilization resistant, and an ideal bioterrorism agent.[1,3]

Spore Architecture and Composition

A closer look at the spore shows significant differences in the composition and location of many biomolecules when compared to a vegetative cell. The cell wall of a gram-positive vegetative cell is characterized by a rigid layer of peptidoglycan. This layer is relatively easy to penetrate, either through the use of enzymes such as lysozyme (which breaks the 1,4-glycosidic bonds in the peptidoglycan), or by extreme changes in osmotic pressure or pH. Vegetative cells are also susceptible to desiccation, heat, radiation, and sterilization.

The spore, in comparison, is more complex with several outer layers that are believed to contribute to its increased resistance. Spore species may contain an outermost layer called the exosporium (Figure 1-1A). All species contain a spore coat, typically comprising an inner and outer layer (Figure 1-1A, B). The exosporium and spore coat are mainly comprised of proteins and glycoproteins. The cortex, a peptidoglycan layer similar to that found in a vegetative cell, is the next layer of the spore. Within the cortex is the dehydrated spore core. The core contains the same parts as the vegetative cell (including the cell wall, cytoplasmic membrane, ribosomes, and DNA). A high concentration of a calcium-dipicolinic acid complex is present in the core of all spores. Bound to the DNA (and unique to spores) are proteins known as small acid soluble proteins (SASPs), which protect the DNA against damage from radiation, desiccation, and dry heat.[4]

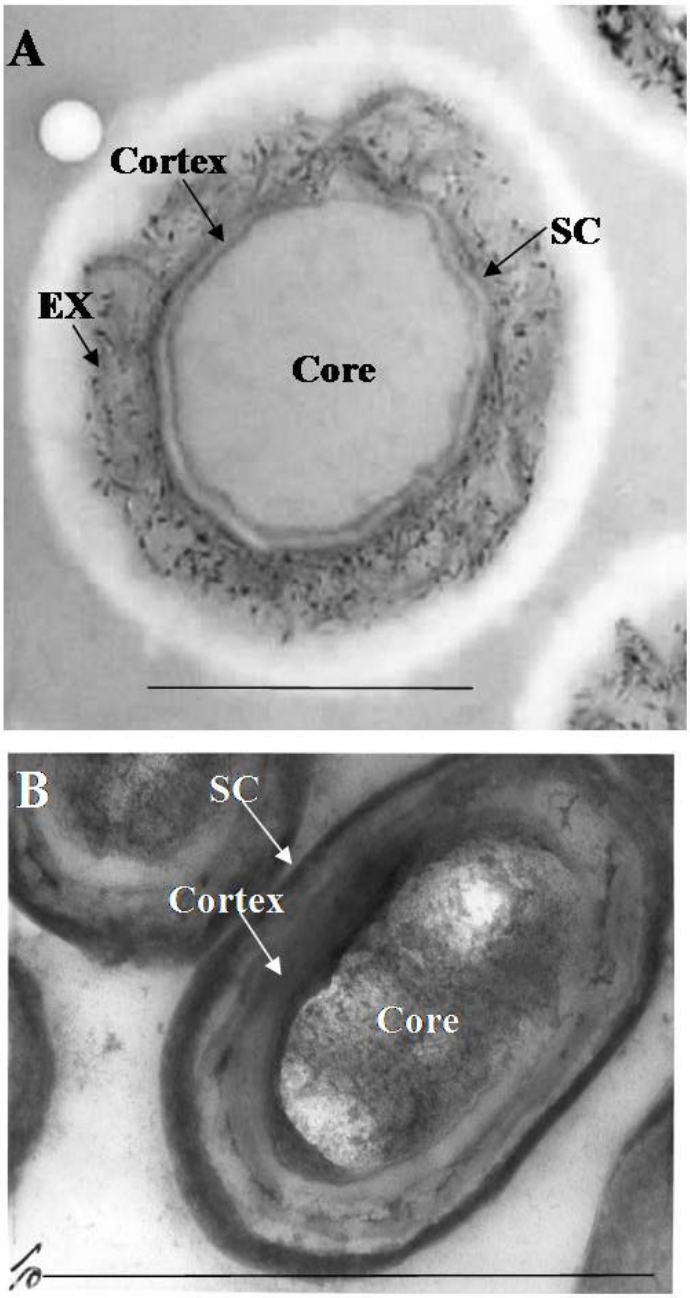

Figure 1-1. Transmission electron microscopy image showing typical spore architecture. A) *B. odysseyi* PTA-4993 which contains an exosporium (EX). B) *B. subtilis* 168 which does not contain an exosporium. The core, cortex, and spore coat (SC) are shown for both spore species.

Sporulation and Germination

Sporulation occurs in a series of stages (0, II, and III-VII) that can be monitored by phase contrast microscopy. A pictorial representation of the sporulation and germination stages is shown in Figure 1-2. The end of exponential growth is considered time 0 for sporulation, and occurs when the cells reach stage II and the cell divides into two asymmetric compartments, each with its own chromosome. The larger division is termed the mother cell or sporangium; the smaller compartment is the forespore. In stage III and IV, the forespore becomes engulfed by the mother cell, the peptidoglycan cortex layer is deposited on the outside of the developing spore, and the SASPs are synthesized within the forespore. During stage V and VI, the spore coat proteins are deposited and the spore reaches maturity with a full arsenal of resistances. Finally, the mother cell lyses, releasing the mature spore during stage VII. This results in the appearance of phase-bright refractile bodies which are observed when using phase contrast microscopy. The entire process of spore formation takes 6-8 hours.[5-7]

When nutrients are returned to the medium, the spore undergoes a process called activation, whose mechanism is not well understood. Germination begins within minutes and can be characterized by a rehydration of the spore core, release of cations and dipicolinic acid, degradation of the SASPs by the germination protease protein (GPR), and the loss of refractility and resistance. Later in germination, the cortex undergoes hydrolysis followed by metabolism and protein synthesis. This is followed by outgrowth, when emergence and elongation occur, during which normal cell division resumes and the coat remnants are discarded. The time for completion of this process is under 1.5 hours.[5-7]

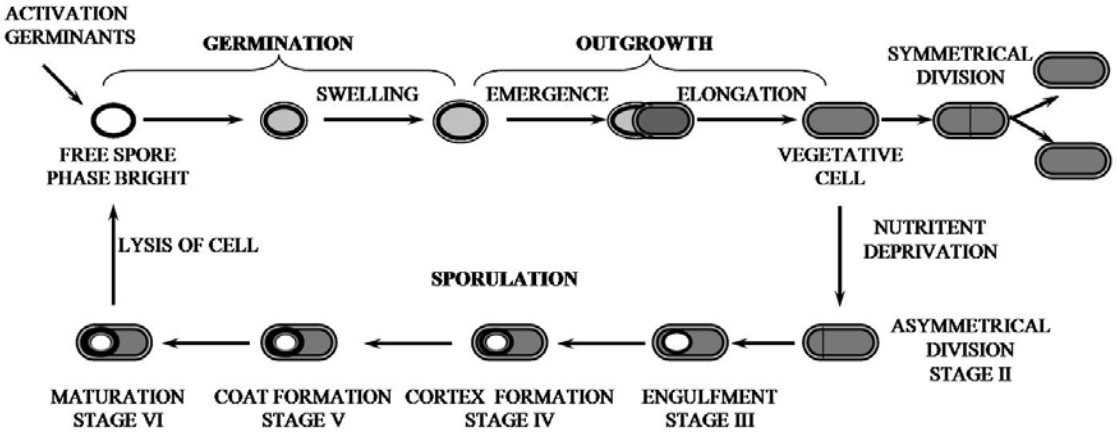

Figure 1-2. Sporulation and germination cycle.

Microbial Identification and Classification

> "The biologist is attuned to the vagarities of living things and does not expect an experiment to be exactly repeatable . . . is surprised by the expected and astounded by the fulfillment of a forecast of prediction. The bacteriologist must never forget that genera and species are artificial concepts and that the bacteria show no interest in their classification"—Sam Cowan, 1978

When developing technology for the identification of microbial species, consideration should be given to the biological system itself, the processes it can undergo, and the chemicals and biomolecules available for analysis. One of the most challenging aspects of any biological system (in analytical terms) is its almost constant potential for change and adaptation. Analytical chemists strive for reproducibility, specificity, selectivity, and detection limits. Traditional analytical approaches are not attuned to dealing with "biological flux" or the vast expanse of diversity that can exist. To be successful, we must do what is less traditional, and accept that biological samples have "a mind of their own." The definition of a successful biological analysis may or may not concur with traditional analytical measures of success.

Traditional techniques for the characterization and identification of microorganisms have relied on lengthy biochemical, nutritional, and physiological testing. Often these tests are inconvenient to prepare and perform, difficult to standardize and interpret, and can be challenging to reproduce.[8-10] Modern techniques for microbial classification and identification have focused on the development of chemotaxonomic and molecular-based methods. These techniques can be classified broadly as genotypic or phenotypic. Examples of genotypic techniques include PCR-based analysis, DNA hybridization, genetic fingerprinting, direct sequencing, and nucleic acid probing. Fatty acid methyl ester analysis, pyrolysis mass spectrometry, whole-cell protein profiling via gel electrophoresis or mass spectrometry, antibody-based methods, and various miniaturized

test kits for determining biochemical and nutritional requirements are examples of phenotypic techniques.[8-10]

Genotypic methods are faster and are usually more reliable than traditional biochemical methods, though they have drawbacks, including the stability of consumables, availability of specific probe sequences, and the associated cost and time required for gene sequencing. Commercially available identification kits and other newly developed technologies, mentioned above for phenotypic analysis, have the advantage of speed and convenience when compared to most traditional methodologies, but still require an incubation period and may have difficulties associated with reproducibility. Protein expression, metabolic profiles, and fatty acid profiles can fluctuate dramatically based on environmental and nutritional variables during different stages of growth.[8-10]

Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOFMS*)*

MALDI-TOFMS has become an essential tool for analyzing a wide array of biomolecules, particularly proteins. The explosion in genomic and proteomic research in the past two decades has placed high demands on instrumentation and techniques for analysis. MALDI-TOFMS has become a primary player in both of these arenas due to its high throughput, sensitivity (femtomole range for most proteins), and handling of complex biological samples.

In MALDI-TOFMS, a solid organic matrix compound is dissolved in an appropriate solvent and combined with a protein sample solution. A small volume (typically 0.5-3.0 μL) is spotted on a stainless-steel target plate and allowed to dry. The target is placed (*via* a probe) into the vacuum chamber of a time-of-flight mass spectrometer. The sample is irradiated with a pulsed laser, resulting in the desorption and

ionization of the analyte molecules. The laser is most often a nitrogen laser which produces photons in the ultraviolet range (337 nm), although infrared lasers are also used. The matrix acts as the vehicle to desorb and ionize the sample molecules with little or no fragmentation. The matrix molecules absorb most of the energy from the photons, and vaporize to form an expanding plume that carries the sample molecules into the gas phase. Ions are accelerated by an electrostatic potential (V) to a certain velocity (v) and a total kinetic energy (E). The ions traverse a field-free drift region where they are separated based on differences in their velocity. A channel electron multiplier detects the ions. The time-of-flight of the ions is recorded and converted to a mass-to-charge ratio, using the TOF relationship:

$$m/z = \frac{2Vt^2}{L^2}$$

where t = time, L = length of flight tube, z = charge, V = accelerating voltage, and m = mass of the ion.[11] To increase the resolving power of the time-of-flight mass spectrometer, a reflectron can be added to focus the kinetic energy of the ions. The reflectron is most effective at relatively low masses, and is more frequently used for peptide analysis below 4 kDa.

MALDI-TOFMS for Bacterial Fingerprinting

MALDI-TOFMS has demonstrated great promise for interrogating microbial species and for identifying proteins using comparative and clinical proteomic approaches.[11-14] Microbial analysis with MALDI-TOFMS dates back to the late 1980s.[15] Many groups have demonstrated the versatility of this technique, from analysis of cell lysates to whole-cells to the analysis of PCR products.[16] MALDI-TOFMS is well-suited for this task due to its rapid analysis time (<1 min/sample), low sample requirements,

sensitivity, and resolving power. Analysis of whole bacterial cells and spores with this technique has given rise to unique protein fingerprints that can be used for identification at the species and (at times) strain level.[16-33] All of the studies rely on the growth and/or sporulation of cells in the laboratory; none are sampled directly from the environment. Vegetative cells generally produce a relatively large number of biomarker proteins that can be used for subsequent pattern recognition or correlation analysis. However, the proteins expressed in vegetative cells are dynamic, and can vary dramatically based on cultural conditions. In contrast, extraction of proteins from spores has been more challenging, providing only a limited number of biomarker peaks when compared to their vegetative counterparts.

The spore associated biomarker peaks detected using MALDI-TOFMS with a UV laser are reported in the range of 3-10 kDa. Various sample pretreatments including the use of infrared (IR) laser irradiation,[34-37] corona plasma discharge,[19,26] sonication,[26] and the addition of 5% trifluoroacetic acid (TFA)[34] or 1M HCl,[38] have been used on spores to increase the number and intensity of biomarkers observed in the spectra. These methods have had some success; however, in most cases, these treatments require longer sample preparation times and visualization of peaks above 10 kDa is still limited. Fenselau *et al.*[39] reported detection limits as low as 5,000 cells/spot; however, only a single protein biomarker was reported. Most conventional MALDI-TOFMS bacterial research has focused on species differentiation without the identification of protein biomarkers, and therefore, the ability to extract strain-specific and pathogen-specific biomarkers has not been thoroughly investigated. In the case of spores, recent studies have identified a limited number of biomarker peaks as SASPs; however, these proteins alone do not allow

for differentiation at the species level. This is evident in the case of *B. thuringiensis* and *B. cereus*.[32,38,40]

In addition to the direct MALDI-TOFMS analysis of whole spores, various groups have recently been engaged in full-scale proteomic analysis of *B. anthracis*[37,41] and *B. subtilis*[37,42] using one or two-dimensional gel electrophoresis in conjunction with MALDI-TOFMS or liquid chromatography mass spectrometry. While these extensive studies improve our understanding of the spore coat and allow for limited species comparisons, they are very time consuming and have not been applied to other *Bacillus* strains. These studies are also limited to species with sequenced genomes.

Published studies show that MALDI-TOFMS of whole bacterial spores is feasible and of practical value, lending speed and higher accuracy to the analysis. MALDI-TOFMS for microbial analysis provides a rapid, relatively simple analysis that is amenable to all species, and is not reliant on previous knowledge of DNA sequences or antibody interactions. MALDI-TOFMS protein profiling is well-suited for high throughput and automation, requires minimal sample preparation, has superior reproducibility, and has much higher resolution than gel-based techniques.

To prove MALDI-TOFMS fingerprinting as a useful technology for bacterial identification, the technique must be able to rapidly differentiate and identify genus-, species-, and strain-specific biomarkers over a wide variety of spores. More importantly, it must be able to identify biomarkers that can differentiate pathogenic strains from nonpathogenic strains. To this end, this research was performed on over 50 species, including 21 type strains and 25 wild-type environmental isolates, and 17 *B. cereus* serovars. To our knowledge, this is the largest collection of *Bacillus* spores ever

evaluated with MALDI-TOFMS protein profiling. This large collection of *Bacillus* affords us several advantages. First, it allowed elucidation of the protein differences that are present at both the species and strain level. Second, it allowed evaluation of the reliability of the biomarkers over a wide range of isolates. Third, it allowed for the determination of the level of distinction needed to differentiate pathogens from other similar nonpathogens that are of the same or related species. This highly specific level of discrimination is required to differentiate nonpathogenic strains (*B. anthracis* Sterne) from pathogenic strains (*B. anthracis* Ames). These biomarkers allowed rapid identification of proteins as targets for molecular probes and other biosensors that are field-portable, robust, small, and both sensitive and highly specific. This could decrease the time and cost of identifying targets for sensor-based counterterrorism systems.

Research Overview

This body of research begins with a description of the optimization of the conditions for the extraction of proteins from spores. Using the optimized extraction conditions the research moved into the analysis phase where over 50 different spore species were analyzed. Chapter 3 addresses the successful differentiation of 11 different *Bacillus* species, the most diverse study of the species to date. This chapter also describes in detail the statistical processing of the spectra using linear correlation and hierarchal cluster analysis. The effect of strain variation within a species is addressed in Chapters 4 and 5 and the criteria for including a strain within a species using linear correlation were established. Chapter 4 deals specifically with using MALDI-TOFMS protein profiling in a polyphasic taxonomy approach for the identification and classification of *B. pumilus* isolates. MALDI-TOFMS protein profiling proved to be more accurate than metabolic profiling and was complementary to *gyrB* sequencing and

DNA hybridization for the identification of these isolates, which included the possible identification of a new species of *Bacillus*. Chapter 5 tackles the differentiation of the *B. anthracis-B. cereus-B. thuringiensis* (BACT) group spores and cells. This is the first investigation of a wide variety of BACT group bacteria (20 strains) with MALDI-TOFMS protein profiling where the results are compared directly with genetic analysis for their bacterial systematics and molecular phylogenetic affiliations.

The next portion of the research focused on the identification of the protein biomarkers that were found to be species specific in the studies outlined in Chapters 3-5. MALDI protein extracts from several *Bacillus* species were analyzed by tandem mass spectrometry techniques to obtain peptide mass tag data in Chapter 6. The first report of the identification of coat proteins from a MALDI extract is included for the *B. subtilis* 168 strain in the study. For organisms that did not have sequenced genomes most of the proteins identified are SASP associated due to the high sequence conservation among these proteins, although some surface associated proteins are identified in several *B. pumilus* species. The final chapter, Chapter 7, addressed the effect of environmental exposures on *Bacillus* spores and demonstrated that species specific biomarker peaks were maintained over most of the conditions analyzed. The additional information provided by the appearance and disappearance of other peaks in the spectra was shown to be useful for source tracking, forensic investigations, and epidemiological studies.

CHAPTER 2
OPTIMIZATION OF MATRIX AND SOLVENT CONDITIONS FOR THE EXTRACTION OF PROTEINS FROM SPORES

Matrix and solvent selection is a critical factor in the success of MALDI-TOFMS. Selection of matrix compounds is empirical and must be evaluated for each application. Solvent systems for MALDI must balance the organic and aqueous phase to maintain solubility of both the matrix and sample, and must optimize crystal formation. An acidic modifier must also be present to maintain a pH less than 4 to promote crystallization of the matrix in the free acid form. The number, quality, and intensity of peaks in a MALDI spectrum can also be affected by the matrix compound, the solvent, and the acidic modifier chosen for the analysis. Both the enhancement and suppression of peaks have been observed in MALDI by changing various components.

The selection of the solvent system in the case of spores is complicated by the presence of a wide range of hydrophobic and hydrophilic proteins in the exosporium, coat layers, and core of bacterial spores, as well as by the rigidity and chemical resistance of the spores. Two main families of proteins are present in spores at relatively high concentrations: the SASPs and the spore coat proteins.

SASPs have recently been evaluated as biomarkers for the identification of spores. Targeting the SASPs requires that the spores be disrupted, allowing for the release of SASPs from the spore core. This is accomplished by using high concentrations of strong acids, such as trifluoroacetic acid (TFA) and hydrochloric acid, to lyse open the spore. Post release, the SASPs are digested with trypsin and are identified using post-source

decay or ion-trap technologies. Although this approach has allowed for identification of SASPs that are species-specific, it is limited to the species that have sequenced genomes, and it fails to differentiate closely related species such as *B. cereus* and *B. thuringiensis* effectively.[32,38,40] This failure is likely due to the high level of sequence homology among the SASP proteins. The spore coat proteins have not been identified in any direct whole-cell analysis experiments. They possess a high level of sequence divergence, which should allow for higher levels of discrimination. To effectively extract and analyze the spore coat proteins, it is necessary to target mainly hydrophobic proteins. On average, 75% of the known proteins located in the spore coats are hydrophobic.[43] In addition, it would be ideal to use a gentle extraction scheme that would not lyse the spore open during treatment, releasing SASP proteins that would dominate the spectra.

Spores are more resilient and difficult to destroy than vegetative cells, making it difficult to design an efficient protein extraction scheme. Most of the published MALDI-TOFMS spore spectra have used a mixture of acetonitrile/water with various concentrations of TFA.[25-27,32,38] Components used to analyze vegetative cells are far more varied, with past studies using α-cyano-4-hydroxycinnamic acid (HCCA), ferulic acid, and sinapinic acid with a variety of solvents including ethanol, isopropanol, and acetonitrile.[16,20,27,29,31,33,44,45] Voorhees *et al.*[46] and Chait *et al.*[47] recommended using a mixture of 17% formic acid, 30% acetonitrile, and ferulic acid to enhance the high mass signal in whole-cell analysis. This mixture has also been shown to be more tolerant of salts and surfactants, an important factor when considering possible contamination from using crude cell samples. Procedures for analyzing hydrophobic proteins also regularly employ the use of a detergent additive to increase the solubility of the proteins.[48-51]

Detergents typically used include low levels of sodium dodecyl sulfate (SDS), Triton X-100, and octylglucoside. However, various studies show that adding detergents can have negative impacts on the MALDI signal.[52]

The goal of this research was to develop a simple, one-step extraction protocol that provides for the maximum availability of biomarkers for analysis. To this end, we systematically evaluated the extraction of proteins from spores by MALDI-compatible solvents. Several common MALDI matrices, acidic modifiers, and organic solvents for analysis spores were evaluated. Detergent additives were also examined as a method of increasing the number of biomarkers extracted from the spore coat. Spectra generated were evaluated based on the following criteria: signal-to-noise ratio, number of discernable peaks, molecular weight range, suppression effects, reproducibility, and homogeneity of crystal formation. The limits of detection of the optimized extraction system were also investigated.

Materials and Methods

Chemicals and Reagents

The evaluated MALDI matrices, purchased from Sigma Chemical Co. (St. Louis, MO), included sinapinic acid (SA), ferulic acid (FA), dihydroxybenzonic acid (DHB), and α-cyano-4-hydroxycinnamic acid (HCCA). All matrices were used as received except HCCA. The HCCA was further purified by preparing a saturated solution of HCCA in warm ethanol, to which 3 parts water was added. The solution was allowed to stand at 4°C overnight. The HCCA precipitate was then filtered and the matrix was allowed to dry in a desiccators.

Organic solvents used included acetonitrile (ACN), methanol, ethanol, and isopropanol. All solvents were HPLC grade from Fisher Scientific Co. (Fairlawn, NJ).

Trifluoroacetic acid (TFA) was from Sigma-Aldrich Chemical (St. Louis, MO). Aldehyde-free formic acid was obtained from Fisher Scientific Co. (Fairlawn, NJ). The two detergents evaluated included N-octylglucoside (OGP) from Sigma-Aldrich Chemical Co. and the acid labile Rapidgest from Waters (Milford, MA). Cytochrome C, myoglobin, bovine serum albumin (BSA), and insulin were used in calibration mixtures and were purchased from Sigma-Aldrich Chemical Co.

The spore suspensions used for analysis were provided by the Biotechnology and Planetary Protection Group at the Jet Propulsion Laboratory. Three strains were used in the extraction protocol development: *B. subtilis* 168, *B. pumilus* 7061, and the wild-type FO-36b, which has been putatively identified as *B. pumilus*. All spore suspensions were between 1 x 10^8 and 1 x 10^9 spores/mL. The spores were stored in sterile water at 4°C before use.

Sample Preparation and Mass Spectrometry

Saturated solutions of the MALDI matrices (typically 10-20 mg/mL) were prepared in the selected solvent system for analysis. Unless otherwise indicated, standard dried-droplet sample preparation was used to prepare the MALDI spots for analysis. The optimum ratio for mixing was found to be 10 parts matrix to 1 part sample, where the initial concentration of spores was 1 x 10^8 to 1 x 10^9 spores/mL. The spore suspension was premixed 1:10 with the matrix solution and 1 μL of the resultant solution was spotted on the MALDI plate. Samples were allowed to air dry, and no further treatments were applied to the spot post deposition.

MALDI analysis was performed on a Bruker Reflex II TOFMS (Bruker Daltonics, Billerica, MA) retrofitted with delayed extraction. The instrument uses a pulsed nitrogen laser (337 nm) for ionization. Ions were collected in the linear mode and were detected

with a HIMASS™ detector (Bruker Daltonics, Billerica, MA). An acceleration voltage of 20 kV was used in conjunction with a 50 ns delay time. For deflecting matrix and other low molecular weight ion signals, a deflector was set at 2,000 Da. All spectra were obtained by the accumulation of 50 laser shots in positive mode. A three point external calibration was performed daily using either a mixture of insulin, myoglobin, and BSA; or insulin and the doubly and singly charged ions of cytochrome C.

Results and Discussion

Initial Studies

Initial studies focused on the use of sinapinic acid as a matrix compound. Sinapinic acid was typically used in combination with 0.1% TFA and 30% ACN for the analysis of proteins. When this combination was applied to the *B. subtilis* 168 spores in this study, very few biomarkers were observed (Figure 2-1A). Changing the acidic modifier to 17% formic acid had a profound effect on the spectrum, increasing the observable number of biomarker peaks from 6-8 barely discernable peaks to approximately 15 well-resolved peaks (Figure 2-1B).

Because the addition of formic acid increased the organic content of the solvent system, a 50% ACN/ 0.1% TFA solvent was also evaluated (Figure 2-2A) and demonstrated little improvement over the previous TFA solvent system (Figure 2-1A). The spectrum in Figure 2-2B represents an analysis where a 0.1%TFA/50% ACN solvent was first deposited and allowed to dry. A mixture of the spore sample with 10 parts 17% formic acid/30% ACN was deposited on top of the matrix layer. This gave a similar spectrum to the premixed dried droplet approach used above (Figure 2-1B), indicating that the effect of the formic acid is likely an enhancement in protein solubility more than an effect of the mechanics of the MALDI deposition.

Using the 17% formic acid/30% acetonitrile solvent system as a "base solvent," several matrix compounds other than sinapinic acid were evaluated. All matrix compounds were dissolved in the base solvent. DHB, unlike the other 2 matrices, was water soluble, and was also tested in 17% formic acid/83% water. Representative spectra from each matrix are shown in Figure 2-3. Ferulic acid (not shown) gave an identical spectrum to the sinapinic acid, and the effects of these 2 matrices will be discussed separately. When compared with DHB and HCCA, sinapinic and ferulic acid matrices gave higher signal-to-noise ratios and the largest range of biomarkers. The higher molecular weight proteins were not evident in the DHB samples (Figure 2-3B and C). However, the 17% formic acid/83% water sample (Figure 2-3C) highlighted additional biomarker peaks found in the 5-10 kDa range.

Acidic Modifier

To further characterize the discrepancy between the two matrix solutions a study was performed to evaluate the effects of the acidic modifier on the resultant MALDI spectra. The following solvent systems were prepared:

- 30% ACN, 70% 0.1%TFA (pH=1.90)
- 30% ACN, 53% H2O, 17% formic acid (pH=1.56)
- 30% ACN, 55% H2O, 15% formic acid (pH=1.58)
- 30% ACN, 60% H2O, 10% formic acid (pH=1.67)
- 30% ACN, 65% H2O, 5% formic acid (pH=1.82)

Sinapinic acid and ferulic acid were dissolved in each of the solvent systems. A calibration mixture (CM-IMB) was prepared containing insulin, myoglobin, and bovine serum albumin at concentrations of 15, 100, and 100 pmole/µL, respectively. A 1 µL aliquot of the calibration mixture was mixed with 24 µL matrix solution. The calibration mixture was used to ascertain suppression effects that do not result from differences in

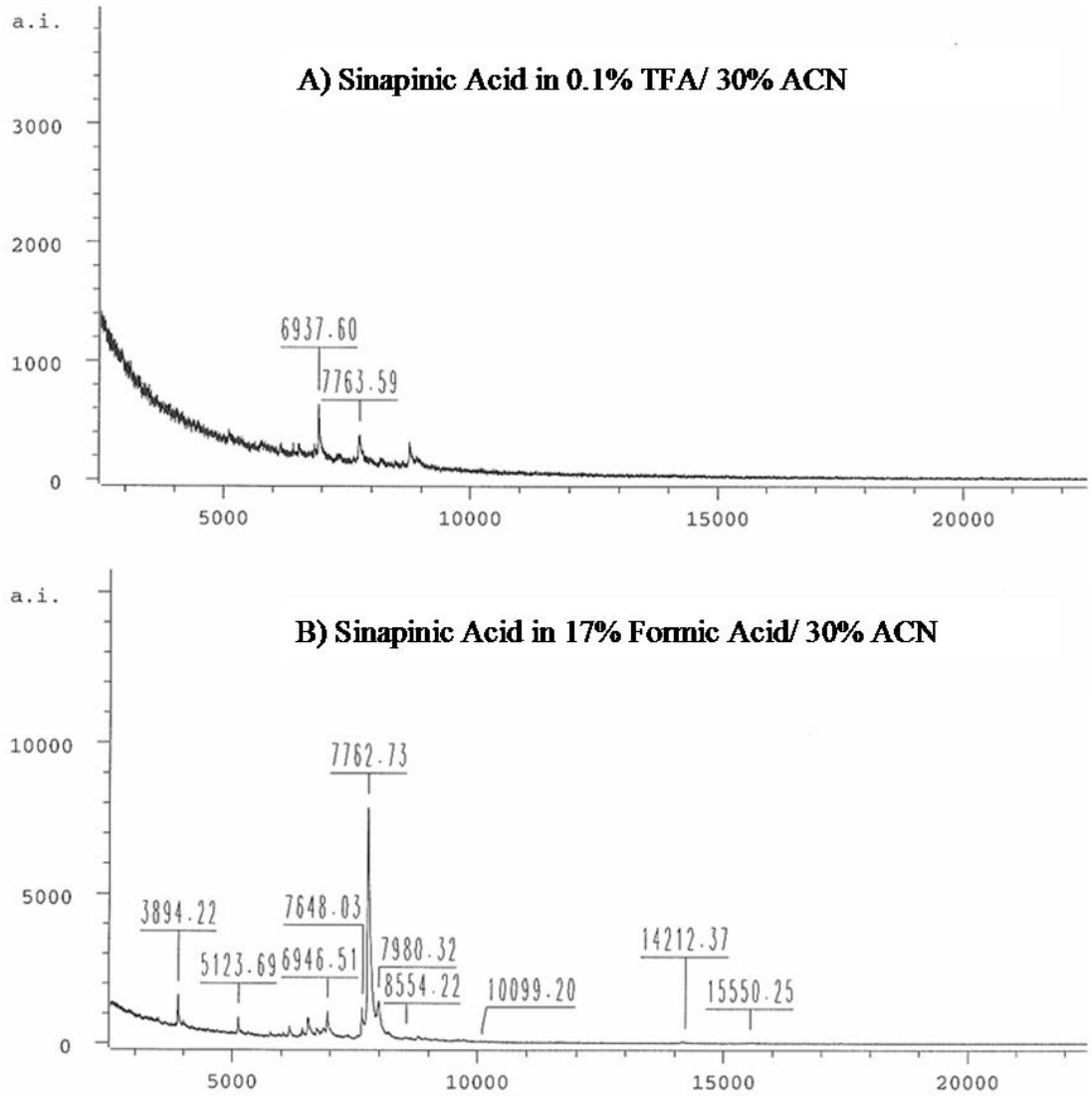

Figure 2-1. Comparison of TFA versus formic acid. A) *B. subtilis* 168 spores in 0.1% TFA/30% acetonitrile. B) *B. subtilis* 168 spores in 17% formic acid/30% acetonitrile.

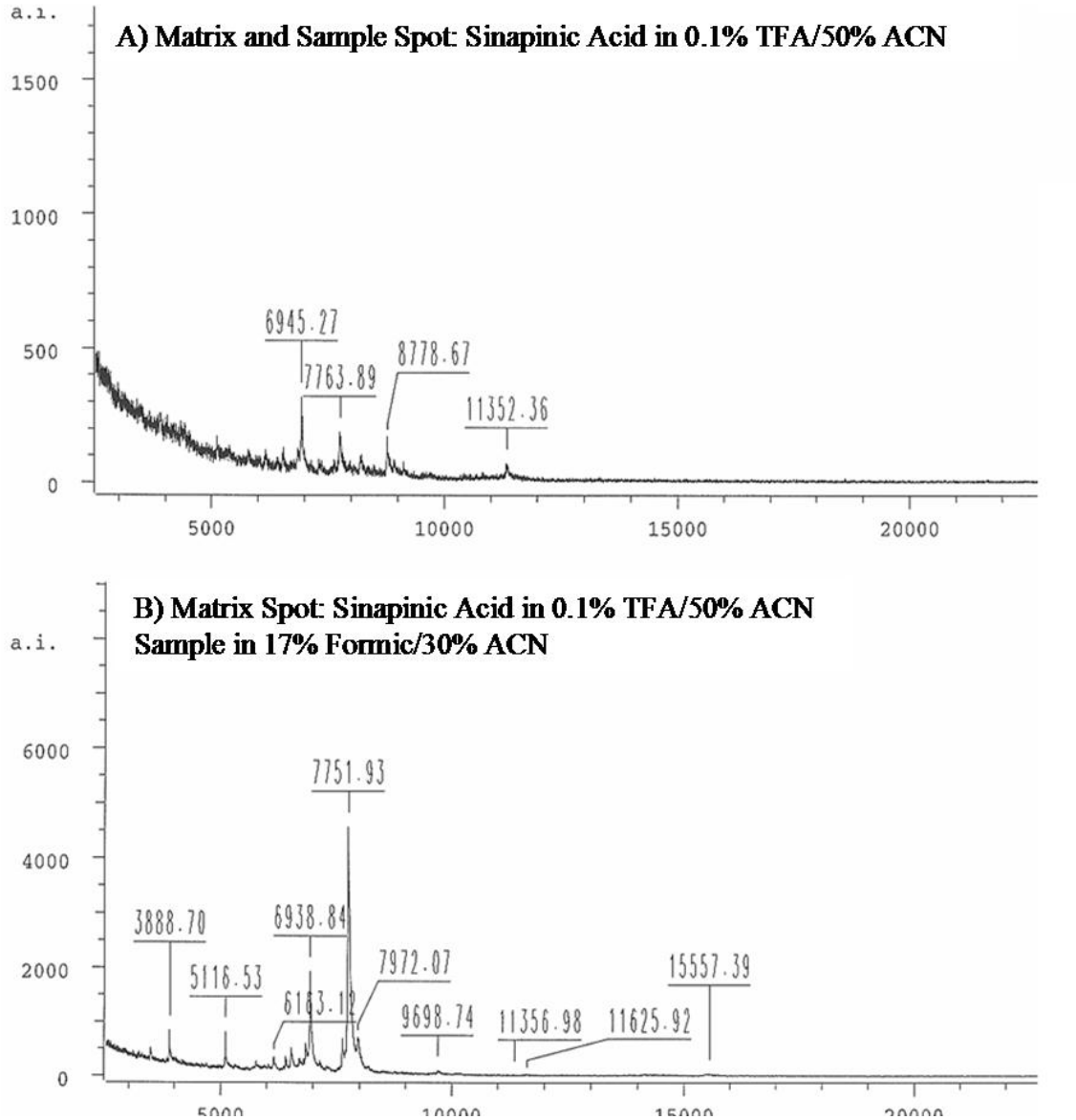

Figure 2-2. Formic acid extraction effect. A) *B. subtilis* 168 spores in 0.1% TFA/50% ACN dried droplet preparation. B) 0.1% TFA/50% ACN matrix layer applied first followed by deposition of *B. subtilis* 168 spores in 17% formic acid/30% acetonitrile.

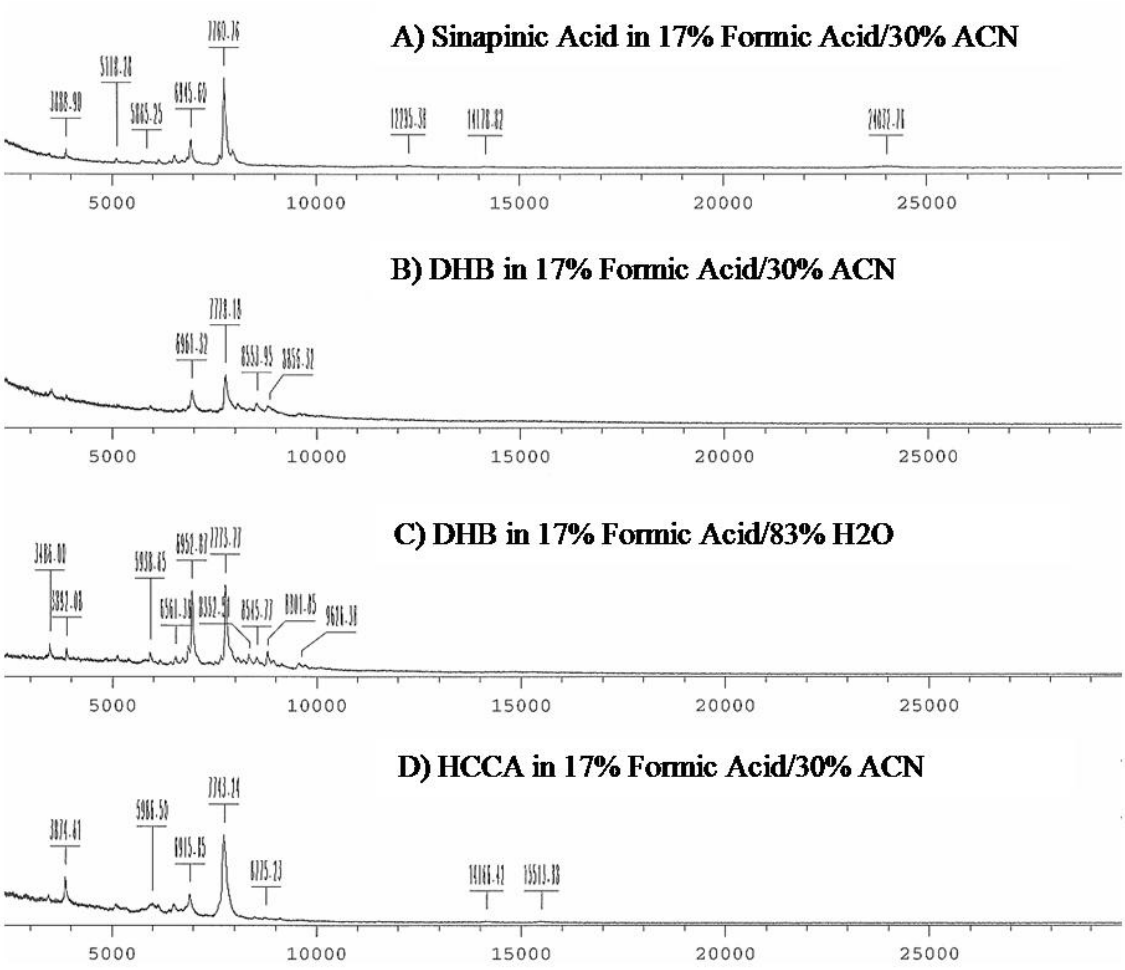

Figure 2-3. Comparison of different matrix compounds and 17% formic acid modifier. A) Sinapinic acid in 17% formic acid/30% ACN B) DHB in 17% formic acid/30% ACN C) DHB in 17% formic/83% water D) HCCA in 17% formic acid/30% ACN

extraction efficiency. Each sample (CM- IMB and *B. subtilis* 168) was spotted in triplicate on the MALDI plate. The signal-to-noise ratio was determined for peaks of interest in each spectrum. In the CM-IMB spectrum, the three peaks of interest were at masses 5,734, 16,952, and 66,432 Da. In the spore spectrum, the peaks of interest were the biomarker peaks at 3,950, 6,648, and 7,760 Da. Data analysis was performed by averaging 3 spectra per spot and obtaining the standard deviation within the spot. The signal-to-noise values from each of the samples were averaged and the standard error was calculated. The resulting data can be seen in Figure 2-4 for the calibration mix and Figure 2-5 for the spore samples. The plots are of the signal-to-noise versus solvent composition for each peak of interest in the spectra.

Sinapinic acid gave higher signal-to-noise ratios overall except in the case of high molecular weight proteins like BSA. The solvent system 30% ACN, 70% 0.1%TFA produced superior results for the calibration mix in both matrices; however, neither produced significant signals from the spore sample. In contrast, both matrices produced superior signals for the spore biomarkers when dissolved in the 30% ACN, 53% H2O, 17% formic acid solvent system. For both matrices, the biomarker signal detected decreased as the percent formic acid was decreased in the solvent mixture. In the formic acid series, there was a decrease in signal for all the proteins in the calibration mix as the percent formic acid increased. These results were in opposition to the spore samples, as signal increased with increasing formic acid. These trends suggested the formic acid aids in the extraction of proteins from the spore coat.

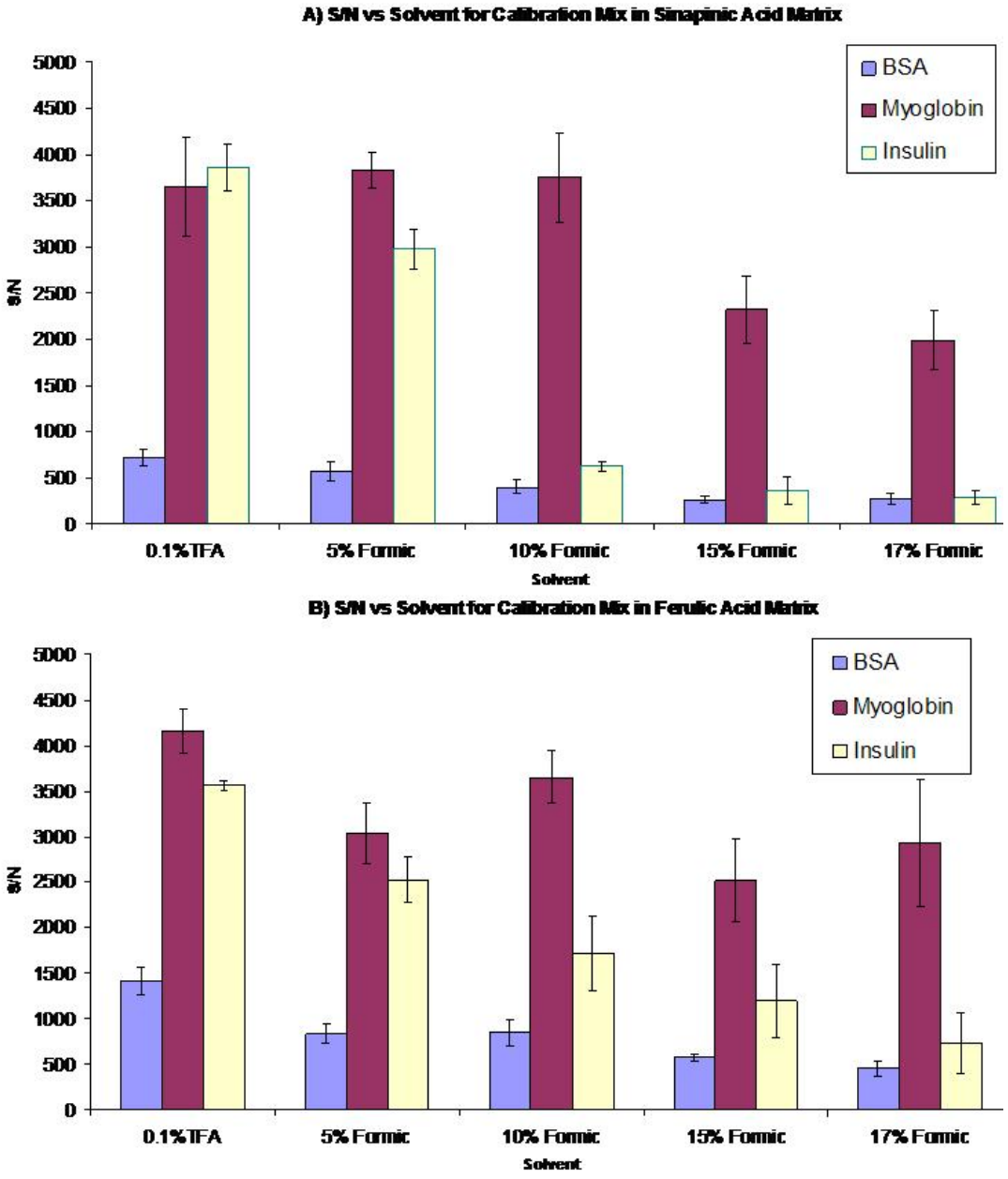

Figure 2-4. Signal-to-noise versus solvent for calibration mix. A) Sinapinic acid matrix. B) Ferulic acid matrix. The solvents, from left to right in each graph are 0.1% TFA, 5% formic acid, 10% formic acid, 15% formic acid, and 17% formic acid in 30 % acetonitrile. The error bars represent the standard error of 9 measurements.

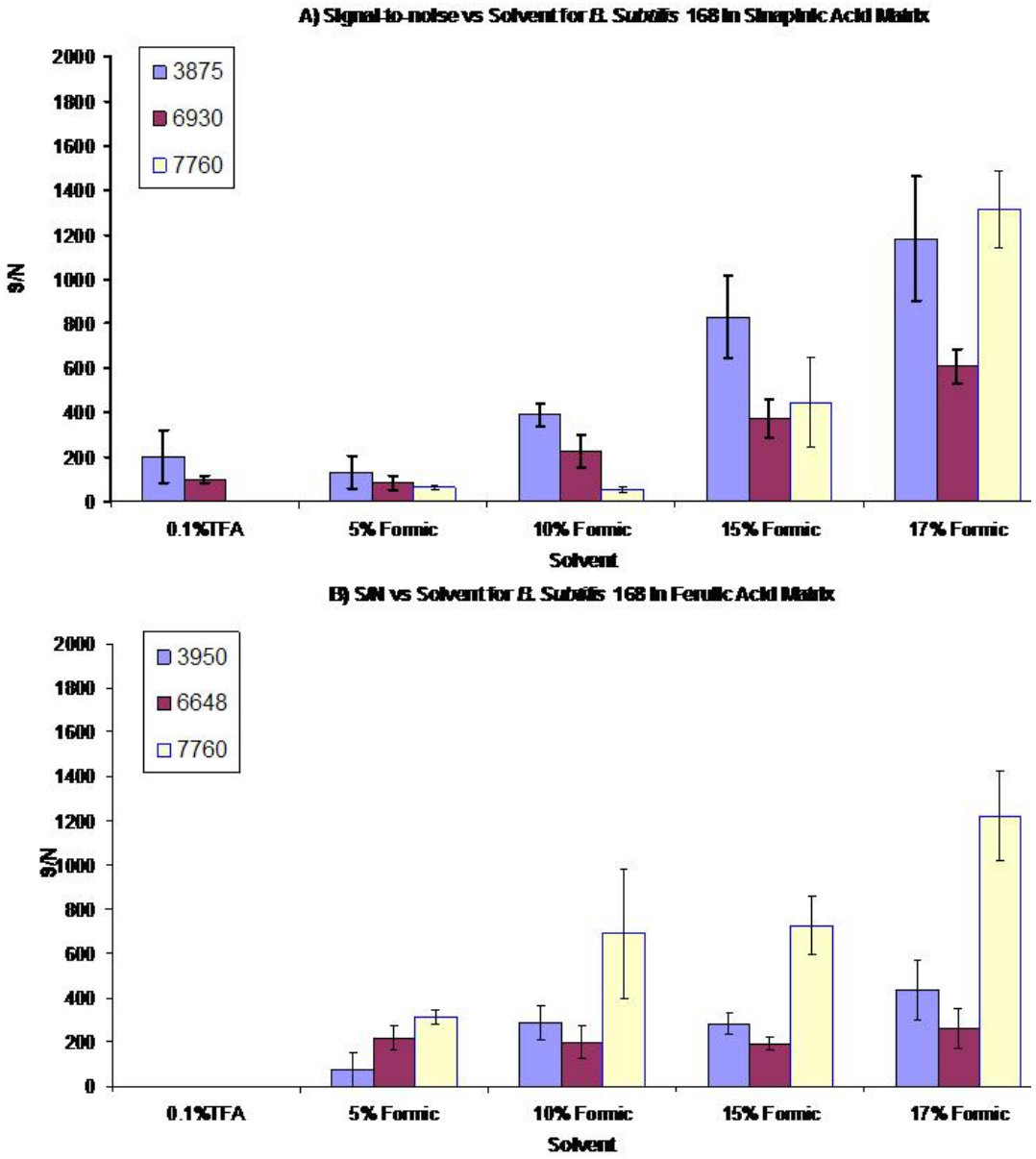

Figure 2-5. Signal-to-noise versus solvent for the *B. subtilis* 168 spore suspension with varying acidic modifier concentrations. A) Sinapinic acid matrix. B) Ferulic acid matrix. The solvents, from left to right in each graph are 0.1% TFA, 5% formic acid, 10% formic acid, 15% formic acid, and 17% formic acid in 30 % acetonitrile. The error bars represent the standard error of 9 measurements.

Because increasing the formic acid concentration seemed to increase the extraction efficiency of the protein biomarkers from the spores, higher concentrations of formic acid were investigated. The percentage of acetonitrile was kept at 30% while the percent formic acid was increased from 10 to 60%. The results are shown in Figure 2-6 for both sinapinic and ferulic acid matrices. Sample spectra from the sinapinic acid samples for each concentration are shown in Figure 2-7. Higher concentrations of formic acid (30 to 40%) enhanced extraction, giving rise to higher molecular weight biomarkers (> 15 kDa) not seen at lower concentrations. This observation was supported by a 1-D gel electrophoresis studies (Figure 2-8) where bands emerged at higher molecular weights with increasing formic acid concentrations. However, MALDI spectral quality declined as formic acid concentrations above 30-40% were used. This decline was attributed to inhomogenieties in crystal formation; as the solvent became more hydrophobic, crystal homogeneity suffered due to spreading of the spot.

Using formic acid as a modifier resulted in spectra with higher signal-to-noise ratios and a significantly greater number of biomarker peaks. The visualization of these higher molecular weight proteins has not typically been seen in other MALDI-TOFMS analyses of whole spores.[25,26,34,35,38] When comparing the two matrices, sinapinic acid clearly was advantageous due to enhanced signal-to-noise ratios. However, the ferulic acid matrix produced spots which were more reproducible. Since the long term goals of this project deal with the statistical treatment of spectra, reproducibility was a critical factor in the success of this methodology. Therefore, signal-to-noise was sacrificed in exchange for better reproducibility and the ferulic acid matrix was used in all subsequent studies.

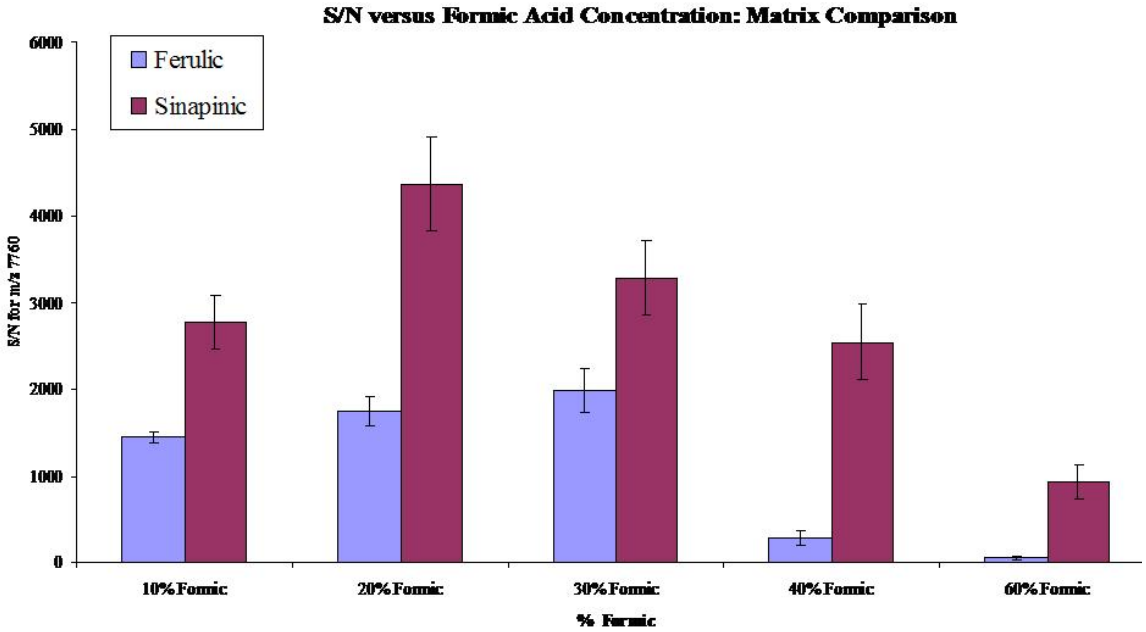

Figure 2-6. Signal-to-noise versus formic acid concentrations from 10 to 60% for the *m/z* 7,760 peak from *B. subtilis* 168. Ferulic acid is shown in blue and sinapinic acid is in red. The error bars represent the standard error across 9 measurements.

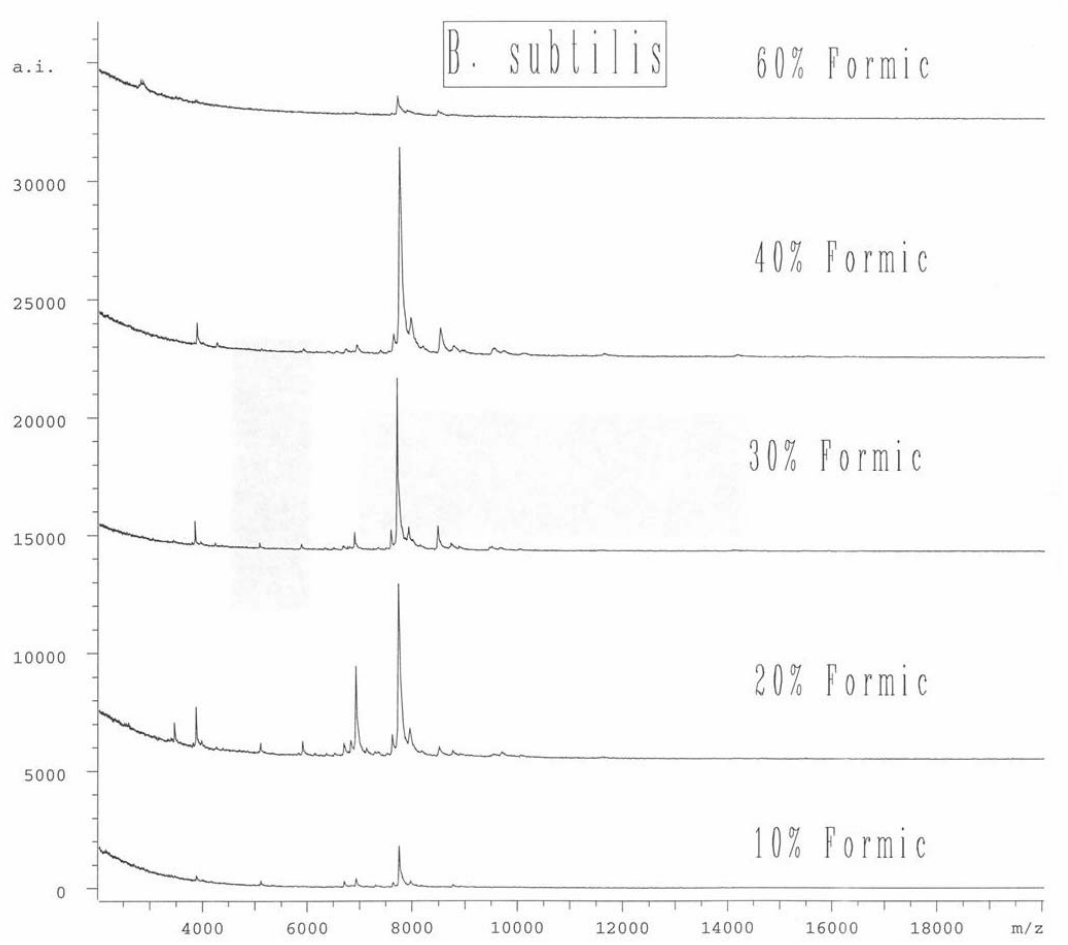

Figure 2-7. Enhancement in biomarker signal for *B. subtilis* 168 by increasing formic acid concentration. From top to bottom, *B. subtilis* in 60%, 40%, 30%, 20%, 10% formic acid. The scale is from *m/z* 2,000-20,000. Note the emergence of higher molecular weight proteins as formic acid concentration increased.

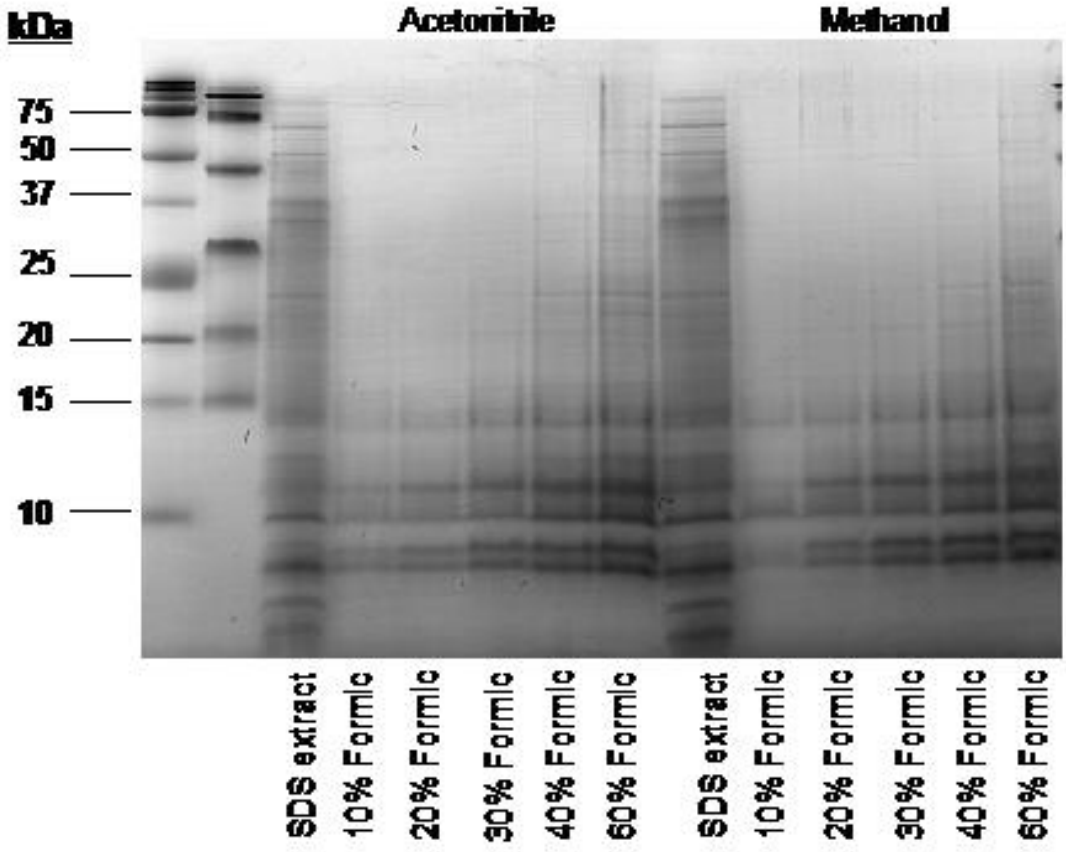

Figure 2-8: 1-D Gel showing the effects of increasing formic acid concentration on extraction of *B. subtilis* 168 spores compared to SDS solublized extract. This is shown for 2 organic solvents, 30% acetonitrile and 30% methanol.

The higher formic acid concentrations also had a significant impact on the biomarkers that were extracted, possibly due to differences in the hydrophobicities of the proteins. This was highlighted when analyzing the FO-36b spore sample. As formic acid concentration increased, the biomarker peak at 7,620 peak was favored over the 7,250 peak (Figure 2-9 and 2-10). This peak was one of the few peaks that differed between spores of FO-36b and *B. pumilus* 7061 and was critical for differentiation of these two strains.

Organic Modifier

The effect of the organic solvent was also investigated. The acetonitrile preparation was compared with methanol, ethanol, and isopropanol. A range of formic acid concentrations was studied with each of the different solvents. Similar results were obtained for all three of the spore lines evaluated. A graph of the results for *B. subtilis* 168 is shown in Figure 2-11 for methanol and isopropanol with acetonitrile for comparison. Results with ethanol were nearly identical to those with isopropanol and are not shown in the graph. Overall, the effect of the organic modifier on the spectra was minimal. Similar trends were noted for the formic acid concentrations as seen with ACN before. The signal increased for samples with up to ~30% formic acid, and then decreased again due to poor MALDI spot formation at higher formic acid concentrations. Methanol shows an advantage in signal-to-noise ratio but, as indicated by the error bars, didn't give results as reproducible as when acetonitrile was used. Isopropanol and ethanol were very similar to acetonitrile in signal-to-noise, although the MALDI spots were less reproducible and tended to spread over the plate. Therefore, acetonitrile was generally used as the organic solvent in our studies.

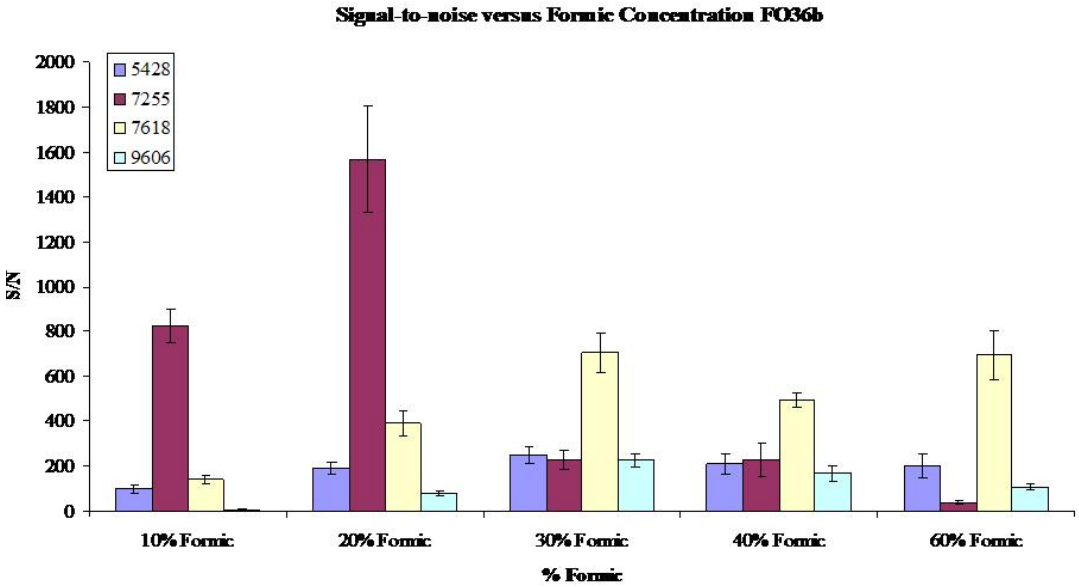

Figure 2-9. Signal-to-noise versus formic acid concentrations from 10-60% for 4 biomarker peaks from FO-36b spores. Ferulic acid was used as a matrix and 30% acetonitrile is the organic solvent. The error bars represent the standard error of 9 measurements.

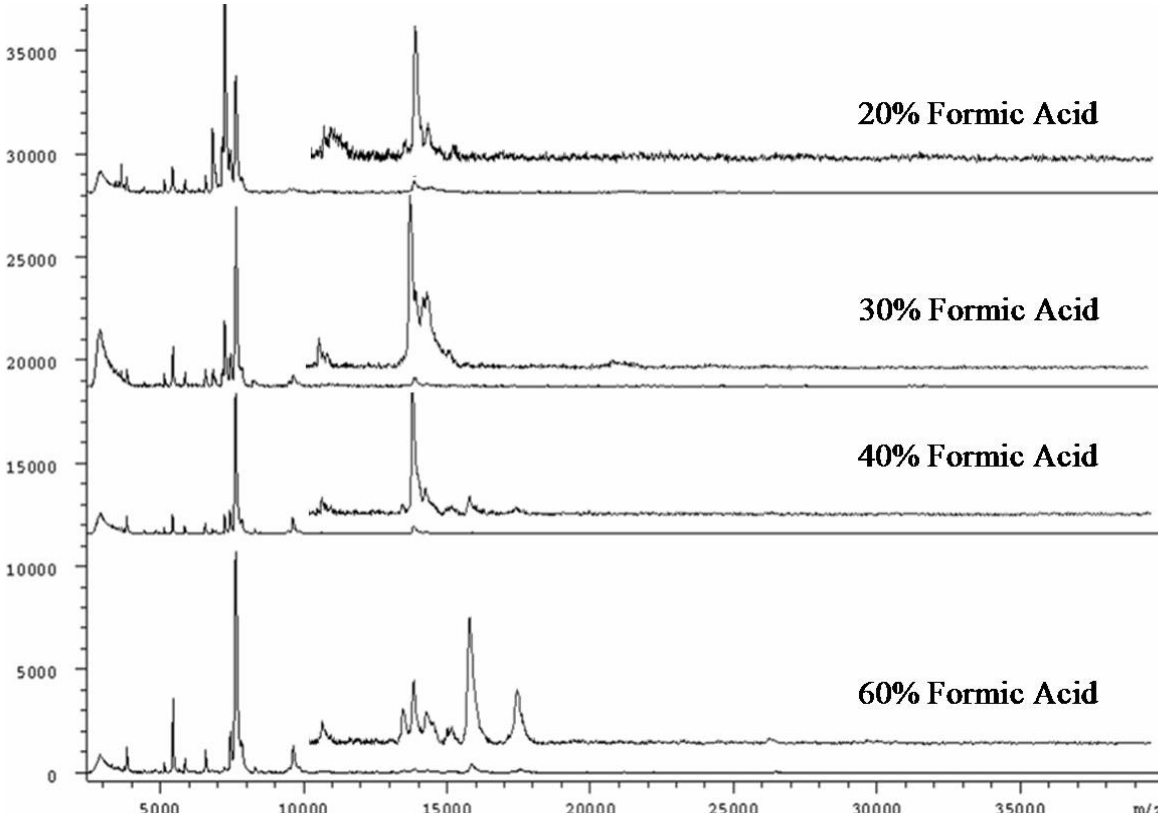

Figure 2-10. Enhancement in biomarker signal for FO-36b spores by increasing formic acid concentration. From top to bottom, FO-36b in 20%, 30%, 40%, and 60% formic acid. Each spectrum is displayed from *m/z* 2,500-40,000. At 20% formic acid the 7,250 Da peak is the base peak in the spectra. As the formic acid concentration increased to 30% and higher, the 7,620 Da peak became the base peak in the spectra.

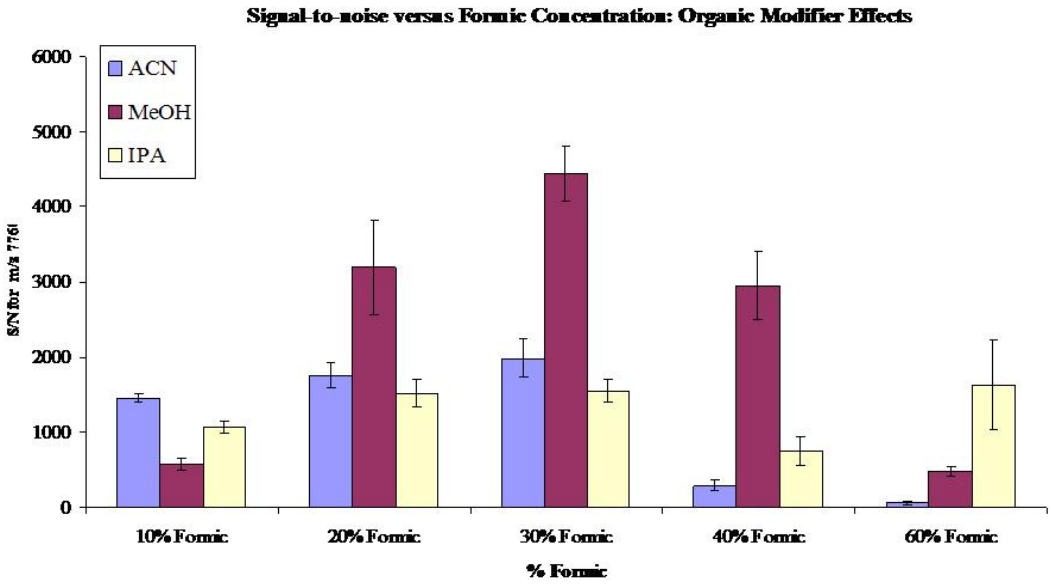

Figure 2-11. Signal-to-noise versus formic acid concentrations: organic modifier effects. Comparison of acetonitrile (ACN), methanol (MeOH), and isopropanol (IPA) as the organic modifier in the MALDI solvent using *B. subtilis* 168 as the sample.

Detergent Additives

Detergent additives were also examined as a mechanism for increasing the number of biomarkers extracted from the spores. Two MALDI compatible detergents were examined, N-octylglucoside (OGP) and the acid labile Rapigest. Sodium dodecyl sulfate (SDS) was not evaluated as a detergent additive because MALDI compatible concentrations of SDS are less than 0.01%. During the purification of a spore preparation, 0.05% SDS was used to clean the spores and did not disrupt the spore coat proteins.[7] Therefore, lower concentrations of SDS would have no effect on protein solubilization. A 0.1% TFA/50% ACN solution was used as the matrix solution instead of formic acid in order to ascertain whether it was the detergents which improved solubilization,.

OGP was added directly to the matrix solution at concentrations ranging from 0.425-68 mM. Figure 2-12 shows sample spectra from *B. subtilis* 168 spores treated with 0.1%TFA/50% ACN matrix preparation alone in A, and with increasing OGP concentrations of 13.6 mM OGP added in B, and 68mM OGP added in C. Spores treated with a 30% formic acid/30% acetonitrile solvent are shown for comparison in D. OGP enhanced the extraction of proteins from the spores as indicated by the higher signal-to-noise in the spectra with OGP added. However, even with the highest level of OGP, this enhancement was still lower than the formic acid treatment (D). The OGP also did not allow for the detection of the higher molecular weight peaks in the spectra.

Rapigest was also evaluated at a concentration of 0.1% and was used per the manufacturer's instructions. This involved boiling the spore sample in the detergent and then adding 50 mM hydrochloric acid to degrade the Rapigest. As a control, a spore sample in water was also boiled and added to the formic acid matrix. The Rapigest

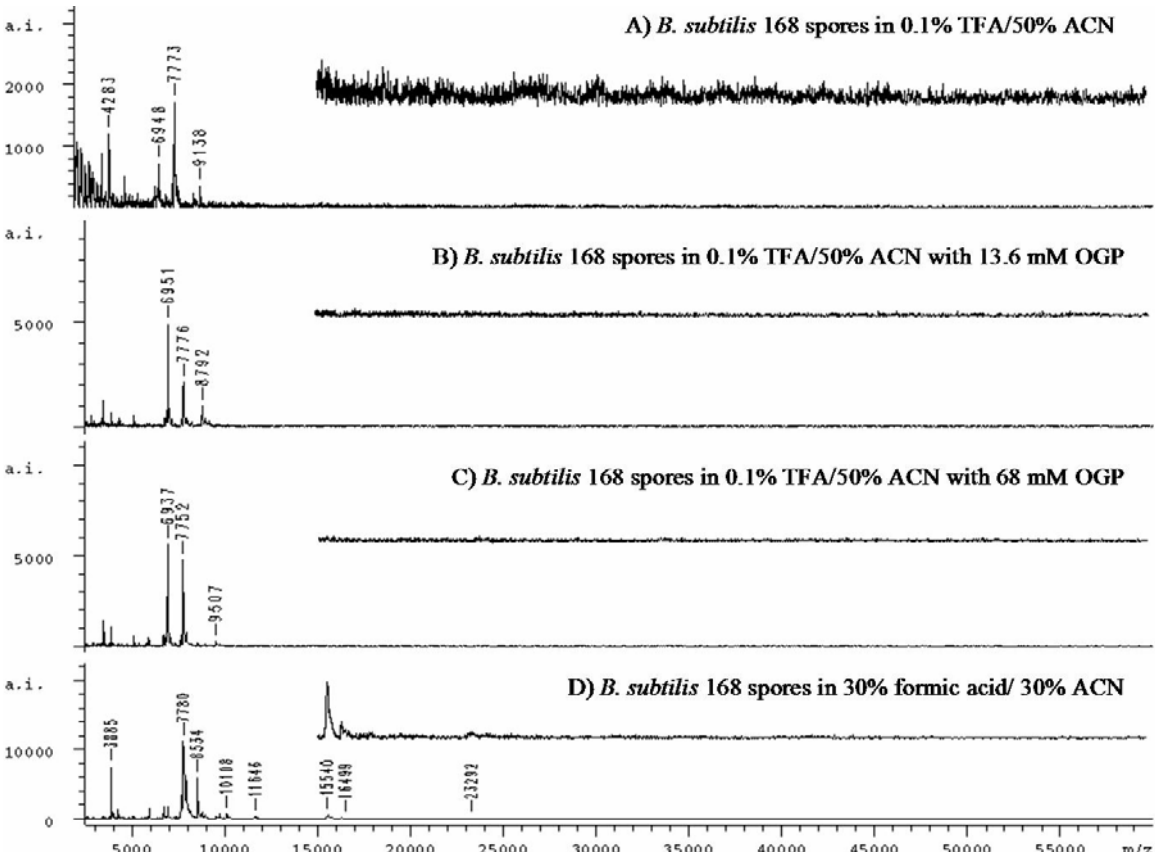

Figure 2-12. Treatment of spores with OGP detergent. A) *B. subtilis* 168 spores in 0.1% TFA/50% ACN. B) *B. subtilis* 168 spores in 0.1% TFA/50% ACN with 13.6 mM OGP. C) *B. subtilis* 168 spores in 0.1% TFA/50% ACN with 68 mM OGP. D) *B. subtilis* 168 spores in 30% formic acid/30% acetonitrile. The OGP improved extraction from the spores; however the formic acid treatment was still superior.

treatment significantly altered the visible biomarkers spectra (Figure 2-13, top). The dominating biomarker became the 9,130 peak, which had not been previously observed. This spectrum is similar to that obtained when concentrated TFA or HCl treatments were used to lyse spores for SASP extraction, and these peaks correspond to the molecular weights of the major SASPs in *B. subtilis* 168. Since Rapigest required degradation with high acid concentrations prior to analysis, the SASPs would be expected to dominate the spectra. In the formic acid matrix, the boiled spores gave rise to an even higher molecular weight protein at ~43 kDa that had not been previously discernable, indicating that a boiling step might aid in the solubilization of additional proteins.

Characteristics of Optimized Solvent Extraction System

The best combination of solvents evaluated for the analysis of spore was a 30% formic acid/30% ACN solvent system with ferulic acid as the matrix. This solvent system represented a compromise between signal-to-noise, reproducibility, and the availability of a wide range of low and high molecular weight biomarkers for analysis. Higher formic acid concentrations allowed for the extraction of higher molecular weight proteins; however, inconsistencies in the MALDI spot formation were detrimental to analysis. At a concentration of 30% formic acid, the MALDI spot formation was homogeneous and consistently gave a good crystal layer for the analysis. This treatment was rapid and did not require any additional sample preparation or spot treatment. It was also relatively inexpensive in comparison to the use of detergents such as Rapigest and OGP.

Unlike other sample preparation procedures, such as treatments for SASP extractions, in which the spores were not viable post-treatment, the *B. subtilis* 168 and FO-36b spores in our studies remained viable for up to 1 hour in 30% formic/30%

acetonitrile. After 1 hour in the solvent there was a 2 decade reduction in growth. In contrast, spores from *B. pumilus* 7061 were affected by the formic acid treatment and there was a 2 decade reduction in growth after only 10 minutes of treatment. The difference in viability between the three spore strains might be explained by differences in the permeability of the spore coats by the MALDI solvent. This could also be a result of storage conditions and storage times.

Limit of Detection Study for MALDI Spore Preparations

Although a limit of detection for bacterial cells has been reported in the literature (5,000 cells/spot), no comprehensive study has been compiled. To assess a more realistic limit of detection for whole-cell analysis of spores by MALDI, an experiment was designed to determine the minimum number of spores necessary to obtain useful spectra. Obtaining a useful spectrum depends on a number of factors including the signal-to-noise ratio and the number of biomarker peaks discernable in the spectrum. Dilutions of the spore suspension were made in water and the resulting solutions were mixed with the ferulic acid matrix described above. Figure 2-14 is an overlay of the spectra collected for the spore sample at each of the dilutions in the series. The results from the spore samples show a drastic decrease in spectral information very early in the dilution series. Nearly all spectral information is lost when there are fewer than 50,000 cells on the spot. The peak at approximately 7,760, the most abundant biomarker for the spores, remains barely discernable at 5,000 cells/spot; and no other peaks are seen in the spectrum. The ability to identify the spores at lower concentrations will be dependent on the statistical approaches used.

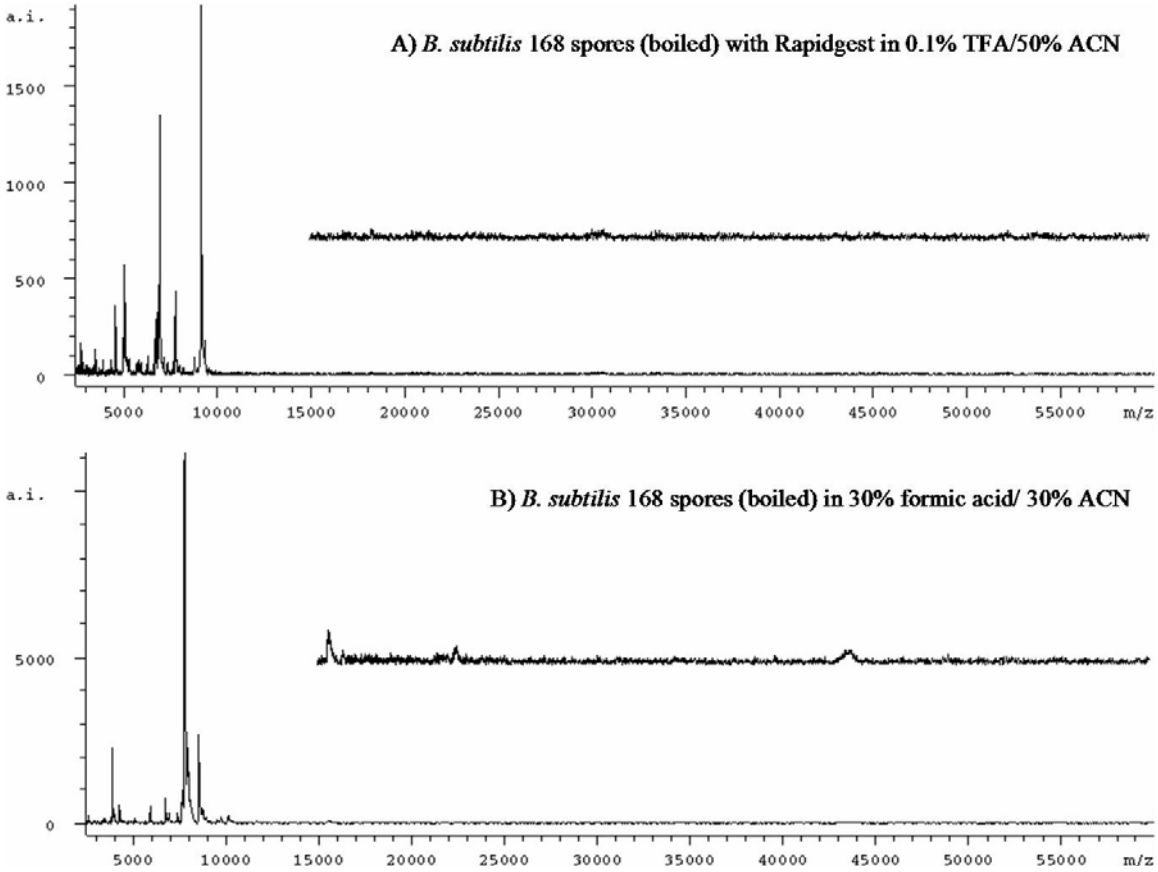

Figure 2-13. Treatment of spores with Rapigest detergent. A) *B. subtilis* 168 spores in boiled with Rapigest and analyzed in 0.1% TFA/50% ACN matrix solvent. B) *B. subtilis* 168 spores boiled in water and analyzed in 30% formic/ 30% ACN. The Rapigest treatment resulted in the release of SASP proteins from the spores. With the additional boiling step, a 43 kDa protein was also extracted with the formic acid treatment.

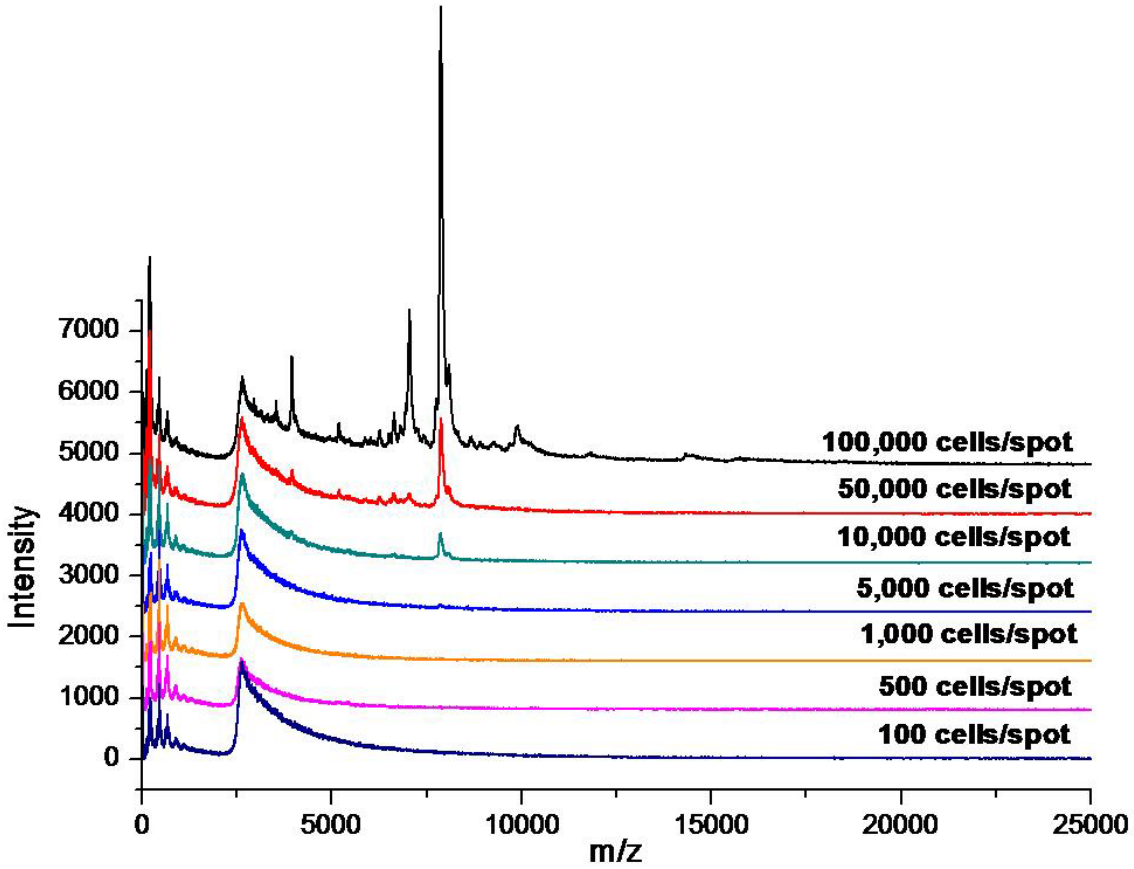

Figure 2-14. Limit of detection for *B. subtilis* 168 spores. The dilution series runs from 100,000 cells/spot to 100 cells/spot. At 5,000 cells/ spot the 7,760 Da biomarker is barely discernable in the spectra.

CHAPTER 3
SPECIES DIFFERENTIATION OF A DIVERSE SUITE OF *BACILLUS* SPORES AND CELLS WITH MASS SPECTROMETRY BASED PROTEIN PROFILING

Introduction

In order to overcome the problems involved with phenotypic characterization, 16S ribosomal RNA (16S rDNA) analysis has been used for decades to more accurately define the phylogenetic affiliation of the given test microorganism.[53] However, being highly conserved, the 16S rDNA molecule at times cannot differentiate closely related microbial species.[54,55] Therefore, alternative biomarkers[56] or a suite of protein profiling methods would be useful to effectively differentiate closely related microbial species.

In this chapter, the versatility of Matrix-Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry (MALDI-TOFMS) protein profiling for the species differentiation of a diverse suite of *Bacillus* cells and spores is demonstrated. MALDI-TOFMS protein profiles of fourteen different strains of *Bacillus*, encompassing eleven different species, were evaluated. *Bacillus* species selected for MALDI-TOFMS analysis represented the spore-forming bacterial diversity of typical class 100K clean-room spacecraft assembly facilities.

A majority of the MALDI-TOFMS research directed at *Bacillus* has focused on only a few spore species. These include *B. anthracis* and its closely related species *B. thuringiensis, B. cereus,*[55] *B. atrophaeus* (formally called *B. globigii*)[57], an anthrax surrogate, and *B. subtilis*, whose genome has been completely sequenced and has been thoroughly examined by molecular biological methods.[25,26,38,58] Very little research

attention has been given to other *Bacillus* species, which naturally occur in the environment. The nonpathogenic *Bacillus* spores, which are ubiquitous in the environment, are the most likely source of interference for any detection technique and have the highest potential to produce false positives.

To demonstrate the versatility of MALDI-TOFMS protein profiling for the identification of a variety of spores and cells, a subset of *Bacillus* species isolated from various NASA spacecraft assembly facilities (class 10 to 100K clean rooms) was used in this study. The optimized one-step sample treatment and MALDI-TOFMS preparation was used to obtain spectra rapidly with a wide range of protein biomarkers, including several higher molecular weight (10-25kDa) protein species for the spores. A library of MALDI-TOFMS spectra was created from the 16 different spores and vegetative cells of the *Bacillus* species, the most diverse study of the genus reported to date. Linear correlation analysis was used to identify all *Bacillus* species evaluated. The results obtained from MALDI-TOFMS protein profiling of these *Bacillus* species were compared with 16S rDNA sequences for their bacterial systematics and molecular phylogenetic affiliations.

Materials and Methods

Bacterial Strains

Bacillus strains used in this study and their source are listed in Table 3-1. Fourteen strains consisting of 11 *Bacillus* species were studied. The type strains of *B. atrophaeus, B. licheniformis, B. megaterium, B. mojavensis, B. thuringiensis, B. pumilus,* and *B. subtilis* were procured from the American Type Culture Collection (ATCC, Manassas, Virginia). *B. subtilis* 168 was received from Wayne Nicholson, Univ. of Arizona and the *B. anthracis* 34F2 vaccine strain was from M. Satomi, National Institute of Fisheries,

Japan. *B. odysseyi*, *B. licheniformis* KL-196, *B. niacini* 51-8C, *B. megaterium* FO-38, and *B. psycrodurans* were isolated from several NASA spacecraft and assembly facilities surfaces. Bacterial isolation procedures from spacecraft and assembly facilities surfaces were described elsewhere[59,60]. Identity of the test organisms was determined based on 16S rDNA sequencing for the environmental isolates; for the ATCC strains, those sequences available in the GenBank database were used[61]. The 16S rDNA sequences of the environmental isolates have been deposited in the GenBank nucleotide sequence database.

Table 3-1. List of *Bacillus* species used in this study

Name	Strain Number	Source	Remarks
B. anthracis	34F2	Inst. of Fisheries, Japan	Vaccine strain
B. atrophaeus	9372	ATCC	Surrogate to *B. anthracis*
B. licheniformis	14580	ATCC	Most predominate species in clean room facilities
B. licheniformis	KL-196	JPL-SAF	Class 100K clean room floor, JPL
B. megaterium	14581	ATCC	
B. megaterium	FO-38	JPL-SAF	Clean room air particulate
B. mojavensis	51516	ATCC	
B. niacini	51-8C	KSC, SAEF-II	Mars Odyssey assembly facility floor
B. odysseyi	34hs1	KSC, SAEF-II	Mars Odyssey spacecraft surface
B. psycrodurans	VSE1-06	KSC, PHSF	Mars Exploration Rover assembly facility air particles
B. pumilus	7061	ATCC	Second most predominate species in clean room facilities
B. subtilis	168	University of Arizona	Genome fully sequenced
B. subtilis	6051	ATCC	Type species of *Bacillus* genus
B. thuringiensis	10792	ATCC	Insecticide producing bacteria and phylogenetically unseparable from *B. anthracis*

Abbreviations: ATCC, American type culture collection; SAF, Spacecraft assembly facility; SAEF-II, Spacecraft assembly and encapsulation facility-II; PHSF, Payload handling and spacecraft assembly facility; JPL, Jet Propulsion Laboratory; KSC, Kennedy Space Center

Sporulation of Bacillus isolates

A nutrient broth sporulation medium (NSM) was used to produce spores.[7,62] A single purified colony of the strain to be sporulated was inoculated into the NSM liquid medium. After 1 to 3 days of incubation at 32^{o}C under shaking conditions, cultures were

examined using phase-contrast microscopy to determine the level of sporulation. Microcosms that attained >99% of spores were further purified to remove vegetative cells or cell debris as previously reported.[7] The purified spores were suspended in sterile deionized water and stored at 4°C in glass tubes until analyzed. Before the analysis, spore suspensions were adjusted to give an optical density of 0.6 at 600 nm, which resulted in suspensions that were between 10^8 to 10^9 spores/mL.

Preparation of Vegetative Cells

A stock culture of each *Bacillus* species was streaked for isolation on tryptic soy agar (TSA) plates. *B. anthracis* 34F2, *B. subtilis* 168, *B. pumilus* 7061, and *B. thuringiensis* were also streaked on nutrient agar (NA) plates and Luria-Bertani (LB) plates for a study of the effect of different growth media on the spectra. The plates were incubated at 32°C for 16 hours except in the case of the media study where the plates were incubated for 24 hours. Single purified colonies were removed from the plate with a sterile loop and were placed in 100 µL of a phosphate buffered saline (PBS) solution. Most colonies were approximately 2 mm in diameter. If larger colonies were present only a 2 mm portion was removed for washing and analysis. The cells were vortexed in PBS for 15 minutes and then were pelleted by centrifugation for 10 minutes at 9600 x g. The supernatant was removed and the cell pellet was used for subsequent analysis.

Sample Preparation for Mass Spectrometry

A saturated matrix solution was prepared by dissolving 20 mg of ferulic acid into a 1 mL solution of 30% acetonitrile, 30% formic acid. As described in Chapter 2, this solvent system was selected due to the higher signal-to-noise, consistent crystallization, and better ability to differentiate across the various bacterial species. This effect was due to a combination of an increased number of biomarker peaks and the higher molecular

weight range of these peaks in the spectra.[63,64] A 2.5 µL aliquot of the spore suspension (0.6 OD_{660}) was added to 22.5 µL of the matrix solution. This mixture was vortexed briefly and then 1 µL of the sample containing both spores and matrix compound was removed and spotted on a SCOUT26 MALDI plate (Bruker Daltonics; Billerica, Ma). For the preparation of vegetative cells, 25 µL of the matrix solution was mixed directly with the cell pellet. This solution was sonicated for 3 minutes and then vortexed for 3 minutes. A 1 µL aliquot of the vegetative sample was then placed on the MALDI plate for analysis. Spots were allowed to air dry. No further treatments were applied to the spots once dried. Spots were prepared in duplicate for each sample mixture. Sample preparation required only a few minutes per sample.

Mass Spectrometry Analysis

MALDI-TOFMS analysis was performed on a Bruker Daltonics Reflex II Mass Spectrometer (Bruker Daltonics, Billerica, Ma) retrofitted with delayed extraction. The instrument was operated in the linear mode. A nitrogen laser (337 nm) pulsed at a frequency of 5 Hz irradiated the sample. Spectra were obtained in positive ion mode with a delay time of 50 ns. The acceleration voltage was 20 kV. An ion deflector was used to deflect low mass ions that would saturate the detector. The deflector was set at 2,500 Da. The laser intensity was adjusted to just above the threshold for ion formation for each sample. The instrument was calibrated daily using external calibration with a mixture of bovine insulin and equine cytochrome C. All spectra represent the accumulation of 50 laser shots. Ten spectra were collected from each spot on the MALDI plate. A total of 20 spectra were collected per sample.

Spectral Processing and Statistical Methodology

Prior to statistical analysis, each spectrum was baseline-corrected and smoothed using a ten point Savitzky-Golay smoothing algorithm. Normalized spectra were converted into ASCII files for statistical processing. Because linear correlation is invariant with respect to a linear transformation of spectra, the relative, not absolute, intensities were important for correlation analysis. Statistical analysis of the data was performed using linear correlation software developed in house using Visual Basic 6.0.[65-67] Spectra from the mass spectrometer were imported into the software as ASCII files and libraries were created using the average of the 20 spectra collected per sample (10 spectra per spot). Correlation analysis was performed on a point-to-point basis based on the following equation for the Pearson correlation coefficient, r:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}}$$

where $\bar{x}$ is the mean of x_i's and $\bar{y}$ is the mean of y_i's. The x_i's and y_i's are the intensities at the i-th pixel of the detector which in this case corresponds to the *m/z* (i=1….N; for the *m/z* range 2,500-60,000 N is approximately 16,000 points). The x_i's, belong to an analyzed spectrum, and the y_i's belong to one of the library spectra. The spectrum consisting of x_i's is correlated against each spectrum in the library (different sets of y_i's) and the closest match with the highest correlation coefficient indicated a similarity of this spectrum with the corresponding library spectrum. Conversely, the difference between this and other correlation coefficients signified spectral dissimilarities. To quantify the level of significance of these differences, a Student's t-

test was applied. Student t values were calculated differently depending on whether the two distributions had the same or different variances. To check this, an F-test was applied (F denoting the ratio of the variances) as the basis of t-values. The probabilities that two distributions of correlation coefficients had different means were calculated.

A reference library of spores and vegetative cells, consisting of the average spectrum created from the 20 spectra collected for each sample, was produced for all of the fourteen species evaluated in this study. The individual spectrum and the average spectrum obtained for each of the 14 strains were then compared to the MALDI-TOFMS spectra stored in the library to elucidate the bacterial speciation. To evaluate the reproducibility of the technique, a separate set of MALDI-TOFMS spectra were collected and averaged for all of the different species of spores in this study. The averages of these separate analyses were compared with the library spectra. To address batch-to-batch variability, *B. subtilis* 168 spore cultures prepared at different times over the course of two years were analyzed and compared to the library spectra. In the case of vegetative cells, colonies from 3 different agar plates were analyzed to ascertain the effect of different growth media and incubation times on the spectra.

In conjunction with the correlation analysis, hierarchal cluster analysis (HCA) was used to help visualize and categorize the different species. The HCA analysis was performed using the commercially available statistical software SPSS (Chicago, IL). Dendrograms were produced based on the Pearson correlation value between the spectra using the nearest neighbor method (single linkage). To help visualize the peak patterns for the spectra, Surfer 8.0 from Golden software (Golden, CO) was used to create an image map with 10 Da resolution from the average spectra for each species. The image

map provided a 3D representation of the spectra where the color of the band was indicative of peak intensity.

Results and Discussion

Incidence of Spore-Forming Microbes from Spacecraft Associated Environments

Among several hundred aerobic spore-forming bacteria isolated from several spacecraft and associated facility surfaces, >90% of the isolates were found to be phylogenetically affiliated with the members of the genus *Bacillus.* [59-61] *B. licheniformis (*25%) and *B. pumilus (*16%) were the most prevalent *Bacillus* species isolated. Since *B. licheniformis* was the most prevalent *Bacillus* species in the environment and *B. subtilis* is the type species of the *Bacillus* genus, multiple strains of these species were included in this study. An additional wild-type strain of *B. megaterium* was included as well. To avoid confusion about the identity of the bacterial species, wherever possible, authentic type strains were procured from the culture collection and used. All tested *Bacillus* species fall into the RNA group I except *B. psychrodurans* and *B. odysseyi,* which are in RNA group II.[68,69] Group I includes aerobic *Bacillus* species that produce acid from a variety of sugars including glucose and whose spores are ellipsoidal and do not swell the mother cell. Group II *Bacillus* species are also aerobic; however, they do not produce acids from sugars and even though they also produce ellipsoidal spores, they swell the mother cell. As the *Bacillus* species of other rDNA groups were not isolated from class 100K clean-room facilities,[59-61,70] the characterization of the species by MALDI-TOFMS was restricted to the sixteen members of these two rDNA groups.

Molecular Phylogeny of Spore-Forming Microbes

The sequence similarities based on 16S rDNA sequences of the various *Bacillus* species tested are shown in Table 3-2. These sequences were either obtained from the

GenBank database or were sequenced in previous studies.[59-61,70] The similarities in 16S rDNA nucleotide sequence between the tested *Bacillus* species, recognized by GenBank "BLAST" searches, were between 91 and 99%. A sequence variation of ~9% was found between rDNA groups 1 and 2 *Bacillus* species. A very high sequence variation within a well-described genus is not uncommon. Further analyses indicated that *B. atrophaeus* shares a close phylogenetic relationship with several *Bacillus* species such as *B. mojavensis, B. pumilus* and *B. subtilis (*>97.5%). Similarly, *B. licheniformis* wild-type

Table 3-2. 16S rDNA sequence similarities for the various *Bacillus* species studied

Bacteria	*B atrophaeus* X60607	*B licheniformis* AF387515	*B licheniformis* X68416	*B megaterium* X60629	*B mojavensis* AB021191	*B odysseensis* AF526913	*B psychrodurans* VSE1 06	*B pumilus* AB020208	*B subtilis 1* 68 rrnA	*B subtilis* X60646	*B thuringiensis* X55062
***B atrophaeus* X60607**	**100**										
***B licheniformis* AF387515**	**96.9**	**100**									
***B licheniformis* X68416**	**98.5**	**98.3**	**100**								
***B megaterium* X60629**	**94.4**	**92.7**	**94.1**	**100**							
***B mojavensis* AB021191**	**99.3**	**96.7**	**98.4**	**94.1**	**100**						
***B odysseyi* AF526913**	**92.0**	**90.1**	**91.5**	**93.4**	**91.8**	**100**					
***B psychrodurans* VSE1 06**	**91.8**	**90.5**	**91.5**	**92.9**	**91.5**	**95.4**	**100**				
***B pumilus* AB020208**	**97.6**	**94.9**	**96.3**	**94.3**	**96.9**	**91.8**	**92.4**	**100**			
***B subtilis* 168 rrnA**	**99.4**	**96.9**	**98.6**	**94.1**	**99.7**	**91.6**	**91.4**	**97.2**	**100**		
***B subtilis* X60646**	**99.3**	**96.7**	**98.3**	**94.1**	**99.6**	**91.6**	**91.2**	**96.9**	**99.8**	**100**	
***B thuringiensis* X55062**	**95.2**	**92.9**	**94.2**	**94.7**	**94.3**	**92.8**	**92.0**	**94.3**	**94.2**	**94.3**	**100**

strain KL-196 and B. mojavensis, as well as two *B. subtilis* strains tested in this study showed >98% 16S rDNA sequence similarities. Such high 16S rDNA sequence similarities was also noticed (>99%) in the case of the two *B. subtilis* strains sequenced and B. mojavensis. This clearly showed that 16S rDNA sequence analysis was not useful in differentiating these closely-related species of the genus Bacillus. The species

identities of all these strains were confirmed by DNA-DNA hybridization (M. Satomi, personal communication). The two strains of *B. licheniformis* and *B. subtilis* showed >70% DNA-DNA hybridization dissociation values and exhibited >98.5% 16S rDNA sequence similarities. When all these species were grouped together, the maximum-likelihood based phylogenetic tree showed two major clusters (M. LaDuc, K. Venkateswaran, personal communication). One cluster consists of *B. megaterium, B. odysseyi, B. psychrodurans,* and *B. thuringiensis,* where the spores of these species contained an additional structure called exosporium around the spore outer coat. The second cluster formed by the other species tested did not contain an exosporium.

MALDI-TOFMS Spore Profiles

A representative spectrum from each *Bacillus* species analyzed in this study is shown in Figure 3-1 A-N. The mass spectra are presented with *m/z* values from 3,000-25,000. The *m/z* region from 9,500-25,000 is amplified (see inset of each spectrum) to aid in visualization of the less abundant peaks present at higher *m/z*. The observation of proteins at higher *m/z* is seldom reported in other MALDI-TOFMS analyses of whole spores.[25,26,34,35,38] We hypothesize that the appearance of large proteins at high *m/z* is due to optimization of the solvent system used in this study.

From the spectra, we were unable to identify an obvious *Bacillus*-ubiquitous biomarker with the sample preparation protocol adapted in this study. A peak at 14,500 *m/z* was present in most of the spore spectra obtained except for that of the *B. licheniformis* ATCC 14580 type strain, its wild-type strain KL-196, and *B. anthracis* 34F2. The absence of a genus specific biomarker might be due to the extraction protocol used in this study, post translational modifications of proteins that may differ between the strains, or the need for more sophisticated spectral comparisons of the different species.

All of the spores have a group of peaks in the *m/z* region from 6,500-8,000. *B. licheniformis* ATCC 14580, *B. licheniformis* KL-196, *B. psychrodurans*, *B. odysseynsis* and *B. megaterium* ATCC 14581, and *B. megaterium* FO-38 all have an additional group of peaks between *m/z* 5,000-6,500 that was not observed in the other spectra. It was challenging to obtain good spectra from the *B. odysseyi* samples as shown by the lower signal-to-noise in the spectra. This could have been a result of glycoproteins present in the exosporium layers. Glycoproteins can be challenging to analyze due to difficulty in the ionization of the sugar moieties and the inherent heterogeneity of glycosolations. An expected result was the level of similarity between the strains of the same species. *B. licheniformis* ATCC 14580 type strain (Figure 3-1C) and its wild-type strain KL-196 (Figure 3-1B), *B. subtilis* 168 (Figure 3-1I) and ATCC 6051 (Figure 3-1J), and *B. megaterium* ATCC 14581 (Figure 3-1D) and *B. megaterium* FO-38 (Figure 3-1M) have very comparable MALDI-TOFMS profiles upon visual inspection. The spectra for the B. *licheniformis* pair were very similar except for a difference in intensity of the *m/z* 7,260 peak and the presence of different higher molecular mass species in *B. licheniformis* 14580. The *B. subtilis* pair has the same pattern in that there was a difference in peak intensity for the peak at *m/z* 6,936 and variation in the masses observed above *m/z* 10,000. Similar patterns were also observed in the *B. megaterium* pair. This observation supports the theory that it is important to examine a wide variety of *Bacillus* spores before assigning definitive genus, species, and strain specific protein biomarkers.

Linear correlation analysis provided a means of statistical comparison of the spectra. Correlation values close to 1 indicate that the fingerprint patterns of two organisms are very similar. Table 3-3 shows the linear correlation values for the

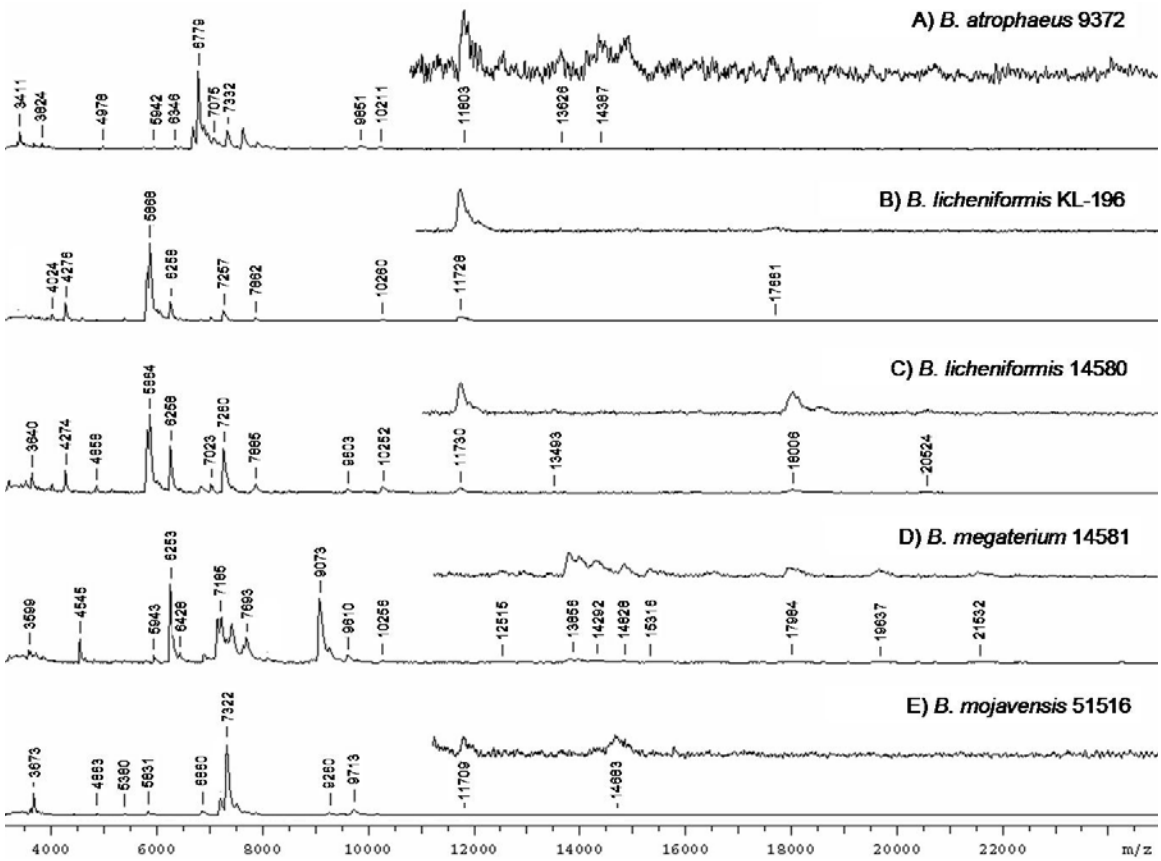

Figure 3-1. MALDI-TOFMS protein profiles of the 14 *Bacillus* spore species analyzed in this study. The mass range depicted is from *m/z* 3,000-25,000. The higher molecular mass region from *m/z* 9,500-25,000 is amplified 4x (see inset of each spectrum) in order to visualize the higher molecular weight peaks that are present but are at much lower abundance in the samples. A) *B. atrophaeus* ATCC 9372. B) *B. licheniformis* KL-196. C) *B. licheniformis* ATCC 14580. D) *B. megaterium* ATCC 14581. E) *B. mojavensis* ATCC 51516. F) *B. odysseyi* ATCC PTA-4993. G) *B. psycrodurans* VSE1-06. H) *B. pumulis* ATCC 7061. I) *B. subtilis* 168. J) *B. subtilis* ATCC 6051. K) *B. thuringiensis* ATCC 10792. L) *B. anthracis* 34F2. M) *B. megaterium* FO-38. N) *B. niacini* 51-8C.

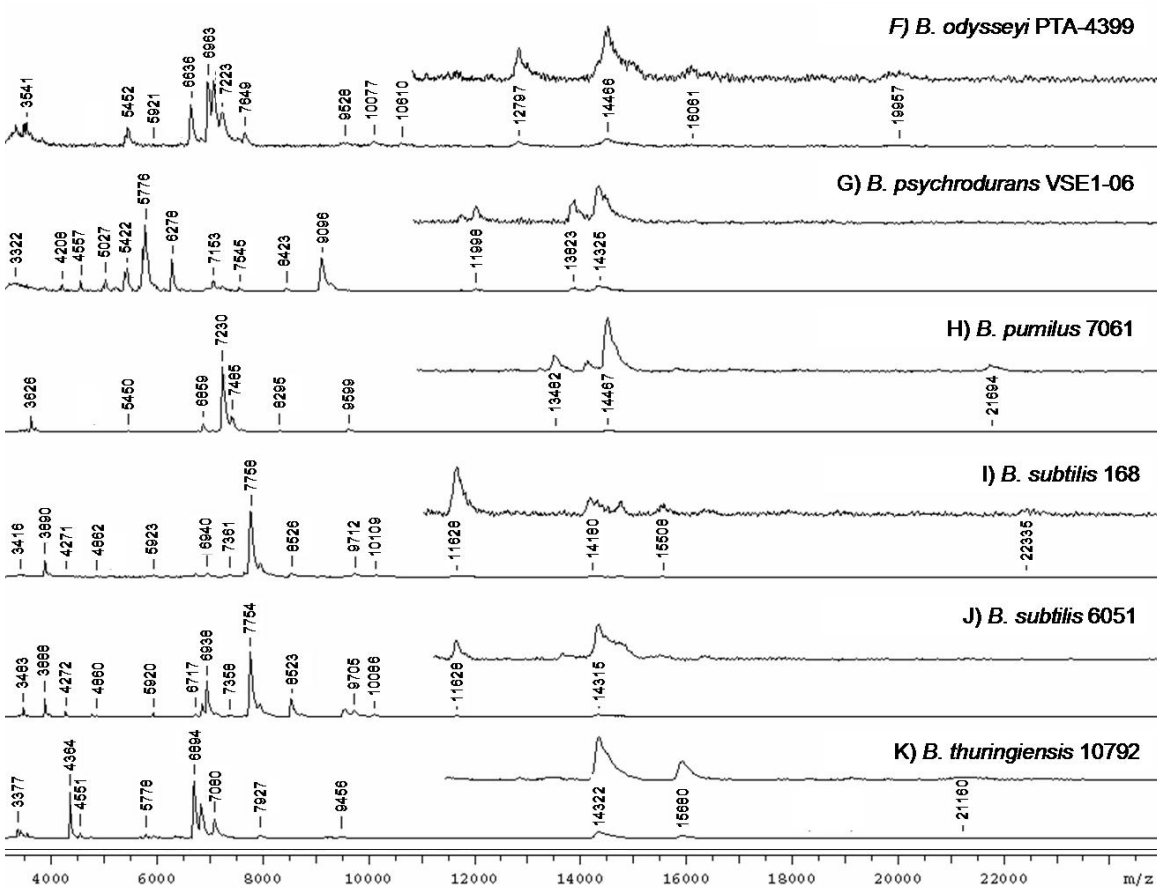

Figure 3-1. Continued.

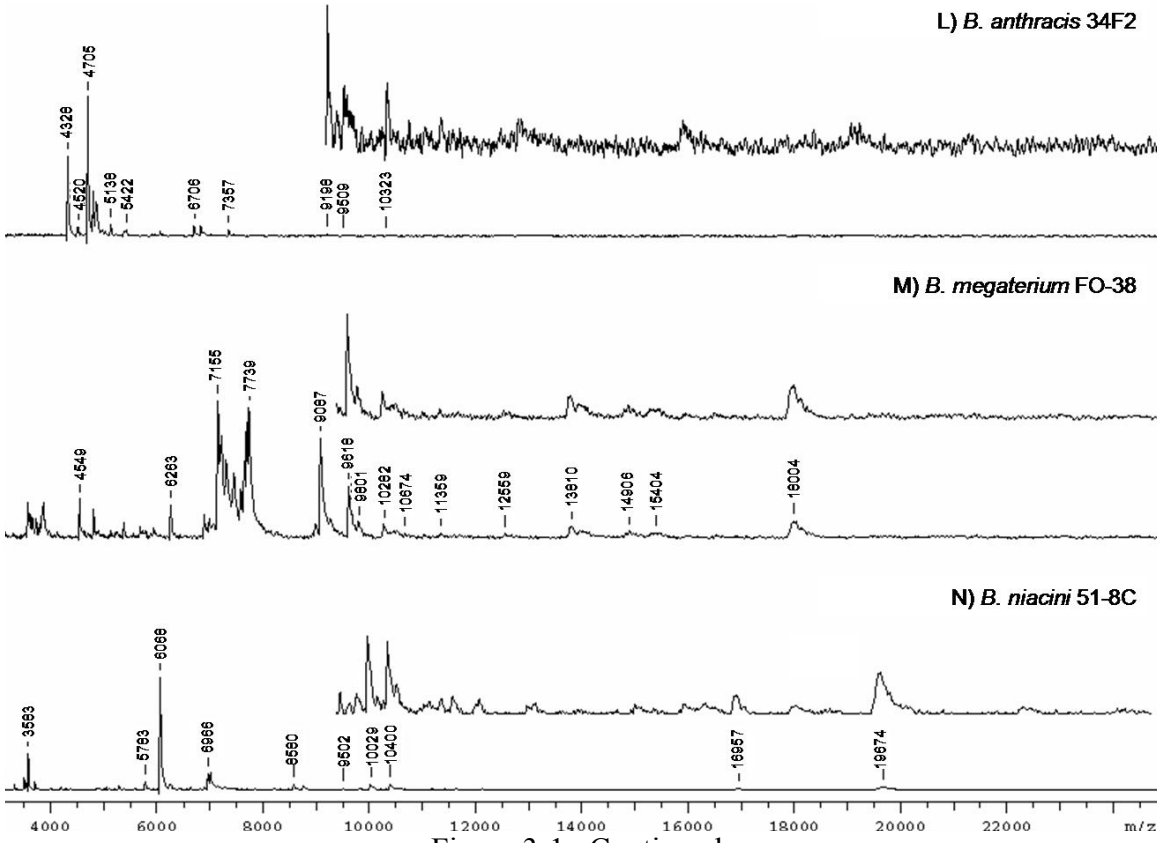

Figure 3-1. Continued.

MALDI-TOFMS spectra of the various *Bacillus* spore species evaluated when compared to the library spectra. Each of the 20 individual spectra from each species was searched against the user generated average library spectra. All individual spectra were successfully identified as their corresponding species and strain. These results were verified by applying a Student's t-test to the data. Using the t-test, we confirmed that we could differentiate all the species studied at the 95% confidence level. Figure 3-2 shows the correlation results of the 20 individual *B. atrophaeus* spectra when searched against

Table 3-3. Correlation values based on MALDI-TOFMS protein profiling of the spores of the *Bacillus* species in this study

Bacterial spores		34F2	ATCC 9372	ATCC 14580	KL-196	ATCC 14581	FO-38	ATCC 51516	51-8C	34hs1	VSE1-06	ATCC 7061	168	ATCC 6051	ATCC 10792
		B. anthracis	*B. atrophaeus*	*B. licheniformis*	*B. licheniformis*	*B. megaterium*	*B. megaterium*	*B. mojavensis*	*B. niacini*	*B. odysseyi*	*B. psychrodurans*	*B. pumulis*	*B. subtilis*	*B. subtilis*	*B. thuringiensis*
B. anthracis	34F2	1													
B. atrophaeus	ATCC 9372	0.05	1												
B. licheniformis	ATCC 14580	0.01	0.09	1											
B. licheniformis	KL-196	0.01	0.03	0.90	1										
B. megaterium	ATCC 14581	-0.01	0.14	0.30	0.09	1									
B. megaterium	FO-38	0.00	0.16	0.14	0.00	0.74	1								
B. mojavensis	ATCC 51516	0.01	0.20	0.23	0.07	0.37	0.45	1							
B. niacini	51-8C	0.02	0.05	0.13	0.12	0.06	0.03	0.02	1						
B. odysseyi	34hs1	0.01	0.35	0.15	0.05	0.26	0.28	0.21	0.17	1					
B. psychrodurans	VSE1-06	0.00	0.04	0.45	0.42	0.35	0.19	0.05	0.06	0.16	1				
B. pumulis	ATCC 7061	0.00	0.14	0.27	0.04	0.45	0.48	0.44	0.02	0.34	0.07	1			
B. subtilis	168	-0.01	0.09	0.04	0.01	0.11	0.41	0.07	-0.01	0.05	-0.01	0.02	1		
B. subtilis	ATCC 6051	0.01	0.23	0.05	0.01	0.11	0.36	0.07	0.03	0.29	0.01	0.05	0.88	1	
B. thuringiensis	ATCC 10792	0.11	0.52	0.06	0.03	0.07	0.05	0.05	0.03	0.33	0.08	0.09	0.02	0.13	1

the library spectra. The y axis represents the linear correlation values obtained and the x-axis represents the 1^{st}-5^{th} ranks (hits) from the library. At each rank, the standard deviation of the measurement across the 20 spectra is represented by the error bars. The graph demonstrates that for rank 1 (*B. atrophaeus*), we have very high correlation values (0.96±0.02). For the next best hit, *B. thuringiensis,* the correlation values are much lower (0.51±0.02). Since none of the correlation values approach the *B. atrophaeus* hit, we can confirm the differentiation of *B. atrophaeus* from all of the other strains in the library.

The linear correlation method applied here also allows for differentiation of the species whose MALDI-TOFMS profiles are almost indistinguishable upon visual inspection, including the type strain and wild-type strains of *B. subtilis, B. megaterium,* and *B. licheniformis*. Figure 3-3 shows the correlation results of *B. subtilis* 168 versus the library spectra as described above. The 2nd rank (or hit) is much closer than in the case of *B. atrophaeus*, the values for the first rank are 0.98±0.02 and the second rank are 0.86±0.02. The second rank represents *B. subtilis* 6051, the other *B. subtilis* strain in this study. With statistical treatment of the data, the 2 strains were still able to be differentiated at the 95% confidence interval. The close correlation values of 0.88±0.02 for the *B. licheniformis* pair and 0.86±0.02 for the *B. subtilis* pair illustrate that close correlation values indicate a relationships between the organisms. However, with statistical treatment of the data, differentiation at the strain level in these two examples can still be obtained.

To ascertain the robustness of the technique, separate spectra collected and averaged from the same spore culture were examined. All 16 species were correctly identified by comparison to the library spectra (r=0.85-0.98). This result was consistent whether the individual spectra themselves or averages of the individual spectra were used to search the library. In addition to the new preparations from the same culture, four batches of spores of *B. subtilis* 168, prepared at different times over the course of 2 years, were also compared against the library spectra. All of the *B. subtilis* 168 spores were correctly identified as the *B. subtilis* 168 from the library, regardless of the batch or storage time (r=0.92-0.98).

Aligning the correlation results from the MALDI-TOFMS profiles (Table 3-3) with the 16S rDNA sequence analysis (Table 3-2) shows that the MALDI-TOFMS profiles were complementary to 16S rDNA analysis. Using MALDI-TOFMS spore profiles of these organisms, we were able to differentiate all of the species studied confidently, whereas there are several species including *B. subtilis* 168, *B. licheniformis*, *B. mojavensis*, and *B. atrophaeus* that 16S was unable to differentiate at the species level. MALDI-TOFMS analysis on these species would allow for differentiation at the species level. Comparing the MALDI protein profiles with the phenotypic groupings was challenging due to the large diversity in the number and range of the peaks across the spectra for all of the species studied. In general, spores with an exosporium resulted in spectra that had more peaks over a broader range than the non-exosporium organisms. On average, the phenotypic group IV organisms had more peaks than the group II organisms, with the exception of *B. megaterium* and its wild-type FO-38. Cluster analysis was applied to the data to allow the relationships based on protein profiles between the different species to be visualized. The results of the single linkage cluster analysis using the SPSS software package are shown in Figure 3-4 combined with an image map of the spectra for visualization.

MALDI-TOFMS Vegetative Profiles

The bulk of this work was focused on the analysis of spores; however, the same technology was applicable for the analysis of vegetative cells. For routine analysis and identification (not direct environmental sampling), samples would likely be cultured prior to analysis. One of the advantages, other than speed, of the MALDI technique developed

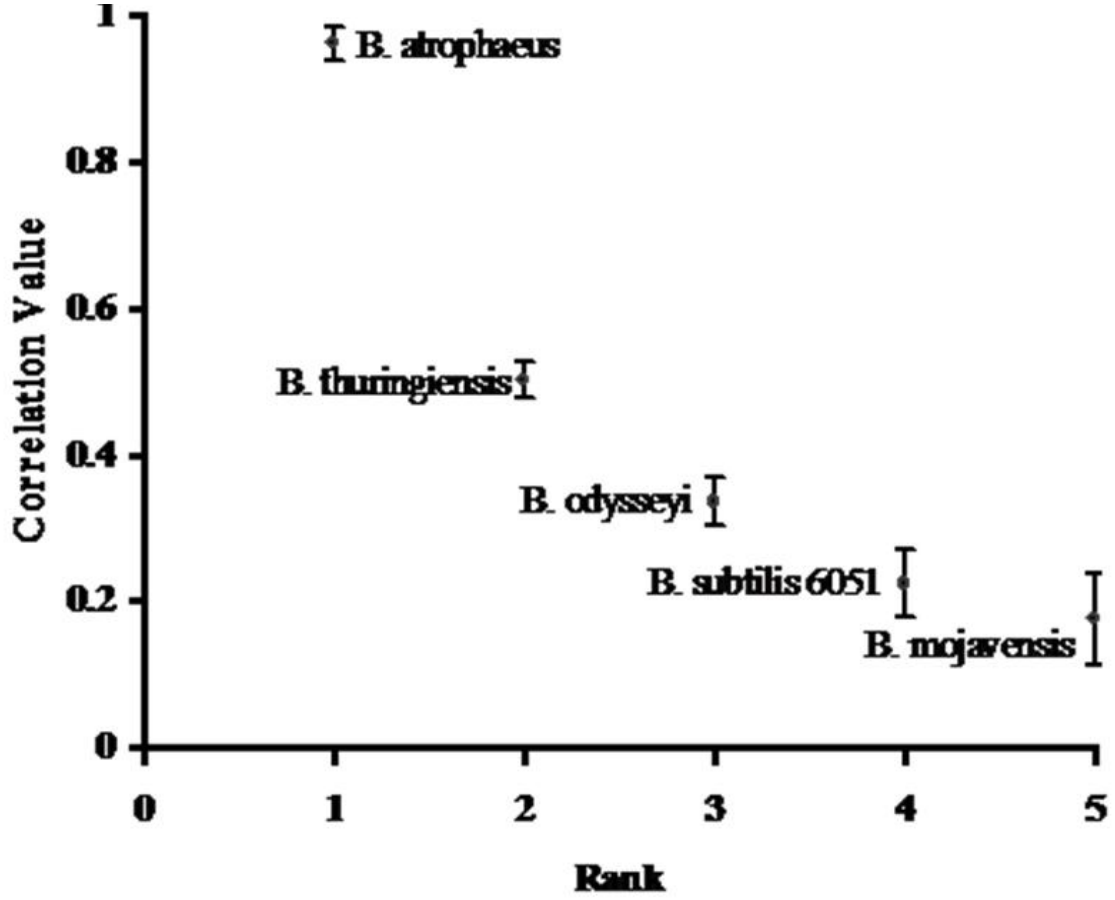

Figure 3-2. Correlation results of the 20 individual *B. atrophaeus* ATCC 9372 spectrum when searched against the library. The y axis represents the linear correlation values obtained and the x-axis represents the 1^{st}-5^{th} ranks (hits) from the library. At each rank, the standard deviation of the measurement across the 20 spectra is represented by the error bars.

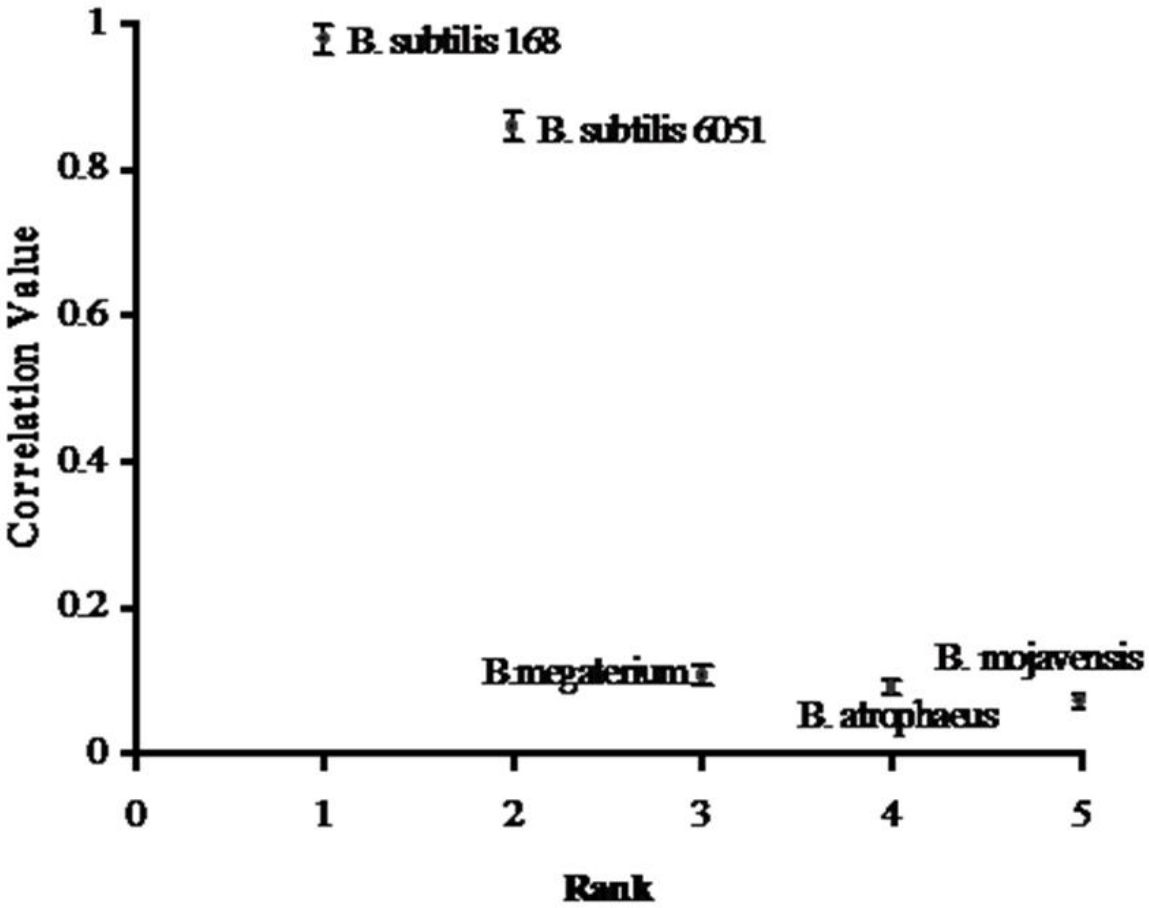

Figure 3-3. Correlation results of the 20 individual *B. subtilis* 168 spectrum when searched against the library. The y axis represents the linear correlation values obtained and the x-axis represents the 1st-5th ranks (hits) from the library. At each rank, the standard deviation of the measurement across the 20 spectra is represented by the error bars.

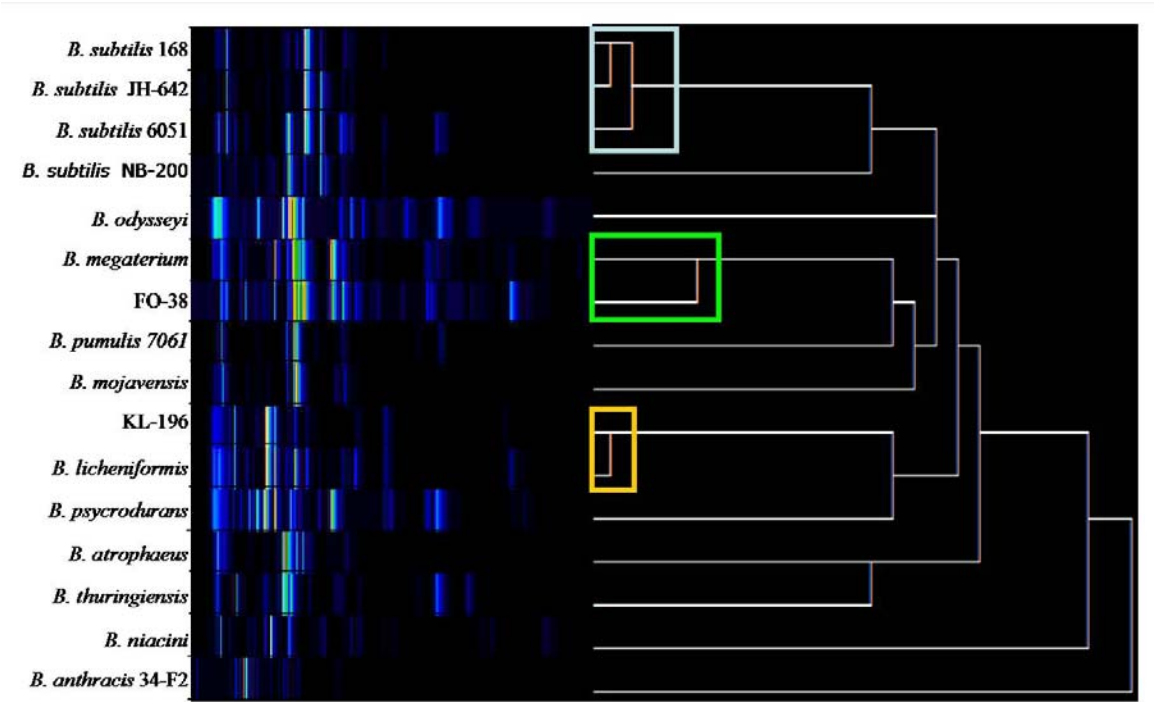

Figure 3-4. Visualization of the spectra in-line with the dendrogram for the spores in this study. The dendrogram is based on a single linkage scheme. Peak intensity is indicated by brighter colors in the image map. The dendrogram is highlighted to show that the closest clusters are between strains of the same species.

here was the amenability to the presence of numerous cultures on a single plate, which could be removed individually with a sample loop for analysis. This could eliminate the need for several incubation steps while trying to isolate a single organism.

To determine the scope of the current methodology the vegetative cells of the different species were also examined. The same extraction protocol with no modifications was used on the vegetative cells and should be effective since spores should present the more difficult challenge for protein extraction. The vegetative cell spectra for each of the species analyzed are shown in Figure 3-5 from *m/z* 2,500-60,000. The region from *m/z* 20,000-60,000 is amplified by 4x to highlight the upper molecular weight region of the spectra. Profiles of the vegetative cells have protein biomarker peaks that extend to a much higher range than their corresponding spore spectra. The vegetative profiles also have a greater number of peaks than the spores. The spectra obtained in this study have similar numbers of peaks as vegetative cell spectra in other studies where formic acid and ferulic acid were used in the matrix. No protein peaks were observed to overlap between the vegetative and spore spectra from the same strain, particularly due to the presence of peaks above 30 kDa in most the vegetative cell spectra.

Correlation analysis of the vegetative cells gave us very similar results to those of the spores. Complete differentiation of the different strains examined was possible and the highest correlation values were found between the *B. subtilis* and *B. licheniformis* pair (Table 3-4). Repeat analysis of different colonies from the plates gave correlation values of 0.71-0.85 with their corresponding library spectra. Since the vegetative cell spectra

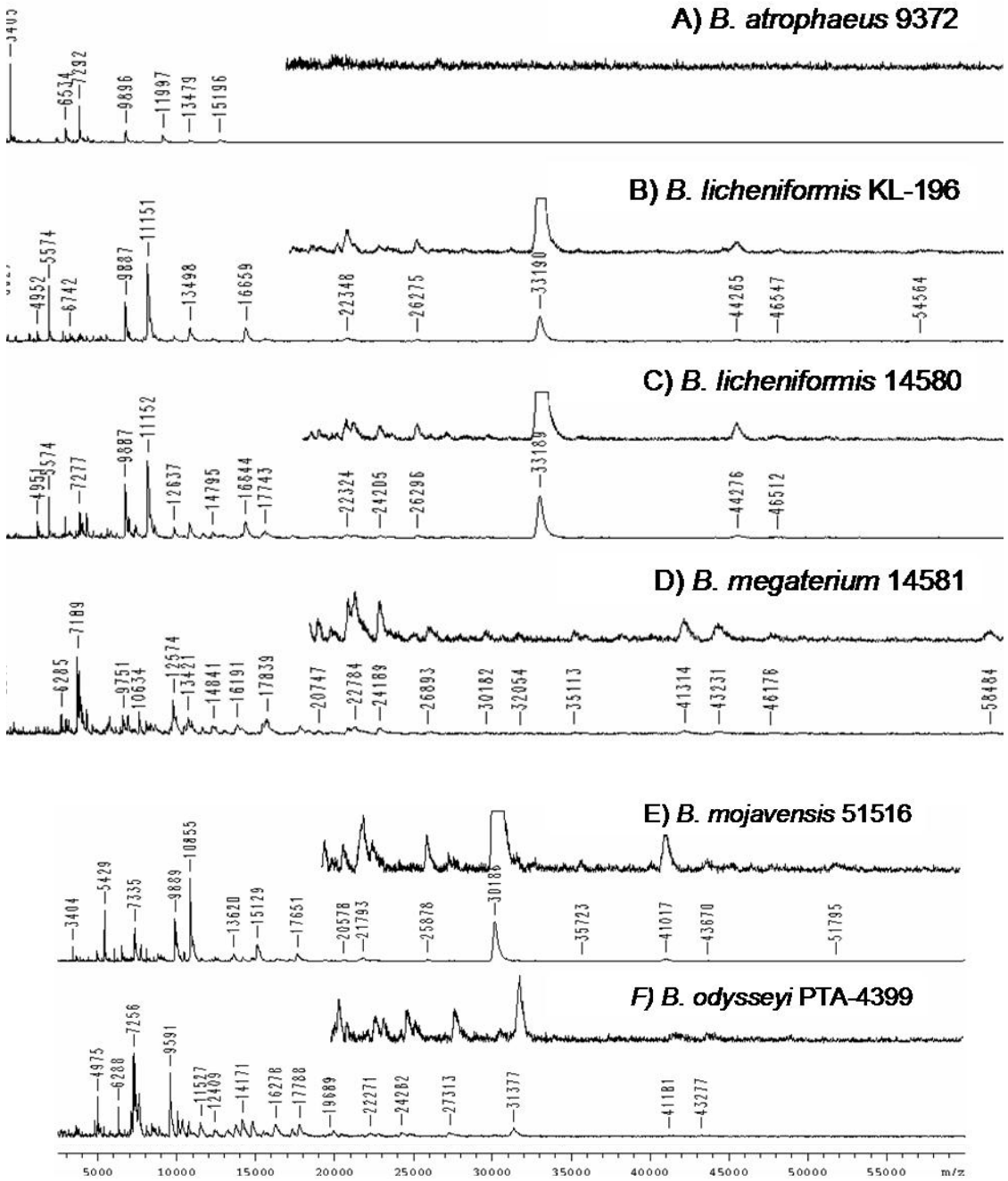

Figure 3-5. MALDI-TOFMS protein profiles of the 14 *Bacillus* vegetative species analyzed in this study. The mass range depicted is from *m/z* 2,500-60,000. The higher molecular mass region from *m/z* 20,000-60,000 is amplified 4x (see inset of each spectrum) in order to visualize the higher molecular weight peaks that are present in the samples. A) *B. atrophaeus* ATCC 9372. B) *B. licheniformis* KL-196. C) *B. licheniformis* ATCC 14580. D) *B. megaterium* ATCC 14581. E) *B. mojavensis* ATCC 51516. F) *B. odysseyi* ATCC PTA-4993. G) *B. psycrodurans* VSE1-06. H) *B. pumilus* ATCC 7061. I) *B. subtilis*168. J) *B. subtilis* ATCC 6051. K) *B. thuringiensis* ATCC 10792. L) *B. anthracis* 34F2. M) *B. niacini* 51-8C.

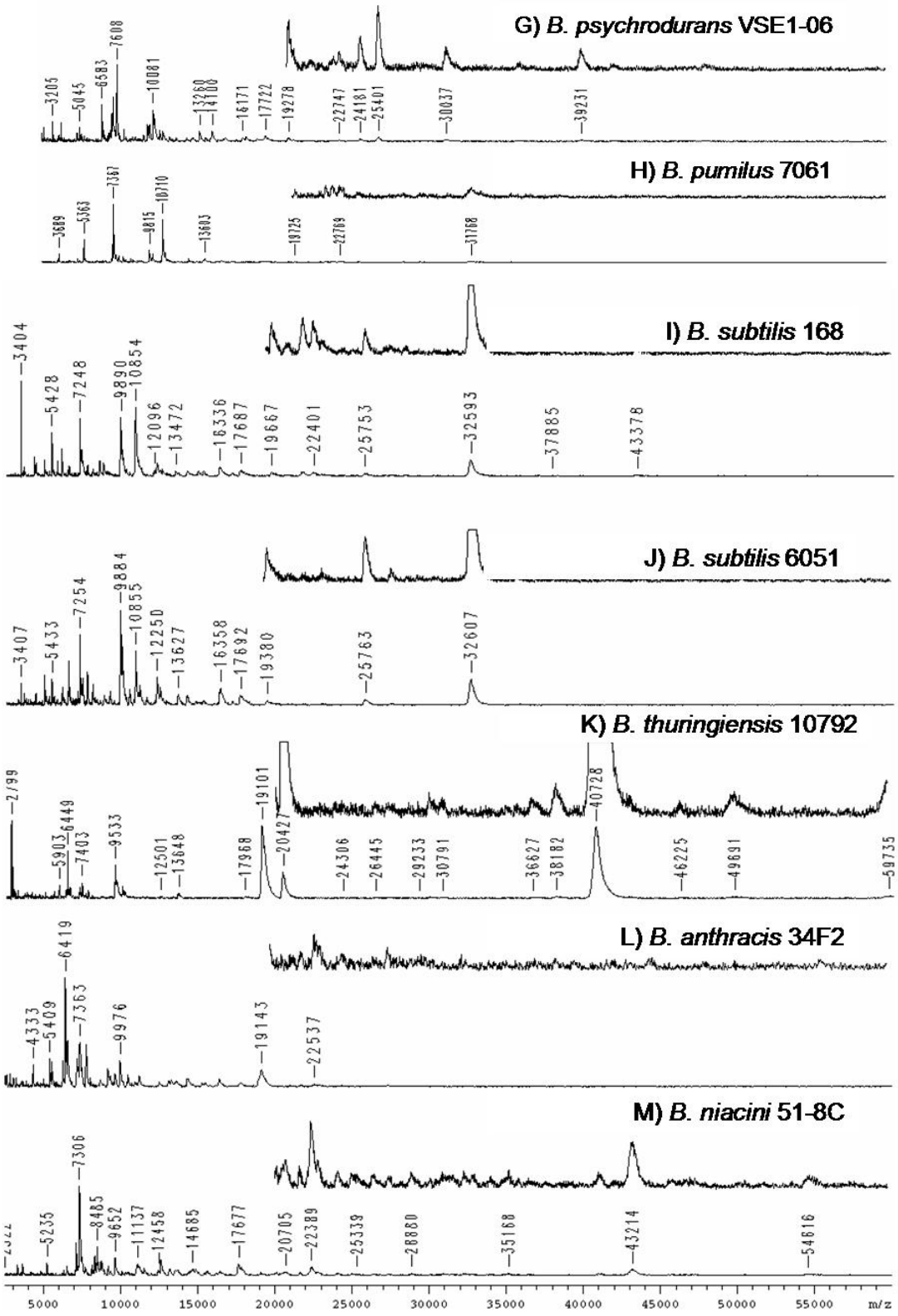

Figure 3-5. Continued.

are more complex than the spore spectra, hierarchical cluster analysis (HCA) in conjunction with visualization using image mapping was used to observe patterns in the spectra (Figure 3-6). The HCA analysis divided the vegetative cells into two clusters, one consisting of *B. atrophaeus*, *B. mojavensis*, the *B. licheniformis* pair, and *B. subtilis* pair and the second cluster containing the other species. *B. thuringiensis* had fewer peaks in comparison with the other vegetative cells and, therefore, was not similar to either cluster. The second cluster had more peaks overall than the first cluster and contained the RNA group 2 organisms. All of the species in the first cluster contained a biomarker peak at 9,890 Da which was not present in the second cluster. In addition, all the species in the first cluster except the *B. licheniformis* pair had a biomarker peak at 3,405. There were no obvious biomarker peaks observed in the second cluster.

Table 3-4. Correlation values based on MALDI-TOFMS protein profiling of the vegetative cells of the *Bacillus* species in this study

Bacteria (vegetative)		34F2	ATCC 9372	ATCC 14580	KL-196	ATCC 14581	ATCC 51516	51-8C	34hs1	VSE1-06	ATCC 7061	168	ATCC 6051	ATCC 10792
		B. anthracis	*B. atrophaeus*	*B. licheniformis*	*B. licheniformis*	*B. megaterium*	*B. mojavensis*	*B. niacini*	*B. odysseyi*	*B. psychrodurans*	*B. pumulis*	*B. subtilis*	*B. subtilis*	*B. thuringiensis*
B. anthracis	34F2	1												
B. atrophaeus	ATCC 9372	0.23	1											
B. licheniformis	ATCC 14580	0.16	0.14	1										
B. licheniformis	KL-196	0.09	0.07	0.93	1									
B. megaterium	ATCC 14581	0.37	0.27	0.25	0.15	1								
B. mojavensis	ATCC 51516	0.16	0.21	0.22	0.19	0.21	1							
B. niacini	51-8C	0.40	0.32	0.20	0.13	0.61	0.24	1						
B. odysseyi	34hs1	0.37	0.25	0.15	0.06	0.53	0.21	0.57	1					
B. psychrodurans	VSE1-06	0.37	0.26	0.20	0.10	0.49	0.23	0.49	0.63	1				
B. pumulis	ATCC 7061	0.31	0.21	0.18	0.10	0.45	0.29	0.49	0.52	0.39	1			
B. subtilis	168	0.18	0.53	0.29	0.24	0.28	0.67	0.25	0.32	0.30	0.34	1		
B. subtilis	ATCC 6051	0.22	0.32	0.43	0.35	0.30	0.58	0.20	0.29	0.33	0.30	0.80	1	
B. thuringiensis	ATCC 10792	0.27	0.01	-0.02	-0.02	0.03	-0.01	0.04	0.07	0.06	0.03	-0.01	0.00	1

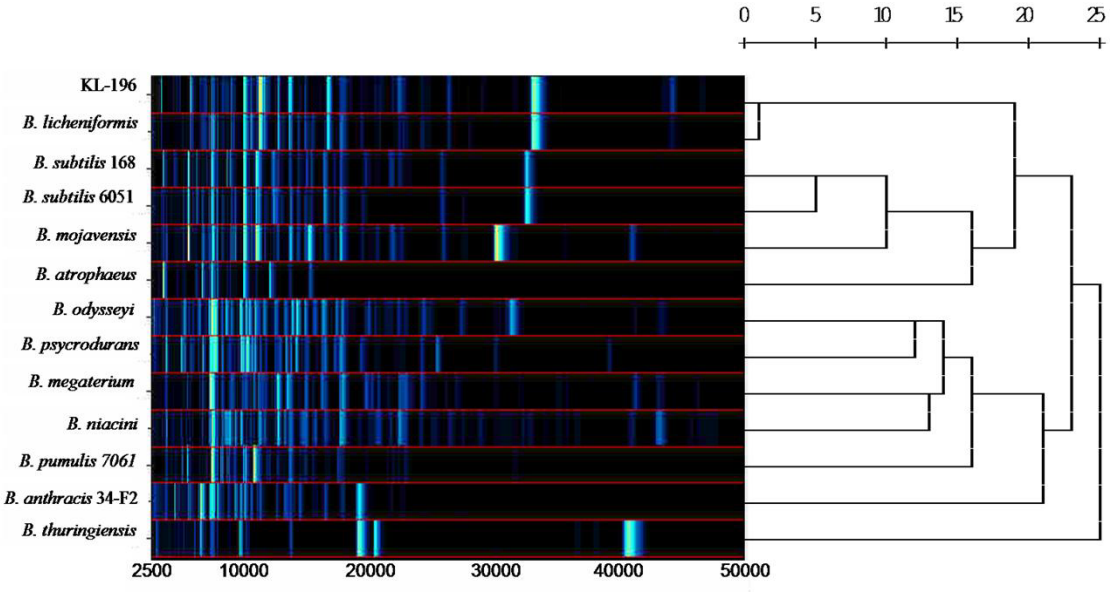

Figure 3-6. Visualization of the spectra in-line with the dendrogram for the vegetative cells in this study. The dendrogram is based on a single linkage scheme. Peak intensity is indicated by brighter colors in the image map.

Four strains (*B. subtilis* 168, *B. anthracis* 34F2, *B. pumilus* ATCC 7061, and *B. thuringiensis* ATCC 10792) were chosen to further study the variability of growth media on the spectra from vegetative cells. These strains were cultured on 3 media, TSA, NA, and LB, under identical incubation conditions. Upon observation, there was a significant difference in the size of the colonies on the different plates. The largest colonies were on the TSA plates, followed by NA and then LBA plates. Every effort was made to remove similar size samples for each strain by only using a small portion from the edges of the larger colonies. Table 3-5 shows the results of the correlation analysis of these samples with each other. Figure 3-6, 3-7, 3-8, and 3-9 show the average spectra from each species on each of the growth plates.

Table 3-5. Correlation values based on MALDI-TOFMS protein profiling of vegetative cells of select *Bacillus* species incubated on three different growth media

Bacteria (vegetative)	Media	*B. anthracis* 34F2 LB	*B. anthracis* 34F2 NA	*B. anthracis* 34F2 TSA	*B. pumilus* 7061 LB	*B. pumilus* 7061 NA	*B. pumilus* 7061 TSA	*B. subtilis* 168 LB	*B. subtilis* 168 NA	*B. subtilis* 168 TSA	*B. thuringiensis* 10792 LB	*B. thuringiensis* 10792 NS	*B. thuringiensis* 10792 TSA
B. anthracis 34F2	LB	-											
B. anthracis 34F2	NA	0.75	-										
B. anthracis 34F2	TSA	0.73	0.77	-									
B. pumilus 7061	LB	0.31	0.35	0.58	-								
B. pumilus 7061	NA	0.29	0.31	0.44	0.78	-							
B. pumilus 7061	TSA	0.32	0.34	0.56	0.82	0.78	-						
B. subtilis 168	LB	0.14	0.10	0.14	0.24	0.27	0.25	-					
B. subtilis 168	NA	0.46	0.55	0.37	0.06	0.07	0.07	0.03	-				
B. subtilis 168	TSA	0.15	0.14	0.18	0.23	0.27	0.25	0.80	0.05	-			
B. thuringiensis 10792	LB	0.56	0.56	0.70	0.48	0.40	0.48	0.12	0.21	0.17	-		
B. thuringiensis 10792	NS	0.44	0.46	0.37	0.11	0.11	0.12	0.04	0.45	0.09	0.39	-	
B. thuringiensis 10792	TSA	0.41	0.40	0.37	0.14	0.10	0.14	0.03	0.34	0.06	0.63	0.64	-

Abbreviations: LB, Luria-Bertani Agar; NA, Nutrient Agar; TSA, Tryptic Soy Agar

In the case of *B. anthracis* (Figure 3-6) and *B. pumilus* (Figure 3-9), the spectra from all 3 of the growth media maintained correlation values of above 0.70 when compared with each other. In both cases, visual observation of the spectra revealed that the 3 samples had several biomarker peaks in common but the NA sample (middle) had additional biomarker peaks not seen in the TSA (top) or LB (bottom) samples. In *B. subtilis* 168 (Figure 3-7), the TSA and LB samples were similar to each other ($r = 0.80$); however, the NA sample had few biomarkers in common with the other two samples and was significantly different. The *B. subtilis* 168 sample grown on NA also demonstrated the presence of higher molecular weight biomarkers not seen in the other samples. The *B. thuringiensis* cells grown on the different media (Figure 3-8) were very dissimilar, supported by very low correlation values and visual observation of the spectra. The most significant difference in these samples is seen in the range above 30 kDa and below 7 kDa. Several biomarkers, including 2,891 Da, 10,010 Da, and 19,046 Da are common across all *B. thuringiensis* samples in this study. Therefore, it is still possible to identify species specific biomarkers present in several growth media. Notably, the *B. thuringiensis* cells grown on the LB plates had a correlation value of 0.70 with the *B. anthracis* cells grown on TSA even though they do not share the above mentioned biomarkers. The *B. anthracis* and *B. thuringiensis* samples have the 7,350 Da biomarker in common. The differences in the protein profiles on the 3 media could be an effect of the difference in growth phase of the organism or the differences in the proteins expressed due to the nutrients available in the different media.

The cells in the media study were grown for 8 hours longer than the cells in which the initial comparisons of the vegetative cells were done. The media samples (20 hour

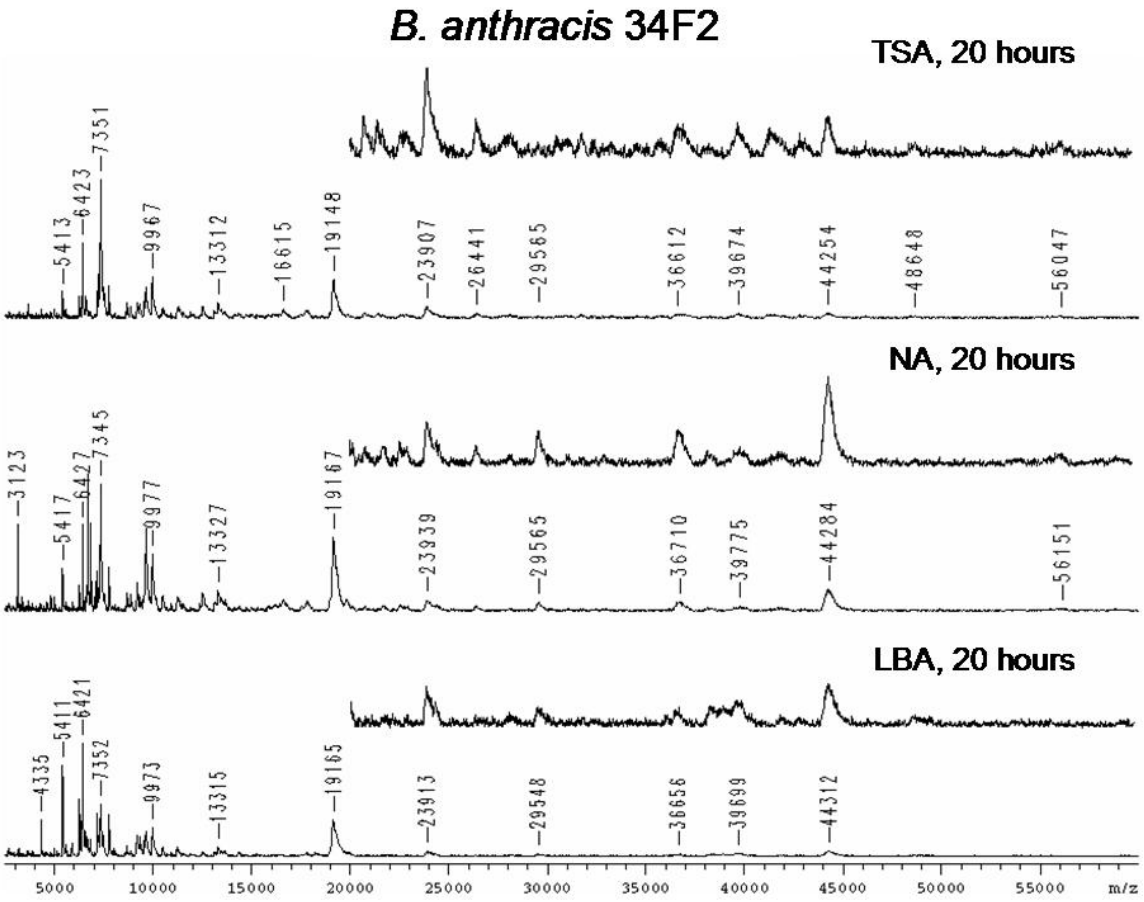

Figure 3-7. MALDI-TOFMS protein profiles of *B. anthracis* 34F2 vegetative cells from different growth media: tryptic soy agar (top), nutrient agar (middle), and Luria Bertani agar (bottom). The mass range is from *m/z* 2,500-60,000. The higher molecular mass region from *m/z* 20,000-60,000 is amplified 4x (see inset of each spectrum) to aid in the visualization of the higher molecular weight peaks that are present.

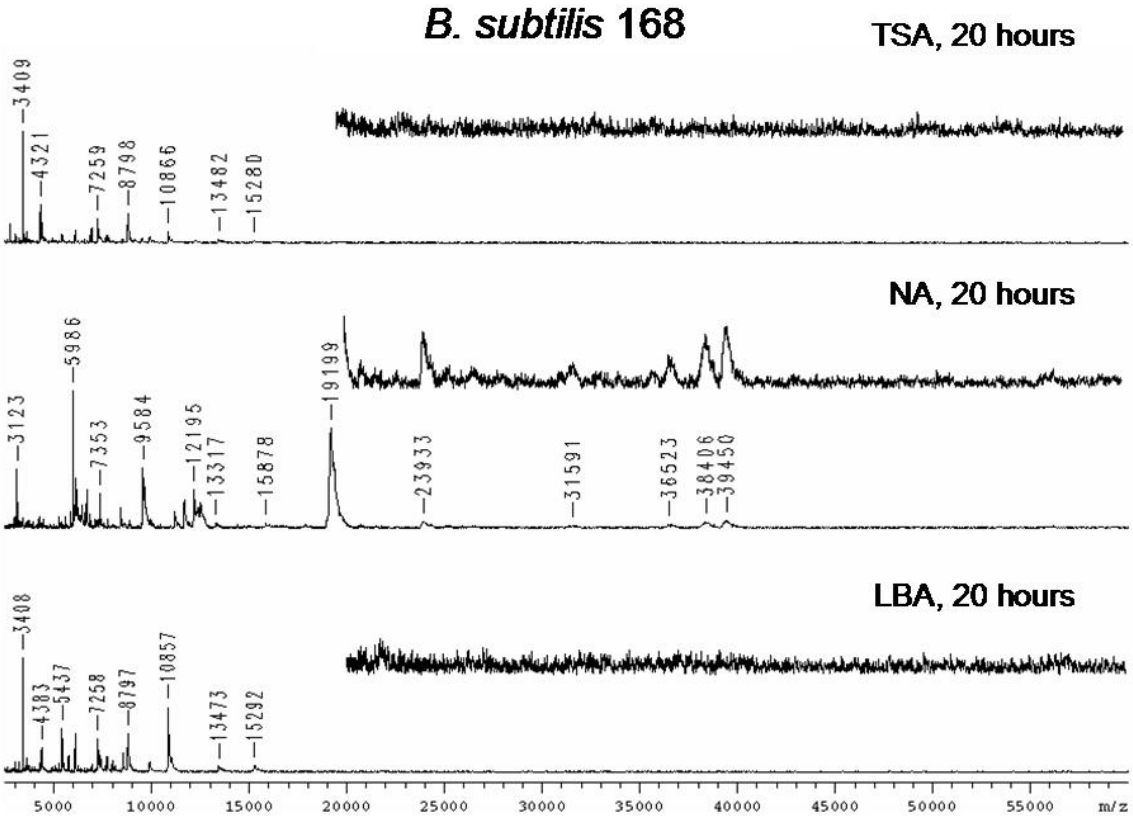

Figure 3-8. MALDI-TOFMS protein profiles of *B. subtilis* 168 vegetative cells from different growth media: tryptic soy agar (top), nutrient agar (middle), and Luria Bertani agar (bottom). The mass range depicted is from *m/z* 2,500-60,000. The higher molecular mass region from *m/z* 20,000-60,000 is amplified 4x (see inset of each spectrum) to aid in the visualization of the higher molecular weight peaks that are present.

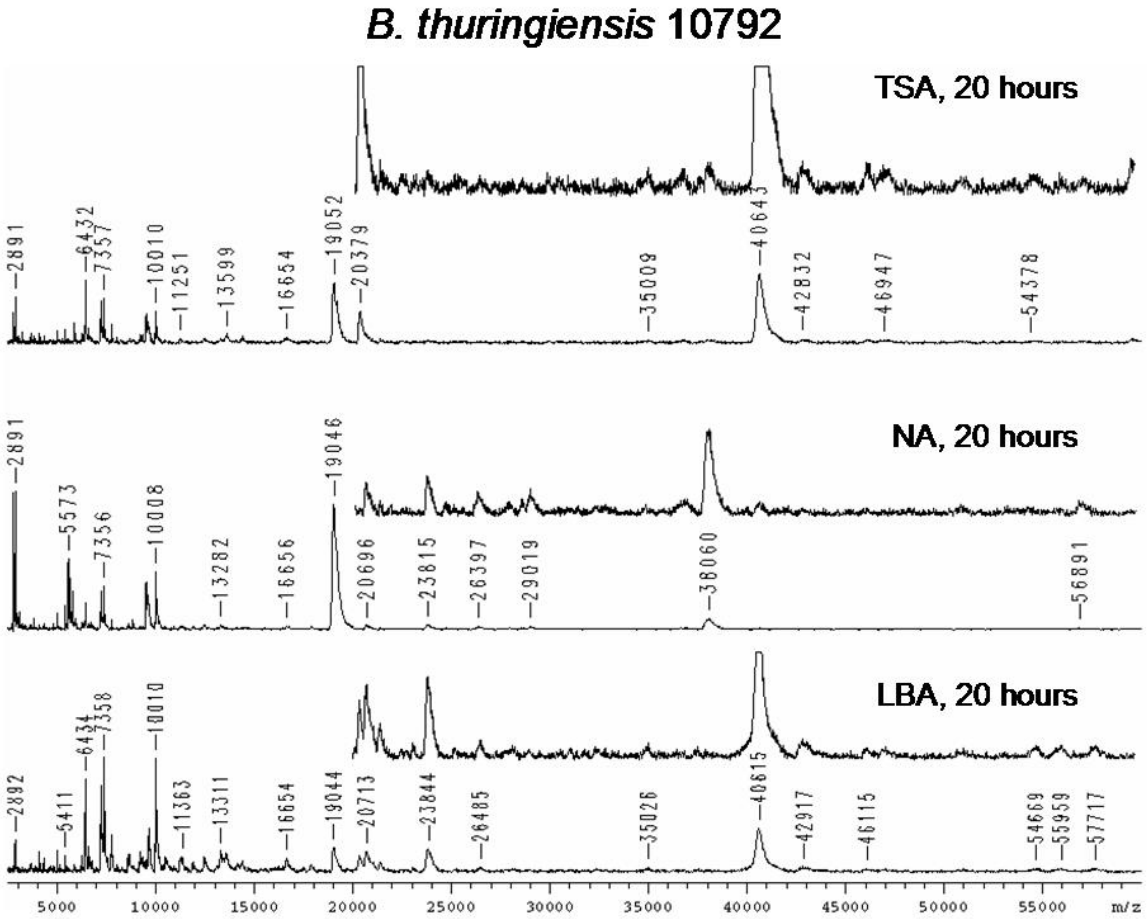

Figure 3-9. MALDI-TOFMS protein profiles of *B. thuringiensis* ATCC 10792 vegetative cells from different growth media: tryptic soy agar (top), nutrient agar (middle), and Luria Bertani agar (bottom). The mass range depicted is from *m/z* 2,500-60,000. The higher molecular mass region from *m/z* 20,000-60,000 is amplified 4x (see inset of each spectrum) to aid in the visualization of the higher molecular weight peaks that are present.

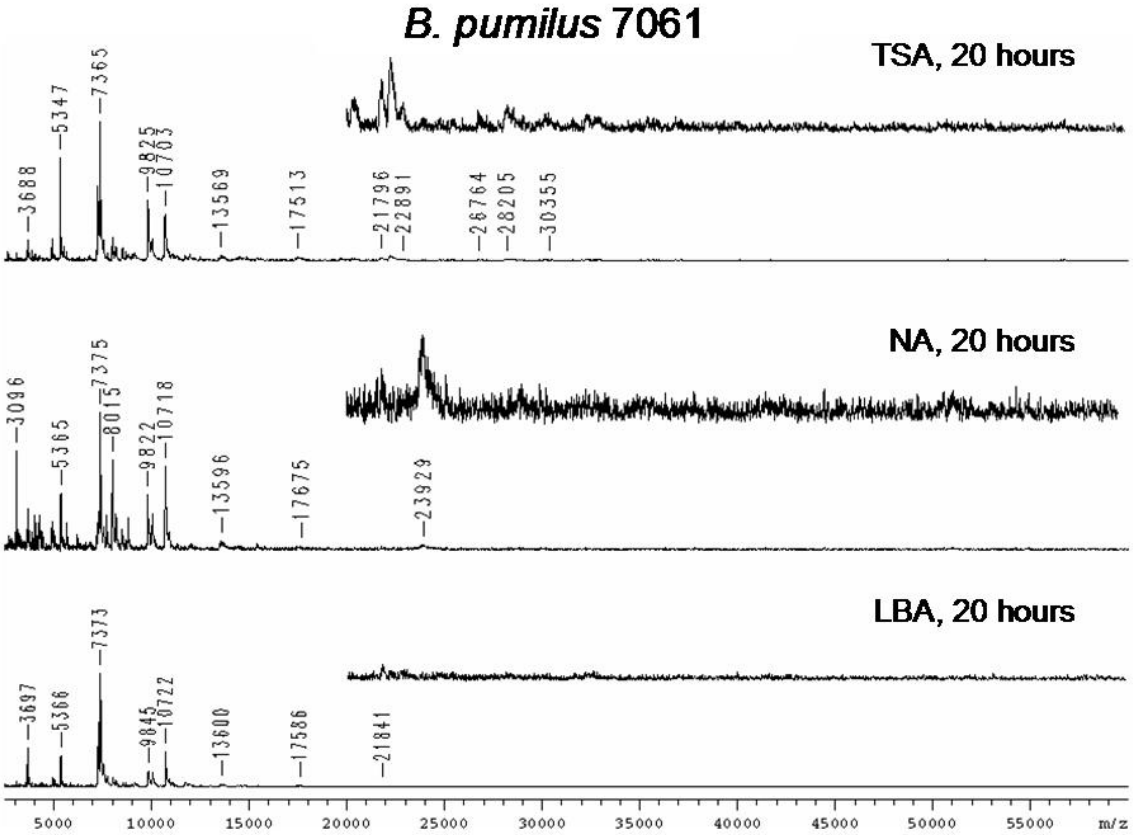

Figure 3-10. MALDI-TOFMS protein profiles of *B. pumilus* 7061 vegetative cells from different growth media: tryptic soy agar (top), nutrient agar (middle), and Luria Bertani agar (bottom). The mass range depicted is from *m/z* 2,500-60,000. The higher molecular mass region from *m/z* 20,000-60,000 is amplified 4x (see inset of each spectrum) to aid in the visualization of the higher molecular weight peaks that are present.

incubation) were compared with the vegetative cell library above (16 hour incubation) to ascertain the difference that growth time can make on the spectra. In the case of *B. pumilus* 7061 and *B. anthracis* 34F2, all of the media samples were positively identified with the *B. pumilus* 7061 and *B. anthracis* 34F2, samples from the previous day. The correlation values for the TSA, NA, and LB samples with the type strain sample on TSA from the previous day were 0.85, 0.76, and 0.83 respectively for the *B. pumilus* 7061 and 0.68, 0.62, and 0.88 for the *B. anthracis* 34F2. In the case of *B. thuringiensis,* the TSA and NA samples were correctly identified with correlation values of 0.88 and 0.58; however, the LB sample again had the highest correlation with the *B. anthracis* sample from the previous day (r = 0.57). Only *B. subtilis* 168 grown on LB was properly identified, with a correlation value of 0.71. The NA sample was closest to B. thuringiensis (r = 0.40) and the TSA sample was closest to B. atrophaeus (r = 0.68), both grown on TSA from the previous study.

Although not all of the species type out properly, the robustness of the technique was highlighted, as cells grown on different media and incubation times still gave relatively high correlation values to a library strain of the same species under different conditions. To determine if the absolute value of r was significant (i.e., does an r of 0.40 indicate a good match?), a more detailed study was needed to determine the range of correlation values encountered when considering the strain variation across a species.

Conclusion

MALDI-TOFMS based protein profiling is a useful, rapid, and sensitive technology to differentiate spores and vegetative cells from closely related microbial species. Although a standardized sample preparation protocol is required, it is obvious that this

technology is promising for species differentiation of a wide variety of bacterial spores and cells.

CHAPTER 4
MALDI-TOFMS COMPARED WITH OTHER POLYPHASIC TAXONOMY APPROACHES FOR THE IDENTIFICATION AND CLASSIFICATION OF *Bacillus pumilus* SPORES

Introduction

To verify the efficacy of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOFMS) protein profiling for identifying and differentiating bacterial species, several strains of *Bacillus pumilus* were examined in a thorough taxonomic study incorporating a polyphasic approach. MALDI-TOFMS protein profiling is rapid, sensitive, and has higher resolution and better reproducibility than gel-based protein or DNA fingerprinting, and has proven effective for bacterial identification.[16,31,33] By carefully controlling extraction conditions combined with suitable software for data analysis, MALDI-TOFMS has the potential to identify and classify previously unidentified environmental isolates. The realization of this potential is dependent on the availability of standardized MALDI-TOFMS profile libraries for comparison of unknown isolates with reference strains. Verification of this microbial classification scheme has not been thoroughly explored, and published studies on this technique have focused solely on bacteria from culture collections and/or blind studies using strains already represented within user generated libraries.[20,71,72] These studies have included members of *Enterobacteriaceae*, *Bacillus*, *Staphylococcus*, *Streptococcus*, and other medically important species.

Similar MALDI-TOFMS studies have been reported on the species and strain differentiation of *Bacillus* spores;[25,26,29,32,73] however, protein profile variation in spores

of a single species by examining several strains isolated from various sources has not been reported to our knowledge. Such research will support the application of MALDI-TOFMS in field applications. In the current study, 16 isolates of putative *B. pumilus* from different spacecraft assembly facilities, the Mars Odyssey spacecraft, and the International Space Station, were characterized using the Biolog system, DNA techniques, and MALDI-TOFMS protein profiling. *B. pumilus* is one of the predominant spore-forming microbes in spacecraft and associated clean room environments.[60] Moreover, resistance of *B. pumilus* spores to various stressors is strain-specific[74] and might be influenced by the environmental factors resulting in the expression of different proteinaceous compounds for protection.[75,76] A one-step sample treatment and MALDI-TOFMS preparation was used to obtain spectra for the creation of a library of spectra from the different isolates of putative *B. pumilus*, providing the most diverse study of a single bacterial spore species using protein profiling reported to date.

The results obtained from MALDI-TOFMS protein fingerprinting of these *B. pumilus* isolates was compared with DNA-DNA hybridization for their bacterial systematics and molecular phylogenetic affiliations. MALDI-TOFMS protein profiling was more accurate than Biolog metabolic profiling, more discriminating than 16S rDNA sequence analysis, and complemented the results of *gyrB* sequence analysis and DNA-DNA hybridization for the identification of the *B. pumilus* spores. This is the first report whereby MALDI-TOFMS generated protein profiles from a set of microbes are compared directly with DNA-DNA hybridization yielding a positive correlation. Unique, cluster-specific biomarker peaks have been identified in the spores of the *B. pumilus* examined in this study. MALDI-TOFMS protein profiling is a rapid and simple analysis

and is demonstrated as a useful taxonomic tool for differentiating spores of the genus *Bacillus*. For practical purposes, it would be ideal (and necessary) to have a publicly available, standardized MALDI profile database, to facilitate the use of the technique as a diagnostic method to differentiate bacterial species.

Materials and Methods

Bacterial Strains

Table 4-1 contains a list of the wild-type bacterial strains used in this study. All ATCC strains were procured from the American Type Culture Collection (Manassas, VA), including the type strains of *B. atrophaeus*, *B. subtilis*, *B. megaterium*, *B. mojavensis*, *B. pumilus*, and *B. licheniformis*. The *B. odysseyi* type strain was from our culture collection and *B. subtilis* 168 was obtained from Dr. Wayne Nicholson at the University of Florida. The source, location, and date of isolation of the 16 wild-type isolates of putative *B. pumilus* are indicated in Table 1 along with the other isolate species used in this study. Bacterial isolation procedures from spacecraft assembly facility surfaces are described elsewhere [59,61].

Sporulation of *Bacillus* isolates

A standard protocol for the production of spores was used in this study [7]. A single purified colony of the strain to be sporulated was inoculated into nutrient broth sporulation medium (NSM) and incubated at 32^{o}C with shaking for ca. 2-4 days, until the cultures reached >99% spores as examined by phase-contrast microscopy. Spore cultures were harvested by centrifugation and purified to remove remnant vegetative cells and cellular debris, as previously reported [7]. The purification protocol involved a lysozyme treatment followed by salt and detergent washes to remove vegetative cellular debris.

Purified spores were adjusted to an optical density of 0.6 at 600 nm and were stored in sterile deionized water at 4°C in glass vials until analyzed.

Table 4-1. Strain designation, grouping, and source of *Bacillus* species in this study

Species	Strain #	Source[a]	Year of Isolation	16S rDNA Genbank Accession Number[b]	Comments
B. pumilus type strain group					
B. pumilus	ATCC 7061^T	ATCC	-	AB020208	Type strain
B. pumilus	ATCC 27142	ATCC	-	n/a	Gamma radiation resistant strain
B. pumilus	0105342-2	ISS-hardware	2000	n/a	International Space Station hardware
B. pumilus	SAFN-029	JPL-SAF	2001	AY167883	Air-lock
B. pumilus	SAFR-032	JPL-SAF	2001	AY167879	Air-lock
B. pumilus FO-036b group					
B. pumilus	FO-033	JPL-SAF	1999	AF234851	Clean room air particulate
B. pumilus	FO-036b	JPL-SAF	1999	AF234854	Clean room air particulate
B. pumilus	SAFN-001	JPL-SAF	2001	AY167886	Entrance floor
B. pumilus	SAFN-027	JPL-SAF	2001	AY167884	Ante-room
B. pumilus	SAFN-036	JPL-SAF	2001	AY167881	Clean room floor
B. pumilus	SAFN-037	JPL-SAF	2001	AY167880	Clean room floor
B. pumilus	KL-052	JPL-SAF	2000	AY030327	Clean room cabinet top
B. pumilus	51-3C	Mars Odyssey	2002	AF526907	Mars Odyssey spacecraft surface
B. pumilus	81-4C	KSC-SAEF II	2002	AF526903	Mars Odyssey assembly facility floor
B. pumilus	82-2C	KSC-SAEF II	2002	AF526902	Mars Odyssey assembly facility floor
B. pumilus	84-1C	KSC-SAEF II	2002	AF526898	Mars Odyssey assembly facility floor
B. pumilus	84-3C	KSC-SAEF II	2002	AF526896	Mars Odyssey assembly facility floor
B. pumilus	84-4C	KSC-SAEF II	2002	AF526895	Mars Odyssey assembly facility floor
Wild-type strains of other *Bacillus* species					
B. cereus	FO-11	JPL-SAF	1999	AY461790	Clean room air particulate
B. licheniformis	KL-196	JPL-SAF	2000	AF387515	Clean room cabinet top
B. niacini	51-8C	KSC-SAEF II	2002	AF526905	Mars Odyssey assembly facility floor
B. odysseyi	PTA-4399	Mars Odyssey	2002	AF526913	Mars Odyssey spacecraft surface
B. psychrodurans	VSE-01	KSC-PHSF	2002	n/a	Mars Exploration Rovers assembly facility air particles

[a]Abbreviations: JPL, Jet Propulsion Laboratory, KSC, Kennedy Space Center, SAF, Spacecraft Assembly Facility, PHSF, Payload Hazardous Servicing Facility, SAEF, Spacecraft Assembly and Encapsulation Facility, ATCC, American Type Culture Collection

b Included sequences reported in various publications were used for comparison (LaDuc et. al. 2003b; Venkateswaran et. at. 2001)

Vegetative Cell Growth

To produce vegetative cells for analysis, cultures were first streaked out on tryptic soy agar plates and incubated overnight at 32°C. A single, isolated colony was then used to inoculate a 5 mL tryptic soy broth culture. This culture was incubated at 32°C with shaking at 250 rpm for 8 hours. A milliliter of the culture was removed from the tube and spun down for 10 minutes at 9,600 x g. The supernatant solution was removed and the pellet was resuspended in phosphate buffered saline solution. The culture was again centrifuged , the supernatant removed, and the remaining cell pellet used for the MALDI analysis.

Metabolic profiling

All isolates were subjected to Gram staining and the presence of spores was confirmed using light microscopy. Metabolic profiling was performed on the various isolates using the Biolog system (Biolog, Foster City, CA). This 96-well microplate method tests for the oxidation of 95 different carbon sources. Protocols for the preparation and analysis of *Bacillus* species were followed per the manufacturer's directions. Microplates were read after 6 and 20 hours of incubation in the 96 well plates. The Microstation hardware was used to read the plates and Microlog 3 software was used to analyze the data.

16S rDNA and *gyrB* sequencing

Chromosomal DNA from each of the isolates was extracted by standard PCIAA and ethanol precipitation protocols[77], and was used as the template for PCR amplification (ca. 10 ng). Universal primers (Eub 8f and Univ.1492r) were used to amplify 16S rDNA fragments, as per established protocols[53]. Procedures developed by Yamamoto and Harayama[56] were followed for *gyrB* amplification. Amplicons were gel-excised, purified with Qiagen columns (Qiagen, Valencia, CA), and sequenced as described elsewhere[59,61]. The phylogenetic relationships of organisms covered in this study were determined by comparison of individual 16S rDNA (www.ncbi.nlm.nih.gov) or *gyrB* (www.mbio.co.jp) sequences to other existing sequences in the public databases. Evolutionary trees were constructed with PAUP software, following maximum- parsimony parameters[78].

DNA-DNA hybridization

Bacterial strains were cultivated in tryptic soy broth (Difco, St. Louis, MO) containing 1.5% glycine by shaking at 30^{o}C for 16 hours. Cells were harvested by centrifugation, resuspended in TE buffer (pH 8.0), and lysed by the addition of 50 μg/mL

labiase (Seikagaku Corporation, Japan) and 1mg/ mL achromopeptidase (Wako Pure Chemicals, Japan). Cell suspensions were incubated at 37°C until they became viscous, at which time chromosomal DNA was purified per standard methods[77]. DNA-DNA homology was performed using the microplate hybridization method[79] with photobiotin labeling and colorimetric detection as described previously[80]. Cluster analysis based on DNA hybridization was performed via the neighbor-joining method[81] using PHYLIP software[82].

MALDI-TOFMS protein profiling

Purified spores were diluted one to ten in a saturated solution of ferulic acid matrix using a 30% formic acid, 40% water, and 30% acetonitrile mixture as the solvent[73]. The mixture was vortexed briefly and 1µL of the sample containing both spores and matrix was deposited on the MALDI plate. For vegetative cell samples, 50 µL of the matrix solution was used to resuspend the cell pellet. The mixture was vortexed briefly and 1µL of the sample containing both vegetative cells and matrix was spotted. Spots were allowed to air dry and no further treatment was applied to the spots post drying. The sample preparation was done in duplicate from each spore and vegetative cell suspension. MALDI-TOFMS protein profiling was performed on a Bruker Daltonics Reflex II Mass Spectrometer (Bruker Daltonics, Billerica, MA) retrofitted with delayed extraction. Positive ion mass spectra were collected in the linear mode using a delay time of 50 ns, an acceleration voltage of 20kV, and a deflector set at 2,500 Da. All spectra represent the accumulation of 50 laser shots. Ten spectra were collected across each spot for a total of 20 spectra per sample. Each spectrum was baseline corrected and smoothed using a ten-point Savitzky-Golay smoothing algorithm prior to statistical analysis.

Statistical processing of MALDI-TOFMS profiles

Linear correlation analysis was performed on software developed in-house with Visual Basic 6.0 as described previously [73]. A library of isolate spore spectra was complied by averaging the 20 spectra collected from each spore sample. The individual and the average spectrum obtained from the strains in this study were compared to MALDI-TOFMS profiles stored in this library and to a user generated library of eleven *Bacillus* species. Higher correlation coefficients are indicative of spectral similarity. To quantify the level of significance of these differences, a simple t-test was applied. A similar procedure was used for the vegetative cells except they were compared only with themselves, rather than with another library.

In addition to the linear correlation analysis, SPSS software (Chicago, IL) was used for performing a hierarchal cluster analysis (HCA) on the average spectrum obtained from each strain in this study. The HCA analysis was based on the Pearson correlation and dendrograms were produced using a single linkage (nearest neighbor) scheme. Spectral visualization is accomplished through the use of Surfer 8.0 mapping software from Golden Software (Golden, CO). Using this mapping software, an image map with 10 Da resolution is produced from the average spectrum from each species. The intensity of the peaks is represented by the color of the bands in the image.

Results

Metabolic fingerprinting of *B. pumilus* strains

Among the 95 carbon substrates tested, N-acetyl-D-glucosamine, inosine, and thymidine were oxidized by all the *B. pumilus* strains but were not oxidized by *B. subtilis* 168. Alpha- and β-cyclodextrin and L-lactic acid were oxidized by *B. subtilis* 168 but were not oxidized by any of the *B. pumilus* strains. All of the 16 wild-type *B. pumilus*

isolates, *B. pumilus* ATCC 27142, and *B. subtilis* 168 reduced maltose, methyl-D-glucoside, palatinose, turanose, pyruvic acid methyl ester and cellibose whereas *B. pumilus* ATCC 7061^T did not.

The Biolog identification system was not able to discriminate the spore forming aerobic bacteria in this study. The system correctly identified only 3 out of 18 strains tested, ATCC 7061^T, SAFN-036, and SAFN-037 as *B. pumilus*. Eight of the *B. pumilus* strains (FO-36b, SAFN-001, SAFN-027, 51-3C, 82-2C, 84-1C, 84-3C, and 84-4C) were incorrectly identified as *B. subtilis* (44%). Metabolic fingerprinting profiles of the remaining 7 *B. pumilus* strains (ATCC 27142, FO-033, KL-052, SAFN-029, SAFR-032, 81-4C, and 015342-2) did not match with any of the species contained in the Biolog metabolic fingerprinting database (39%).

16S rDNA and *gyrB* sequencing

The results of 16S rDNA sequencing and maximum-parsimony analysis rendered no apparent phylogenetic clustering pattern among the isolates tested. All of the *B. pumilus* examined, excluding 84-1C, 82-2C, 84-3C, and SAFR-032, had greater than 97.5% sequence similarity with each other. Strains 84-1C and SAFR-032 exhibited 16S rDNA sequence similarities of >99% with *B. pumilus* ATCC 7061^T and ~96.5% similarities with FO-36b group of isolates. Likewise, 82-2c and 84-3c isolates showed 16S rDNA similarities of ~96.5% with both ATCC 7061^T and the FO-36b group. *Sensu lato*, the strains examined in this study were indistinguishable by 16S rDNA sequence analysis.

The *gyrB* analysis of 12 strains sequenced exhibited two distinct clusters based on maximum-parsimony analysis. The first cluster showed >96.5% sequence similarities among 5 strains, *B. pumilus* ATCC 7061^T, ATCC 27142, 0105342-2, SAFN-029, and

SAFR-032. Likewise, >97.6% sequence similarities was noted among the other 7 strains analyzed, FO-033, FO-036b, SAFN-001, SAFN-027, SAFN-036, SAFN-037, and KL-052. Furthermore, only 90 to 92% *gyrB* sequence similarities were observed between the strains of these two groups.

DNA-DNA hybridization

The results of DNA-DNA hybridization among the *B. pumilus* strains are shown in Table 4-2. The strains tested diverged into two groups when DNA-DNA hybridization

Table 4-2. DNA-DNA hybridization of *B. pumilus* isolates

Species	Strain #	Percentage DNA reassociation values to labelled DNA from *B. pumilus* that are:																	
		SAFN-001	FO-033	FO-036b	81-4C	84-3C	51-3C	82-2C	84-1C	84-4C	SAFN-027	KL-052	SAFN-036	SAFN-037	SAFN-029	SAFR-032	0105342-2	ATCC 7061^T	ATCC 27142
B. pumilus	SAFN-001	-																	
B. pumilus	FO-033	92	-																
B. pumilus	FO-036b	92	93	-															
B. pumilus	81-4C	96	90	96	-														
B. pumilus	84-3C	98	90	91	92	-													
B. pumilus	51-3C	95	93	95	94	94	-												
B. pumilus	82-2C	95	93	93	90	93	90	-											
B. pumilus	84-1C	94	92	93	94	93	95	96	-										
B. pumilus	84-4C	93	90	93	93	93	95	95	90	-									
B. pumilus	SAFN-027	86	80	82	82	84	81	82	82	80	-								
B. pumilus	KL-052	84	80	84	84	85	85	81	81	82	96	-							
B. pumilus	SAFN-036	88	89	86	89	87	86	80	84	85	88	82	-						
B. pumilus	SAFN-037	91	80	86	87	84	86	88	84	88	84	83	96	-					
B. pumilus	SAFN-029	61	58	58	60	61	61	63	58	60	61	58	60	62	-				
B. pumilus	SAFR-032	58	54	54	62	63	57	53	57	54	59	61	57	54	97	-			
B. pumilus	0105342-2	61	55	62	62	63	64	60	60	60	60	61	60	61	79	79	-		
B. pumilus	ATCC 7061^T	65	63	61	61	60	63	60	61	62	66	59	61	66	81	84	85	-	
B. pumilus	ATCC 27142	59	59	60	58	63	60	58	63	62	59	63	57	60	79	76	90	85	-

values were examined by cluster analysis. The type strain group consisted of 5 strains as seen in *gyrB* analysis, *B. pumilus* ATCC 7061^T, ATCC 27142, 0105342-2, SAFN-029, and SAFR-032, which showed more than 76% DNA relatedness within the group. The FO-36b group consists of the remaining 13 strains, FO-033, FO-036b, 51-3C, 81-4C, 82-2C, 84-1C, 84-3C, 84-4C, SAFN-001, SAFN-027, SAFN-036, SAFN-037, and KL-052 where >80% DNA relatedness was observed within this group. Strains of the FO-36b

group formed a much more cohesive cluster than the type strain group. The DNA-DNA reassociation values between the type strain group and the strains in the FO-36b group were less than 66%. Even though the DNA-DNA hybridization values between the strains of the FO-36b group and the *B. pumilus* type strain ATCC 7061^T were 59-66%, we are still deeming this group *B. pumilus*.

MALDI-TOFMS protein profiling of spore samples

The averaged MALDI-TOFMS spectrum from each of the sixteen putative *B. pumilus* isolates and ATCC 27142 were compared sequentially with a user created library containing spectra from 12 strains (10 different type species) of Bacillus spores. The results from the correlation analysis are shown in Table 4-3. The 16 *B. pumilus* isolates tested in this study had low correlation values (0-0.48) with type strains of the ten other Bacillus species examined. Typically, a correlation value of >0.75 together with visual pattern recognition of the protein profiling were considered to define the bacterial species identity. Using MALDI-TOFMS protein profiling as a tool for phenotypic analysis all but 2, or 89% of the spores tested had correlation values of greater than 0.62 with the type strain, *B. pumilus* ATCC 7061T (Table 4-3). FO-033, FO-36b, and SAFN-037 had correlation values ranging from 0.62-0.71 with the type strain. All the other strains examined, except SAFN-029 and SAFN-036, had correlation values of >0.75 with the type strain. The highest correlation value (0.98) was found between the averaged spectra from ATCC 7061T in the library and a fresh batch of ATCC 7061T spores. Strains SAFN-029 and SAFN-036, had correlation values of 0.31 and 0.27 respectively indicating a low degree of spectral similarity with the *B. pumilus* type strain.

Table 4-3. Linear correlation values obtained when comparing the Bacillus species library with the *B. pumilus* strains in this study

Species/Strain		ATCC 9372 T *B. atrophaeus*	FO-11 *B. cereus*	ATCC 14580 T *B. licheniformis*	KL-196 *B. licheniformis*	ATCC 14581 T *B. megaterium*	ATCC 51516 T *B. mojavensis*	51-8C *B. niacini*	PTA-4399 T *B. odysseyi*	VSE1-06 *B. psychrodurans*	ATCC 7061 T *B. pumilus*	168 *B. subtilis*	ATCC 6051 T *B. subtilis*	ATCC 10792 T *B. thuringiensis*
B. pumilus	**SAFN-001**	0.22	0.02	0.25	0.05	0.39	0.33	0.00	0.28	0.10	0.85	0.06	0.08	0.11
B. pumilus	**FO-033**	0.25	0.02	0.28	0.06	0.40	0.39	0.00	0.28	0.06	0.71	0.15	0.15	0.10
B. pumilus	**FO-036b**	0.26	0.01	0.20	0.04	0.39	0.31	0.00	0.26	0.06	0.62	0.17	0.16	0.09
B. pumilus	**81-4C**	0.31	0.01	0.21	0.03	0.41	0.34	0.00	0.30	0.05	0.75	0.14	0.15	0.14
B. pumilus	**84-3C**	0.27	0.01	0.27	0.04	0.44	0.45	0.00	0.32	0.06	0.86	0.10	0.13	0.14
B. pumilus	**51-3C**	0.29	0.02	0.28	0.06	0.43	0.41	0.01	0.30	0.09	0.79	0.13	0.14	0.13
B. pumilus	**82-2C**	0.28	0.01	0.26	0.04	0.45	0.44	0.00	0.32	0.05	0.86	0.11	0.13	0.14
B. pumilus	**84-1C**	0.33	0.02	0.23	0.03	0.44	0.38	0.01	0.33	0.06	0.85	0.10	0.13	0.16
B. pumilus	**84-4C**	0.25	0.01	0.26	0.05	0.44	0.38	0.01	0.34	0.08	0.87	0.09	0.10	0.11
B. pumilus	**SAFN-027**	0.31	0.02	0.23	0.04	0.43	0.36	0.01	0.31	0.07	0.89	0.08	0.10	0.15
B. pumilus	**KL-052**	0.27	0.01	0.21	0.03	0.41	0.33	0.00	0.29	0.05	0.79	0.12	0.12	0.12
B. pumilus	**SAFN-036**	0.19	0.01	0.11	0.06	0.26	0.16	-0.01	0.15	0.11	0.27	0.21	0.16	0.01
B. pumilus	**SAFN-037**	0.35	0.02	0.20	0.05	0.39	0.31	0.03	0.30	0.09	0.68	0.14	0.17	0.16
B. pumilus	**SAFN-029**	0.13	0.01	0.11	0.02	0.16	0.12	0.02	0.14	0.04	0.31	0.26	0.27	0.07
B. pumilus	**SAFR-032**	0.13	0.01	0.17	0.04	0.41	0.34	0.04	0.37	0.15	0.75	-0.01	0.03	0.11
B. pumilus	**0105342-2**	0.19	0.01	0.30	0.04	0.42	0.44	0.03	0.38	0.07	0.94	0.01	0.09	0.12
B. pumilus	**ATCC 27142**	0.12	0.01	0.21	0.04	0.45	0.39	0.03	0.37	0.08	0.83	0.00	0.04	0.09
B. pumilus	**ATCC 7061T**	0.13	0.00	0.27	0.04	0.48	0.46	0.01	0.32	0.07	0.98	0.01	0.03	0.08

Furthermore, the average and individual MALDI-TOFMS spectra from all of the *B. pumilus* isolates were compared with themselves using the user created library, and the correlation results are shown in Table 4-4. MALDI-TOFMS spectra (*m/z* 2,500 to 35,000) of representative strains of type strain group (Figure 4-1), FO-36b group (Figure 4-2) and outlier group (Figure 4-3) are shown. The molecular mass region from approximately 10,000 to 35,000 Da is amplified by 10x to highlight the higher molecular weight peaks which are present at lower intensities. These higher molecular weight proteins may prove important as they are sometimes a strain level distinguishing feature in the spectra obtained from the spores. Using MALDI-TOFMS protein profiling with linear correlation and HCA analysis, the *B. pumilus* isolates in this study are clustered

into two groupings as seen with the *gyrB* and DNA-DNA hybridization analyses. The dendrogram and image map for all of the strains is shown in Figure 4-4.

Table 4-4. Correlation results based on MALDI-TOFMS protein profiles of the *B. pumilus* spore strains in this study

Species	Strain	SAFN-001	FO-033	FO-036b	81-4C	84-3C	51-3C	82-2C	84-1C	84-4C	SAFN-027	KL-052	SAFN-036	SAFN-037	SAFN-029	SAFR-032	0105342-2	ATCC 7061	ATCC 27142
B. pumilus	SAFN-001	-																	
B. pumilus	FO-033	0.96	-																
B. pumilus	FO-036b	0.82	0.90	-															
B. pumilus	81-4C	0.93	0.98	0.95	-														
B. pumilus	84-3C	0.94	0.99	0.86	0.96	-													
B. pumilus	51-3C	0.95	0.99	0.90	0.98	0.98	-												
B. pumilus	82-2C	0.95	0.99	0.88	0.97	0.99	0.99	-											
B. pumilus	84-1C	0.94	0.96	0.88	0.97	0.96	0.97	0.98	-										
B. pumilus	84-4C	0.95	0.97	0.86	0.95	0.97	0.97	0.97	0.97	-									
B. pumilus	SAFN-027	0.95	0.94	0.82	0.94	0.94	0.96	0.96	0.98	0.96	-								
B. pumilus	KL-052	0.94	0.97	0.92	0.99	0.95	0.98	0.96	0.96	0.96	0.96	-							
B. pumilus	SAFN-036	0.55	0.65	0.90	0.73	0.59	0.65	0.62	0.60	0.60	0.52	0.69	-						
B. pumilus	SAFN-037	0.89	0.94	0.96	0.98	0.92	0.96	0.94	0.95	0.92	0.91	0.95	0.80	-					
B. pumilus	SAFN-029	0.28	0.28	0.23	0.27	0.29	0.28	0.28	0.29	0.29	0.30	0.27	0.10	0.24	-				
B. pumilus	SAFR-032	0.63	0.57	0.42	0.55	0.62	0.59	0.64	0.68	0.65	0.70	0.56	0.17	0.53	0.30	-			
B. pumilus	0105342-2	0.84	0.83	0.55	0.74	0.87	0.82	0.85	0.82	0.85	0.84	0.76	0.20	0.66	0.33	0.71	-		
B. pumilus	ATCC 7061	0.85	0.82	0.58	0.76	0.86	0.82	0.86	0.85	0.87	0.89	0.80	0.24	0.69	0.32	0.75	0.93	-	
B. pumilus	ATCC 27142	0.69	0.65	0.46	0.61	0.70	0.66	0.72	0.74	0.72	0.76	0.63	0.16	0.57	0.32	0.94	0.82	0.84	-

The first grouping, the type strain group, contains *B. pumilus* ATCC 7061^{T}, and includes ATCC 27142, 0105342-2, and SAFR-032 with one outlier (SAFN-029). Strains in this group have characteristic peaks in their spectra at *m/z* 6,860 Da, 7,230 Da, and 9,605 Da (Figure 4-1). SAFN-029 (Figure 4-3D), has a correlation value of less than 0.32 for all of the strains examined and is considered an outlier. It is contained within the type strain cluster because of the presence of the three characteristic peaks described above. All of the strains in this type strain group can be differentiated from other *B.* pumilus strains tested at the 97% confidence interval using the Student's t-test. The second FO-36b group contains the remaining *B. pumilus* strains tested in this study. This FO-36b group also has peaks at *m/z* 6,860, 7,230, and 9,605 Da but all strains in this group have an additional peak at *m/z* 7,620 (Figure 4-2). The outlier strain,

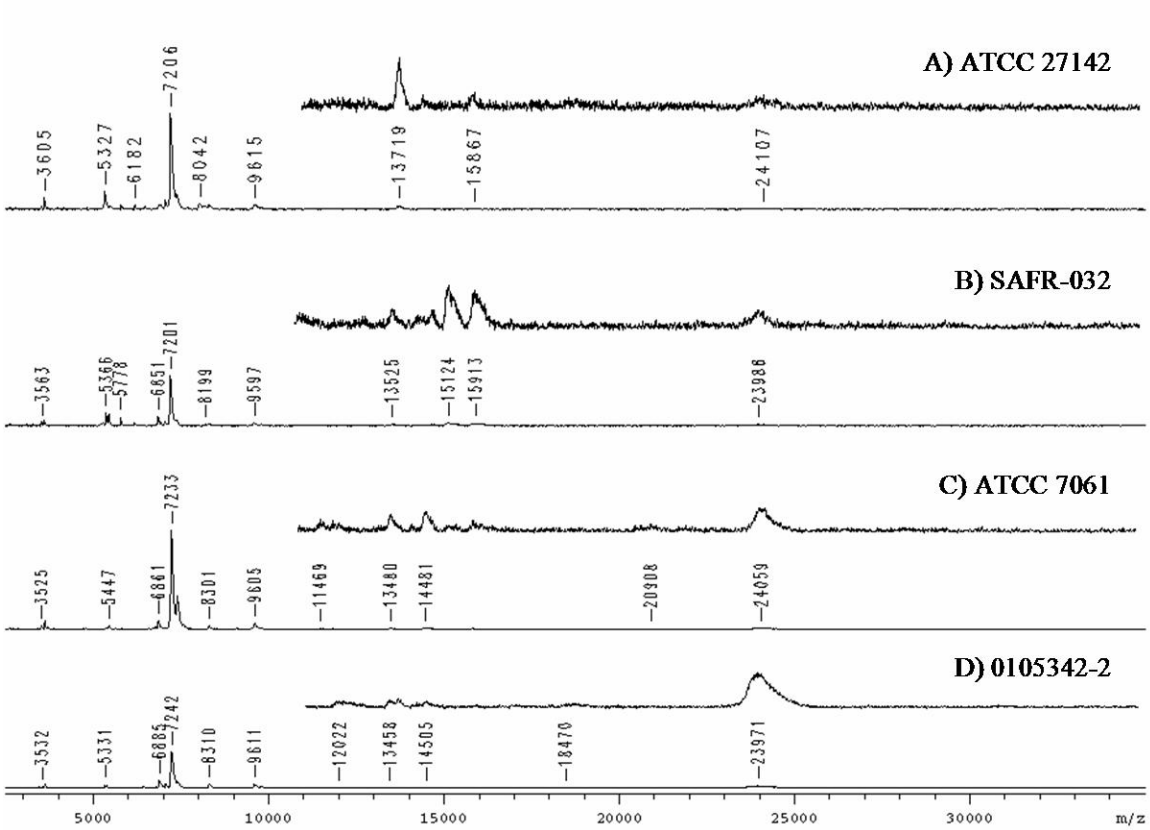

Figure 4-1. MALDI-TOFMS protein profiles of the *B. pumilus* type strain group spores. The mass range depicted is from *m/z* 3,000-35,000. The higher molecular mass region from *m/z* 9,500-25,000 is amplified 10x (see inset of each spectrum) in order to visualize the higher molecular weight peaks that are present but are at much lower abundance in the samples. A) ATCC 27142, B) SAFR-032, C) *B. pumilus* ATCC 7061^{T}, D) 0105342-2.

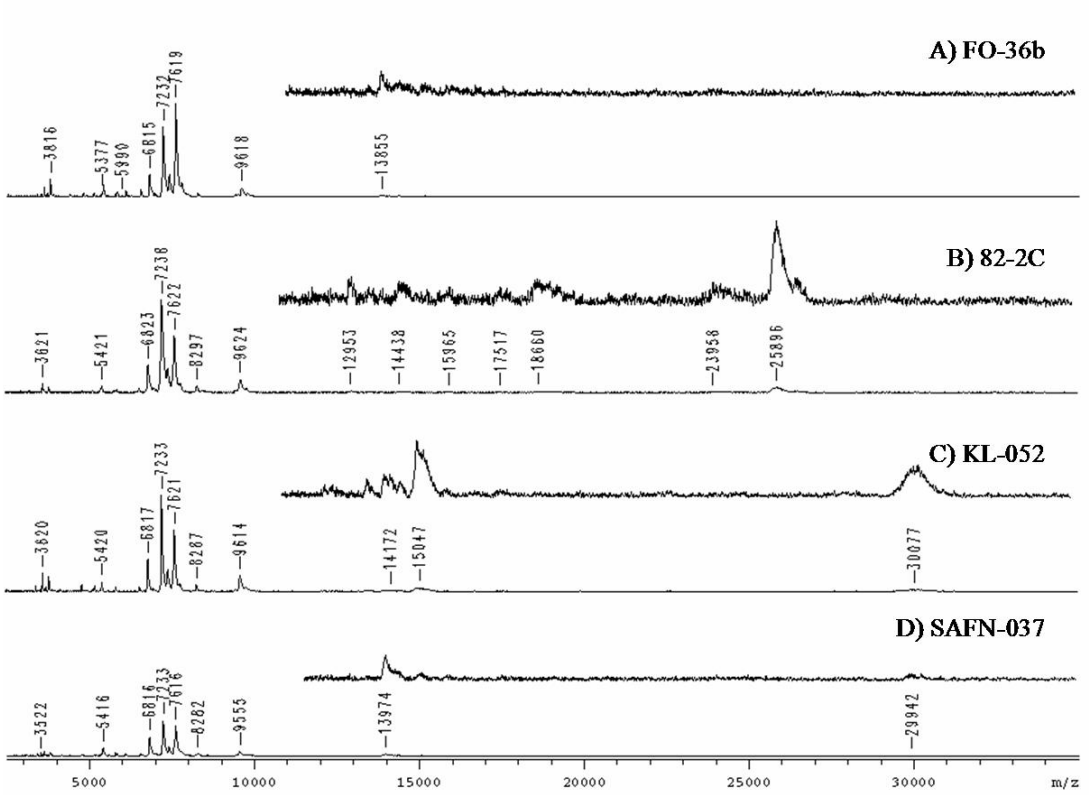

Figure 4-2. MALDI-TOFMS protein profiles from selected spores from the FO-36b cluster. A) FO-36b, B) 82-2C, C) KL-052, D) SAFN-037. Explanations are as given in Fig. 4-1.

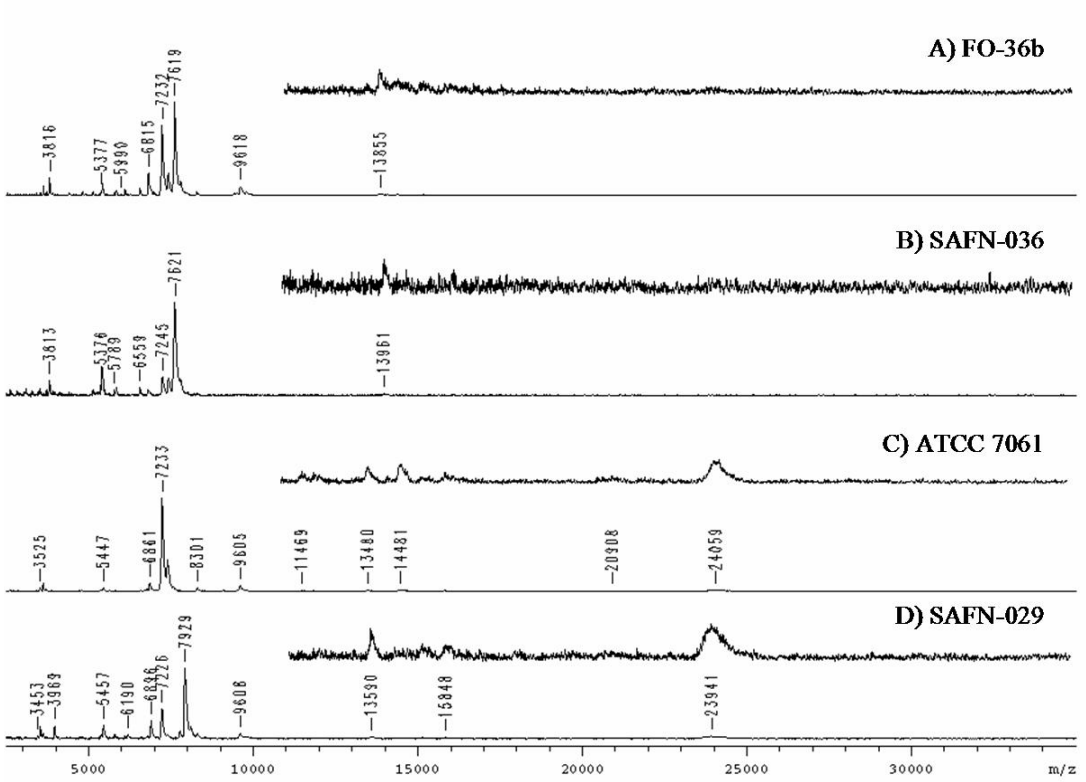

Figure 4-3. MALDI-TOFMS protein profiles comparing *B. pumilus* ATCC 7061T, FO-36b and the two outliers found in this study. A) FO-36b, B) SAFN-036, C) *B. pumilus* ATCC 7061T, D) SAFN-029. Explanations are as given in Fig. 4-1.

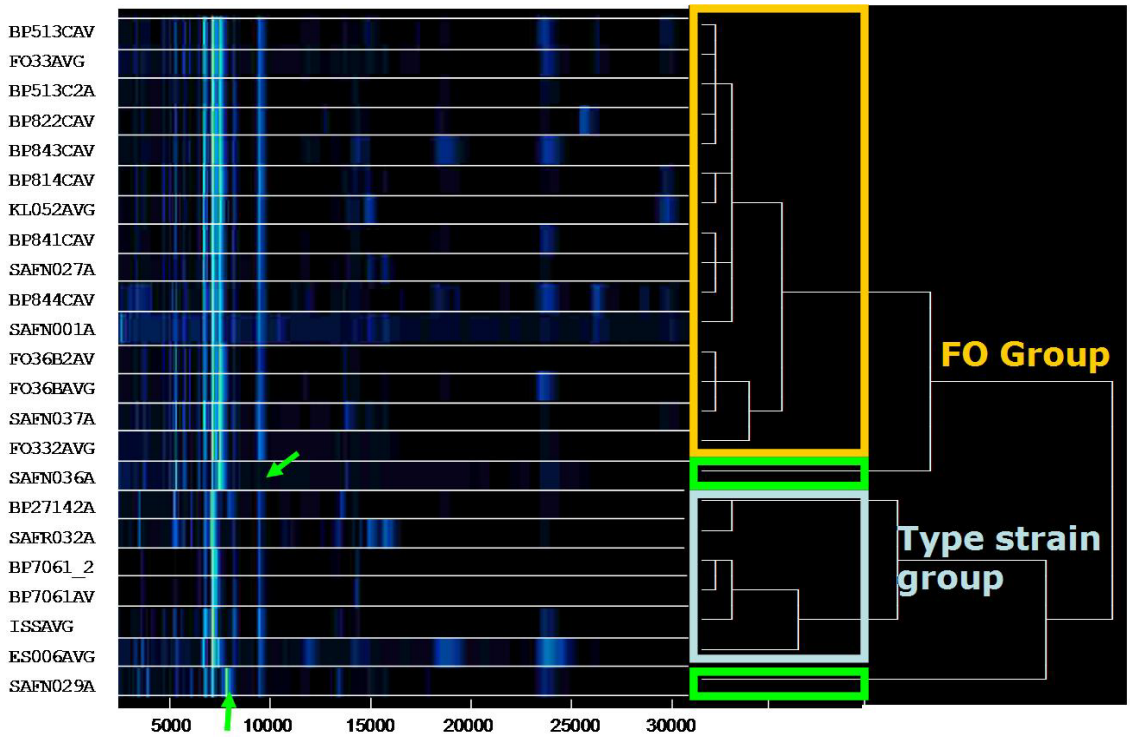

Figure 4-4. Dendrogram and visualization of the spore strains in this study. The spectra are represented as an image map where the intensity is indicated by the band color. The figure is highlighted to show the 2 clusters formed in this study as well as the outliers. The arrows indicate the presence and absence of peaks in the outliers.

SAFN-036, is also included in this FO-36b group. The SAFN-036 profile (Figure 4-3B) does not exhibit the peak at 9,605 Da, which is found in all of the other *B. pumilus* strains examined, but contains the 7,620 peak which is characteristic of this group. Correlation values of 0.52-0.90 are found between SAFN-036 and strains in this second cluster. In contrast to the type strain group, the FO-36 group is very tightly clustered. Many of the isolates in FO-36b group cannot be differentiated at the 97% confidence interval from the other strains in this cluster. However, strains of FO-36b group can be differentiated from strains in the type strain group.

MALDI-TOFMS Protein Profiling of Vegetative Cells

To further characterize the outliers from the spore form of the organism, 8 strains were cultured into vegetative cells and the correlation and HCA analysis was repeated on the spectra from the vegetative cells. *B. pumilus* ATCC 7061^{T}, ATCC 27142, 0105342-2, FO-36b, 82-2C, SAFN-037, and the outliers, SAFN-029 and SAFN-036 were selected for this analysis. The correlation results are shown in Table 4-5. As with the spore

Table 4-5. Correlation results based on MALDI-TOFMS protein profiles of selected *B. pumilus* vegetative cells in this study

Species	Strain	FO-036b	82-2C	SAFN-036	SAFN-037	SAFN-029	0105342-2	ATCC 7061	ATCC 27142
B. pumilus	**FO-036b**	-							
B. pumilus	**82-2C**	**0.74**	-						
B. pumilus	**SAFN-036**	**0.50**	**0.63**	-					
B. pumilus	**SAFN-037**	**0.46**	**0.55**	**0.90**	-				
B. pumilus	**SAFN-029**	**0.38**	**0.37**	**0.35**	**0.37**	-			
B. pumilus	**0105342-2**	**0.29**	**0.37**	**0.28**	**0.28**	**0.78**	-		
B. pumilus	**ATCC 7061**	**0.40**	**0.41**	**0.37**	**0.35**	**0.93**	**0.77**	-	
B. pumilus	**ATCC 27142**	**0.46**	**0.49**	**0.43**	**0.44**	**0.84**	**0.67**	**0.90**	-

samples the vegetative cells are clustered into 2 groups, although the groupings are not as close as the spore clusters. One cluster contains the type strain *B. pumilus* ATCC 7061^T along with ATCC 27142, 0105342-2, and SAFN-029 (Figure 4-5). Among these samples the correlation values ranged from 0.67-0.93. The second group contains FO-36b, 82-2C, SAFN-037, and SAFN-036 (Figure 4-6). The correlation values for this grouping were slightly lower ranging from 0.46-0.90 with the lowest values being evident for SAFN-037 and SAFN-036 with FO36b. SAFN-037 and SAFN-036 have a correlation value of 0.90 with each other.

The spectra of all the vegetative cells are very similar in the region below 11 kDa (Figure 4-5 and 4-6). All of the spectra have biomarker peaks at 9,810 Da and approximately 10,716 Da. In this region the type strain group has a distinct biomarker peak at 5,350 Da while in the FO group the peak is at 5,390 Da. This is one of the most profound differences between the 2 groups in the vegetative cell state. In the higher molecular weight region there are no biomarker peaks observed that are common within the groupings. There appears to be significant within-group variation in this region. Replicate culturing of the vegetative cells gave nearly identical results to the initial trial. Comparison of the replicate trials gave correlation values of above 0.80 for strains within the type strain group and above 0.60 for those within the FO group.

Discussion

To adopt MALDI-TOFMS as a new methodology for identifying closely related bacterial species, it is necessary to compare this technique with several existing genotypic and phenotypic methods. Biolog metabolic fingerprinting did not correctly identify most of the *B. pumilus* isolates tested. Of the genotypic analyses performed, 16S rDNA proved to be the least discriminating for the tested *B. pumilus* isolates. Unlike the slowly

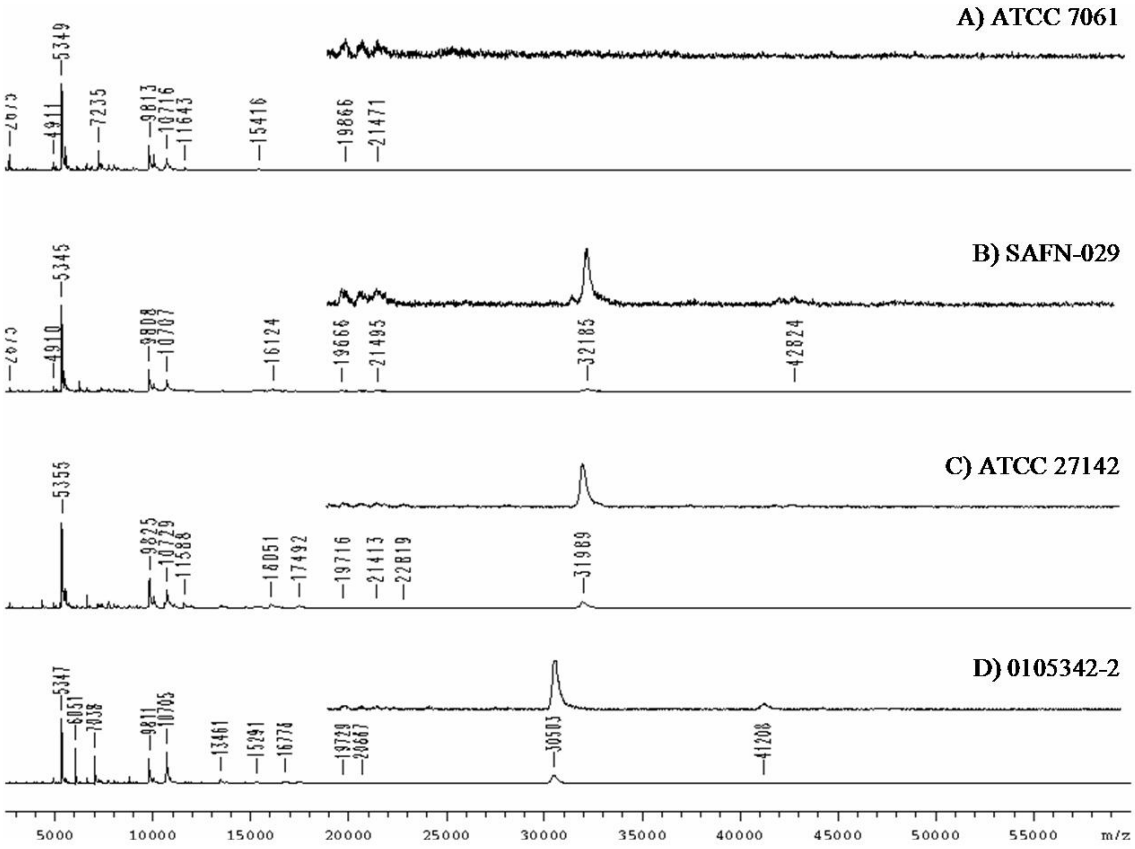

Figure 4-5. MALDI-TOFMS protein profiles of the *B. pumilus* type strain group vegetative cells. The mass range depicted is from *m/z* 2,500-60,000. The higher molecular mass region from *m/z* 20,000-60,000 is amplified 4x (see inset of each spectrum) in order to visualize the higher molecular weight peaks. A) *B. pumilus* ATCC 7061^{T}. B) SAFN-029. C) *B. pumilus* ATCC 27142. D) 0105342-2.

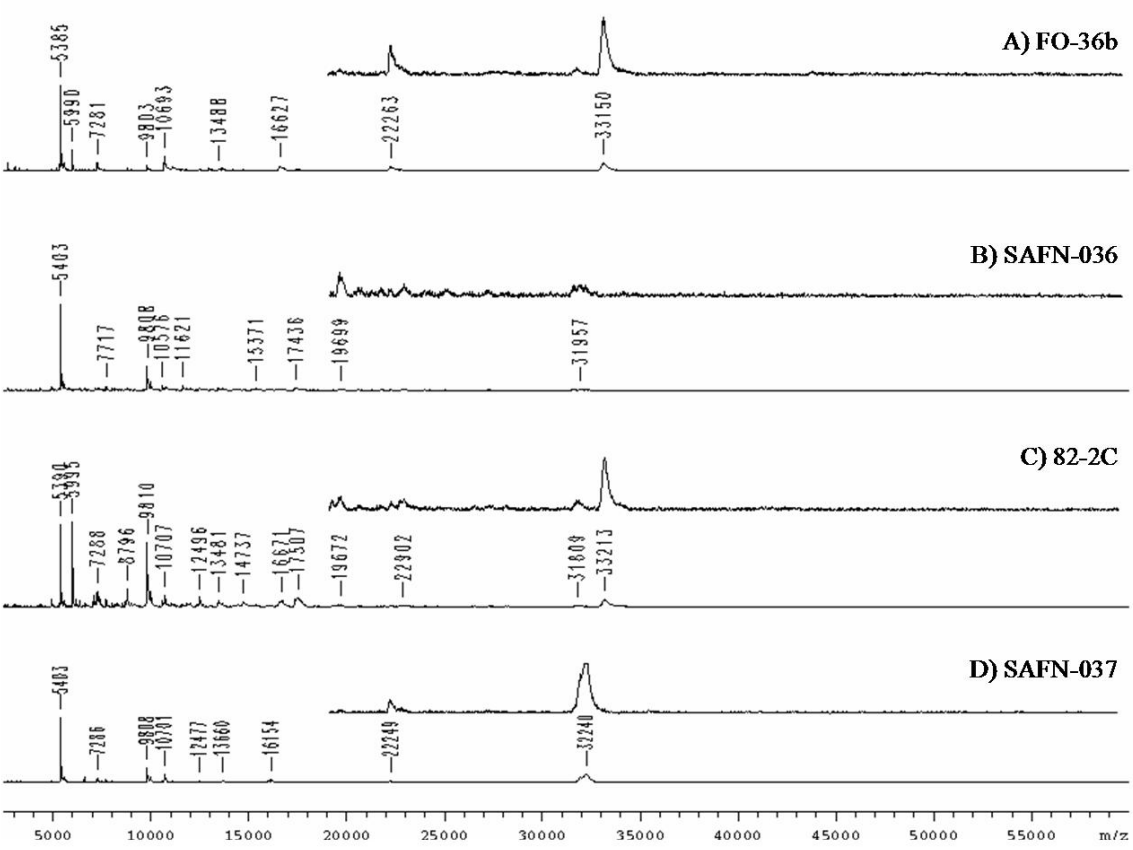

Figure 4-6. MALDI-TOFMS protein profiles of the FO group vegetative cells. The mass range depicted is from *m/z* 2,500-60,000. The higher molecular mass region from *m/z* 20,000-60,000 is amplified 4x (see inset of each spectrum) in order to visualize the higher molecular weight peaks. A) FO-36b. B) SAFN-036. C) 82-2C. D) SAFN-037.

evolving ribosomal RNA gene, the more rapidly evolving *gyrB* gene allowed for distinct clustering of the tested strains into two groups, which are in strong agreement with both the DNA-DNA hybridization and MALDI-TOFMS protein profiling.

To identify the unknown isolates based on MALDI-TOFMS protein profiling, both the linear correlation value and visual peak comparisons were used as a diagnostic tool for differentiating a group of strains. Although high correlation values (typically >0.75) are desirable for definitive species differentiation, low correlation values do not unambiguously exclude an isolate. Rather the delta value for the first and second highest hits should be considered the significant measure of the validity of the bacterial species assignment. The MALDI-TOFMS protein profile from the unknown strains was compared with a library of type strains of different species. The *B. pumilus* type strain was the species with the highest correlation value for every putative *B. pumilus* isolate tested including SAFN-029 and SAFN-036. In all cases except SAFN-029 and SAFN-036, the delta value between the first and second hit was >0.23 (Table 3).

As seen in the *gyrB* and DNA-DNA hybridization methodologies, MALDI-TOFMS analysis exhibited two very distinct and consistent groups among the *B. pumilus* isolates tested. According to the *Ad Hoc* Committee on Bacterial Systematics,[83] DNA-DNA re-association values of >70% between the type strain and the unknown strain are considered as the same species. The 53 to 66% DNA:DNA hybridization values as well as groupings exhibited by MALDI-TOFMS among the isolates tested in this study suggest that these two groups represent a different species of *Bacillus*. However, description of a novel species is beyond the scope of this study.

The most notable difference in the MALDI-TOFMS protein profile between the type strain and FO-36b group is the presence of a peak at 7,620 Da, found only in the FO-36b group strains. Another difference between the two groups is that the FO-36b group forms a tight cluster, which cannot be differentiated at the strain level based solely on MALDI-TOFMS protein profiling. In contrast, the type strain group can be differentiated at the strain level indicating a higher level of strain variation within this cluster. To determine the reproducibility of the MALDI-TOFMS analysis, duplicate analysis of spores from each isolate was performed and gave correlation values ranging from 0.87-0.98 for strains within the FO-36b group and 0.89-0.96 from the type strain group. Analysis of duplicate batches of 10 different strains prepared by various technicians yielded a correlation values of >0.75 for strains within the same group.

SAFN-036 and SAFN-029 are the two species not convincingly identified as either a member of the *B. pumilus* type strain group or the FO-36b group solely by the linear correlation analysis and visual interpretation of the peaks in the spectra was required. In both cases, there is not another type strain within the library that was a closer match to these isolates (Table 4-3). The 9,605 Da peak, absent in the initial SAFN-036 spore batch, was observed at low intensities in subsequent analysis of this isolate from a different batch of spores. The absence of it in the first analysis could be due either to a difference in spore formation or more likely to a problem with the concentration of spores in the first analysis. Although SAFN-036 had a very low correlation value (0.27) with the type strain, ATCC 7061^T, it had correlation values ranging from 0.52-0.90 (Table 4-4) for strains in the FO-36b group. The spectrum from SAFN-029 (Figure 4-3D) has the 3 biomarker peaks in common with ATCC 7061^T cluster; however, its most intense peak at

7,929 Da, not found in any of the other *B. pumilus* spectra, contributes to its low correlation values thus making it an outlier. This intense 7,929 Da peak has not been observed in the spectra of any of the other *Bacillus* species we have examined. Subsequent analysis including a different sporulation of this isolate produced spectra that contained the intense 7,929 Da peak and had correlation values ranging from 0.83-0.91 with the SAFN-029 spectra in the library.

Analysis of a subset of these strain as vegetative cells supported the assignment of the spore outliers to the groupings above. In the vegetative state, SAFN-029 was clearly clustered with the type strain group and had a correlation value of 0.93 (Table 4-5) with *B. pumilus* ATCC 7061^{T}. SAFN-036 was more loosely associated with FO-36b, r = 0.50, but was closely clustered with SAFN-037. Both strains were within the FO-36b cluster due to the presence of the biomarker at 5,390 Da (Figure 4-6). The vegetative cell spectra appear to have more strain variability than the spore spectra for the strains analyzed. Although not addressed in the current study, the intra-species with vegetative cells variation may necessitate a different acceptance number for the correlation value; perhaps $r < 0.50$ for vegetative cells with a delta value of 0.1 would be more indicative of the variation seen in vegetative cell spectra. Further investigation of this is warranted.

The linear correlation analysis and HCA used was sufficient for the confident classification of 14/16 of the spores in this study. It proved critical to take into account the correlation value and the presence and absence of certain biomarker peaks to understand the placement of the two outliers. This limitation and the variability of the absolute correlation value presents a challenge for determining definitive species classification based solely on linear correlation analysis. Software developed at the

Pacific Northwest National Laboratory is currently being implemented in an attempt to overcome these limitations.[84,85] This software uses a peak-picking based algorithm to differentiate MALDI-TOFMS protein profiles and has been shown to be effective for bacterial differentiation of single isolates and of mixtures.[72,86] In addition to improved pattern recognition, the identification of protein peaks which are useful for bacterial systematics, such as the 7,620 Da peak, would be valuable for the definitive differentiation of a given species.

Conclusion

MALDI-TOFMS protein profiling has been demonstrated as a useful taxonomic tool for differentiating spores and vegetative cells of the genus *Bacillus.* This methodology is far more accurate than metabolic profiling, more discriminating than 16S rDNA sequence analysis, and complements the results of *gyrB* sequence analysis and DNA-DNA hybridization. When compared with the genotypic methods used here, the MALDI-TOFMS analysis is much more rapid for isolate differentiation, taking only a few minutes to prepare and analyze. In addition, new isolates can be compared with established libraries to obtain species level identification and to elucidate relationships between bacterial strains, eliminating the necessity of maintaining and growing reference strains for subsequent studies. The addition of automated peak picking algorithms that recognize species-specific biomarker peaks will further strengthen the diagnostic power of this tool.

CHAPTER 5
MALDI-TOFMS PROTEIN PROFILING OF *Bacillus anthracis-cereus-thuringiensis* GROUP SPORES

Introduction

A great deal of attention has been focused recently on the MALDI-TOFMS analysis of *B. anthracis* spores. *B. anthracis*, the causative agent of the disease anthrax, is closely related to a group of bacteria that includes *B. cereus* and *B. thuringiensis*. *B. cereus* is often the organism implicated in food poisoning outbreaks and causes great concern in the food processing and dairy industries. *B. thuringiensis,* by contrast, is of great importance in the agricultural industry as an insecticide used in crop sprays and pesticide treatments and is generally though to be harmless to humans. The precise discrimination and classification of this group of bacteria remains a topic of great debate. The debate is fueled on one side by taxonomists who seek to establish evolutionary relationships and lineages between bacteria. On the other side are practitioners and bacteriologists who are primarily concerned with the ability to deduce the pathogenic, spoilage, or ecological properties associated with a new isolate.[2]

Phenotypic identification of these species depends on virulence factors, including the genes encoding for the Cry toxin crystals in *B. thuringiensis*, the toxin and capsule of *B. anthracis*, and the emetic and enterotoxin genes of *B. cereus*.[2] These genes are encoded by extrachromosomal mobile genetic elements such as plasmids that can be lost or involved in lateral gene transfer between species.[2,87,88] Sequence analysis of the 16S rDNA is not able to differentiate the species within this group.[55,88,89] The use of modern

genotypic methods such as PCR of virulence factors,[90,91] restriction fragment length polymorphisms,[92,93] pulsed-field gel electrophoresis,[2] and analysis of intergenic spacer regions[94] are able to differentiate *B. anthracis* from *B. cereus* but are unable to differentiate *B. cereus* from *B. thuringiensis*. Because of the ambiguities in the phenotypic and genotypic characterization of these organisms, it has been suggested that the three species be considered as one species evolved from a common ancestor.[89,95]

The need to differentiate this group of bacteria, and particularly discriminate potential pathogenic and nonpathogenic strains, regardless of their correct phylogenetic classification, lingers as the goal of most rapid, phenotype-based identification techniques. The many attempts to design rapid methods for this task such as commercial identification tests (API, Biolog, Vitek, and the Microbial ID system) have fallen short in their ability to effectively differentiate this closely related group of bacteria.[2] The clear, unambiguous identification and characterization of members in this group is currently an unmet diagnostic challenge.

MALDI-TOFMS-based protein profiling has demonstrated the ability to differentiate closely related groups of bacteria for a set of *B. pumilus* isolates.[96] However, MALDI-TOFMS based investigations of whole spores from the BACT group to date have been limited in both scope and in the number of strains examined.[25,29,32] The most prominent biomarker peaks highlighted in these studies were in the 6,000-8,000 Da range, with lower weight biomarkers observed in some cases between 2,000-4,000 Da. The lower molecular weight components have been attributed to microbial lipopeptides and spore cortex peptidoglycan while the 6,000-8,000 Da peaks are from small acid soluble proteins (SASPs).[25,32,40,97,98] Little similarity can be seen in the biomarker peaks,

even for identical strains between the different studies. All have failed to effectively differentiate the group, although biomarkers unique to *B. anthracis* strains are reported using both whole proteins and trypsin digests of SASPs.[25,29,32,38,99,100] However, the *B. cereus* and *B. thuringiensis* strains are not differentiated. This chapter reports on an attempt to validate the use of MALDI-TOFMS protein profiling as a phenotypic discriminator for this group of bacteria. A set of 26 unique strains of *B. anthracis, B. cereus,* and *B. thuringiensis* (BACT) were examined for their proteomic-based affiliations.

Materials and Methods

Bacterial Strains

Table 5-1 contains a list of the *B. cereus* isolate and serotype strains used in this study with both the DNA:DNA hybridization and 16S rDNA results reported by LaDuc *et al.*[88] All ATCC strains were procured from the American Type Culture Collection (Manassas, VA) which included *B. cereus* ATCC 14579, *B. mycoides* ATCC 6462, *B. thuringiensis* Berliner ATCC 10792^T, and the type strains contained in the reference libraries. *B. thuringiensis* Kurstaki HD-1, *B. thuringiensis* Aizawai, *B. thuringiensis* Galleriae, *B. thuringiensis* Israeliensis, *B. anthracis* 34F2, and the collection of *B. cereus* serotypes were provided by M. Satomi at the National Institute of Fisheries, Japan.

Sporulation of *Bacillus* Isolates

Two protocols were used for the productions of spores in this study. For the first protocol, production of spores was performed as described previously.[7,101] Briefly, a single purified colony of the strain to be sporulated was inoculated into nutrient broth sporulation medium (NSM) and was incubated at 32^{o}C under shaking conditions for 2-4 days. Spore cultures were harvested once they attained >99% spores as examined using

phase-contrast microscopy. In a second protocol, tryptic soy agar (TSA) plates or nutrient agar (NA) plates were inoculated and the plates were allowed to sit at room

Table 5-1. List of *B. cereus* serotype strains

Serotype[a]	Strain #[b]	Source	Country	Origin	% similarity in 16S rDNA[c]		% similarity in DNA:DNA[c]	
					B. cereus	*B. anthracis*	*B. cereus*	*B. anthracis*
JPL	FO-11	Air particles	USA	JPL-SAF-179	99.0	99.0	53	71
H5	186	Fried rice	Great Britain	PHLS, Great Britain	99.3	99.8	55	71
H6	CAN30	Barbecued chicken	Canada	E. Todd, FRI, Health & Welfare Canada	99.1	99.2	53	78
H7	277	Curry powder	Great Britain	PHLS, Great Britain	99.3	99.7	53	75
H17	202	Uncooked rice	Great Britain	PHLS, Great Britain	99.2	99.7	56	75
H2	4433	Meat loaf	USA	J.M. Goepfert, FRI, UW-Madison	99.3	99.9	59	67
H3	214	Boiled rice	Great Britain	PHLS, Great Britain	99.1	99.7	49	69
H8	4431	Indonesian rice dish	Netherlands	J.M. Goepfert, FRI, UW-Madison	99.2	99.7	52	71
H9	4429	Vanilla pudding	Netherlands	J.M. Goepfert, FRI, UW-Madison	98.2	98.4	55	66
H12	118	Risotto	Great Britain	PHLS, Great Britain	99.2	99.7	49	76
H16	RR43	Uncooked Rice	Great Britian	PHLS, Great Britain	99.2	99.7	56	75
H10	4432	Indonesian rice dish	Netherlands	J.M. Goepfert, FRI, UW-Madison	99.3	99.7	72	50
H11	2140	Neonatal brain abscess	Great Britain	PHLS, Great Britain	99.4	99.8	77	48
H13	167	Prawn curry and rice	Great Britain	PHLS, Great Britain	99.5	99.7	70	53
H14	262	Fried rice	Great Britain	PHLS, Great Britain	99.4	99.8	88	56
H15	RR60	Uncooked rice	Great Britain	PHLS, Great Britain	99.3	99.9	81	56
H18	6833	Uncooked rice	Great Britain	PHLS, Great Britain	99.2	99.7	58	40

a Taylor *et al.*, 1975; b Original strain designiation; c LaDuc *et al.*, 2004

*These strains were originally classified as *B. cereus* based on 16S rDNA sequence analysis. Highlighted are the strains which DNA hybridization showed higher similarities with *B. anthracis or B. cereus*. Further studies on the toxigenic properties of thes

temperature for 1-2 weeks until the plates contained >90% spores. The center of the colonies were removed with a sterile loop and placed in sterile water. After harvesting, spores were further purified to remove remnant vegetative cells or cell debris as previously reported.[102] The purified spores were adjusted to give an optical density of 0.6 at 600 nm and were suspended in sterile deionized water and stored at 4^{o}C in glass tubes until analyzed.

Preparation of Vegetative Cells

A stock culture of each *Bacillus* strain was streaked for isolation on tryptic soy agar (TSA) plates and/or nutrient agar (NA) plates. The plates were incubated at 32^{o}C for 16 hours. Single purified colonies were removed from the plate with a sterile loop and were

placed in 100 µL of a phosphate buffered saline solution. Most colonies were approximately 2 mm in diameter. Cells were removed from the agar with a sterile loop and were vortexed in phosphate buffered saline for 15 minutes and pelleted by centrifugation for 10 minutes at 9600 x g. The supernatant was removed and the cell pellet was used for subsequent analysis.

Fatty Acid Methyl Ester (FAME) Analysis

FAME analysis was performed on the following strains in this study: *B. anthracis* 34F2, *B. thuringiensis* Berliner ATCC 10792^{T}, *B. thuringiensis* Kurstaki HD-1, and the *B. cereus* serotypes H3, H5, H7, H10, H15, H16, H18. FAME analysis was performed using the commercially available Sherlock Identification System by MIDI, Inc. (Newark, DE). The Sherlock Identification System is a fully automated gas chromatographic system which identifies bacteria based on their unique fatty acid profiles. To ensure accurate and reproducible results, protocols for FAME analysis were followed according to the technical note #101 available on the manufacturer's website (www.midi-inc.com/media/pdfs/TechNote_101.pdf, last viewed on April 12 2004).

Cells were cultured for analysis on TSA infused with 5% defibrinated sheep blood for 24 hours at 35°C. FAME analysis was performed on an Agilent technologies 6890N gas chromatograph with a flame ionization detector (FID). An Agilent Ultra 2 (cross linked 5% PH ME Siloxane) capillary column (25 M long, 0.2 mm ID) was used for separation of the fatty acid methyl esters. The peaks from the chromatograph were integrated on a PC and the fatty acid methyl ester composition of the sample is compared with the bioterrorism database using Sherlock pattern recognition software. The bioterrorism library provided by MIDI contained *B. cereus*, *B. anthracis*, and *B. thuringiensis* species and 6 other *Bacillus* challenge organisms.

MALDI-TOFMS Protein Profiling

Samples were prepared and analyzed using MALDI-TOFMS as described previously.[101] Briefly, a saturated matrix solution was prepared by dissolving 20 mg of ferulic acid in a solution of 30% acetonitrile, 30% formic acid. For the spore suspensions (normal and autoclaved), a 2.5 µL aliquot of the spore suspension (0.6 OD660) was added to 22.5 µL of the matrix solution. For the preparation of vegetative cells, 25 µL of the matrix solution was mixed directly with the cell pellet. This solution was sonicated for 3 minutes and then vortexed for 3 minutes. A 1 µL aliquot of the spore or vegetative sample was then placed on the MALDI plate for analysis. Spots were allowed to air dry. No further treatments were applied to the spots once dried. The sample preparation was done in duplicate. An additional MALDI analysis was performed as described above except a 5% trifluoroacetic acid (TFA) and 70% acetonitrile solution was used as the matrix solvent to facilitate the release and analysis of small acid soluble proteins (SASPs) from the spores.

MALDI-TOFMS protein profiling was performed on a Bruker Daltonics Reflex II Mass Spectrometer (Bruker Daltonics, Billerica, MA) retrofitted with delayed extraction. Positive ions were collected in the linear mode using a delay time of 50 ns, an acceleration voltage of 20kV, and a deflector set at 2,500 Da. All spectra represent the accumulation of 50 laser shots. Ten spectra were collected across each spot for a total of 20 spectra per sample.

Statistical Processing

Protocols for spectral processing, library spectra creation, and statistical analysis have been described elsewhere.[101] A library of BACT group spore or vegetative cell spectra was created by taking the average spectra from the 20 spectra collected for each

strain. The individual and the average spectrum obtained from the BACT strains were compared to MALDI-TOFMS profiles stored in the BACT library and to a user generated library of eleven *Bacillus* species.

Based on previous analysis of *B. pumilus* isolates, correlation values (r) of ≥ 0.75 with a delta value between the first and second hit of ≥ 0.1 are regarded as a good match for inclusion of a strain within a species.[96] Correlation values between 0.50-0.75 with delta values ≥ 0.1 may be considered for inclusion within a species providing peaks identified as species specific biomarkers are preserved. Strains with correlation values <0.50 that maintain sufficient delta values and biomarkers may be acceptable but represent an atypical strain of a species and at present require additional manual analysis for species-specific biomarker peaks.

In addition to the linear correlation analysis, SPSS software (Chicago, IL) was used for performing a hierarchal cluster analysis (HCA) on the data. The HCA analysis was based on the Pearson correlation and dendrograms were produced using a single linkage (nearest neighbor) scheme. To help visualize the peak patterns for the spectra, Surfer 8.0 from Golden software (Golden, CO) was used to create an image map with 10 Da resolution from the average spectra for each species. The image map provides a 3-D representation of the spectra where the color of the band is indicative of peak intensity. Clustering and visualization of the spectra was helpful for addressing atypical strains and could be used to justify the inclusion of a strain with low correlation values within a species.

Results

Sporulation of *Bacillus* Isolates

The efficient sporulation of many of the BACT strains proved to be difficult using the standard NSB sporulation protocol. In most cases, engulfed spores within the mother cell could be visualized with phase contrast microscopy. Subsequent lysozyme treatment did not efficiently break the peptidoglycan bonds in the mother cell to allow for release of the spores. Allowing the cell cultures to sit on culture medium for 1-2 weeks resulted in spores that were completely released from the mother cells. Therefore, the plate preparation with TSA media was used for most of the spores in this study. We were unable to obtain spores from *B. mycoides* ATCC 6462^{T} and the *B. cereus* serotypes H6, H9, and H16 using both protocols and only vegetative cell MALDI data is presented for these strains.

FAME Analysis

FAME analysis was performed on 10 of the strains is this study and the resulting similarity index values obtained against organisms in the bioterrorism library are shown in Table 5-2. Criteria that were considered for a good library match were a similarity index of >0.5 and a difference between the first and second match of >0.1, as recommended by the manufacturer. Of the 10 strains analyzed, only 4 met the criteria for consideration as a good library match; the delta values for these strains were above 0.1 and are highlighted in black in the table. These four stains were *B. anthracis* 34F2 identified as *B. anthracis*, the *B. cereus* type strain identified as *B. cereus* subgroup A, the *B. thuringiensis* serotype Kurstaki HD-1 identified as *B. cereus* subgroup B, and the *B. cereus* serotype H7 identified as *B. thuringiensis* subgroup A. The remaining *B. cereus* serotypes and the *B. thuringiensis* type strain did not meet the criteria for a good library

match and generally have high similarity indices with at least two of the species in the reference library.

Cluster analysis of the FAME data using the Sherlock system produced two clusters (data not shown). One contained *B. anthracis* 34F2, H5, and H18. The second cluster contained the *B. thuringiensis* type strain, *B. cereus* type strain, *B. thuringiensis* Kurstaki, H10, H16, H3, H15, and H7.

Table 5-2. Results of FAME analysis for selected BACT strains

EntryName	*B. anthracis* GC subgroup A	*B. anthracis* GC subgroup B	*B. cereus* GC subgroup A	*B. cereus* GC subgroup B	*B. thuringiensis* GC subgroup B	*B. mycoides*	Delta Value
B. anthracis 34F2	0.916		0.308	0.278	0.291		0.608
B. cereus serotype							
H5	0.538			0.532	0.365		0.006
H7	0.474			0.535	0.745	0.474	0.210
H3				0.595	0.66	0.521	0.065
H16			0.786	0.633	0.751		0.035
H10			0.842	0.693	0.767		0.075
H15	0.381		0.466	0.625	0.617	0.425	0.008
H18	0.505	0.307		0.458			0.047
B. cereus ATCC 14579 ^{T}	0.619		0.873		0.613		0.254
B. thuringiensis serotypes:							
Berliner IAM 12077^{T}			0.777	0.634	0.693		0.084
Kurstaki HD-1			0.528	0.798	0.581		0.217

MALDI-TOFMS Protein Profiling of BACT Spores

Using 30% formic acid and 30% acetonitrile as a solvent for analysis, spores in this study produced spectra with diagnostic peaks in the molecular weight range 2-35 kDa. Average spectra from each of the species in this study are shown in Figure 5-1. The most intense peaks in the spectra were typically found from 3-5 kDa although some species, notably the *B. thuringiensis* type strain and several of the serotypes including Galleriae, Aizawaii, and Kurstaki, all had a significant contribution to the spectra from a peak at ~19kDa. The molecular weight peaks above 10 kDa were not observed in these strains in other published studies.[25,29,32]

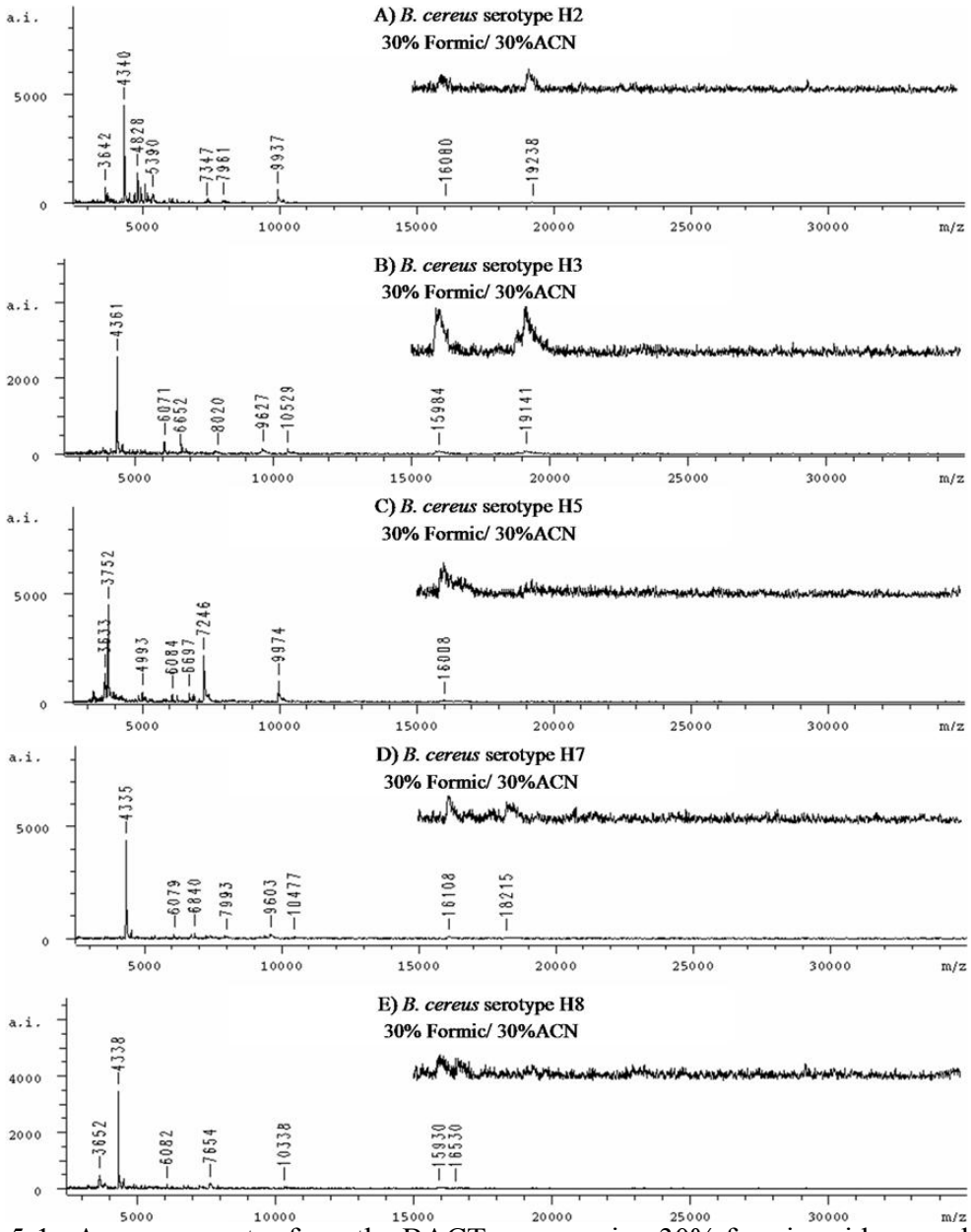

Figure 5-1. Average spectra from the BACT spores using 30% formic acid as a solvent. The mass range depicted is from *m/z* 2,500-35,000. The higher molecular weight region from *m/z* 13,000- 35,000 is amplified 6x in the inset of each spectrum to help visualize the higher molecular weight proteins. A) *B. cereus* serotype H2. B) *B. cereus* serotype H3. C) *B. cereus* serotype H5. D) *B. cereus* serotype H7. E) *B. cereus* serotype H8. F) *B. cereus* serotype H10. G) *B. cereus* serotype H11. H) *B. cereus* serotype H12. I) *B. cereus* serotype H14. J) *B. cereus* serotype H15. K) *B. cereus* serotype H17. L) *B. cereus* serotype H18. M) FO-11. N) *B. thuringiensis* serotype Israeliensis. O) *B. thuringiensis* serotype Aizawai. P) *B. thuringiensis* serotype Kurstaki HD-1. Q) *B. thuringiensis* serotype Galleriae. R) *B. cereus* ATCC 14579^{T}. S) *B. anthracis* 34F2. T) *B. thuringiensis* Berliner ATCC 14579^{T}.

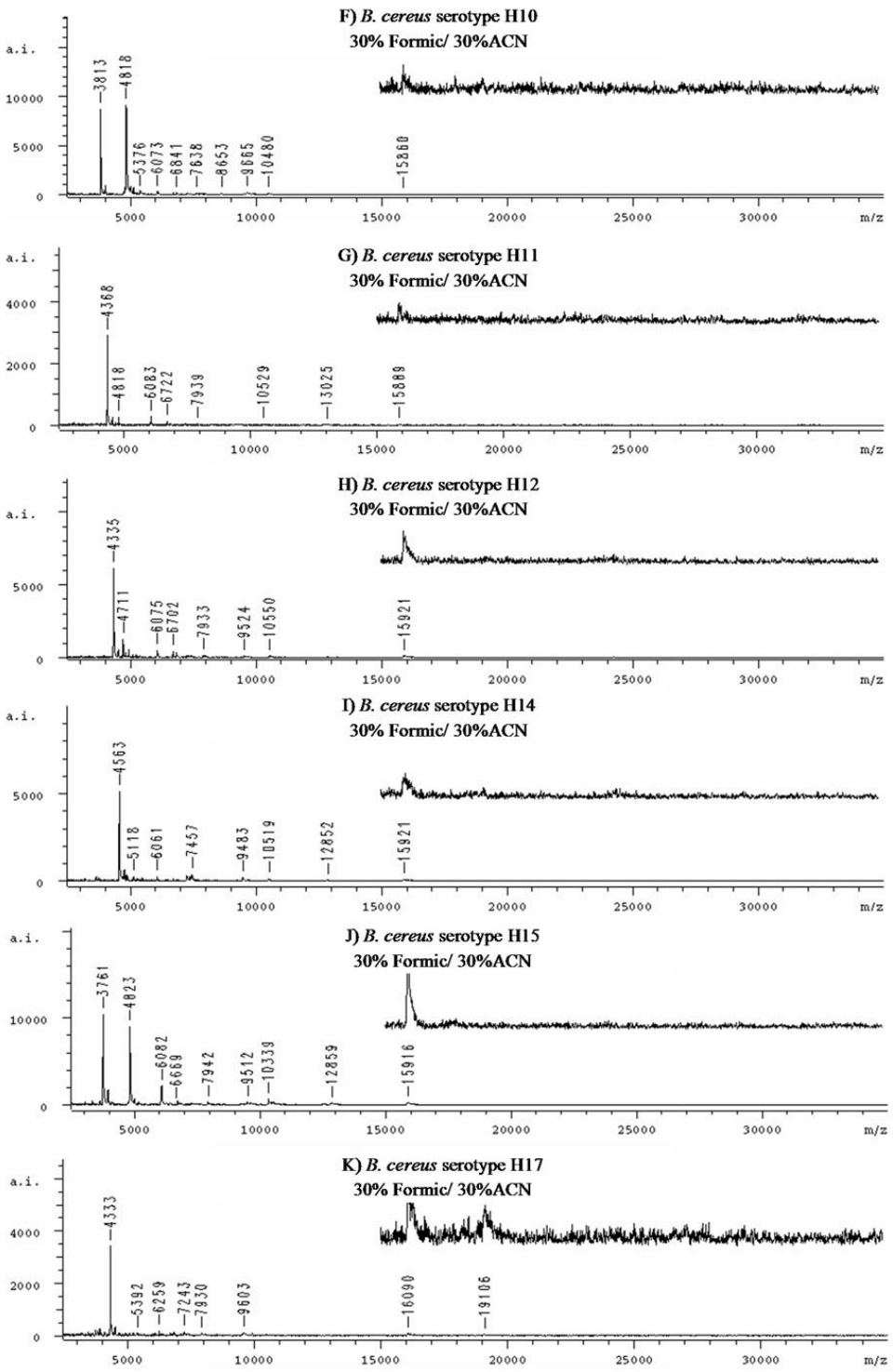

Figure 5-1. Continued.

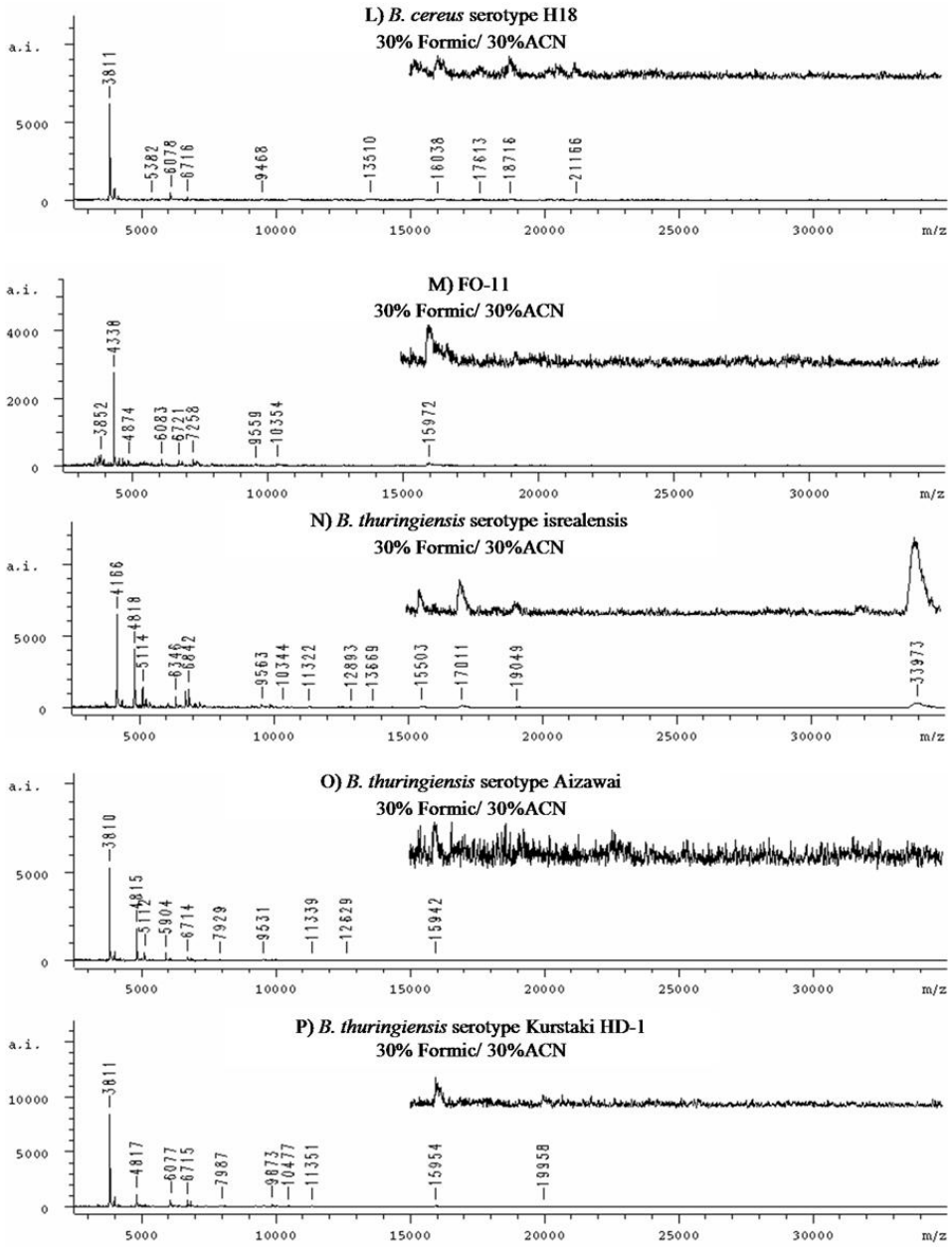

Figure 5-1. Continued.

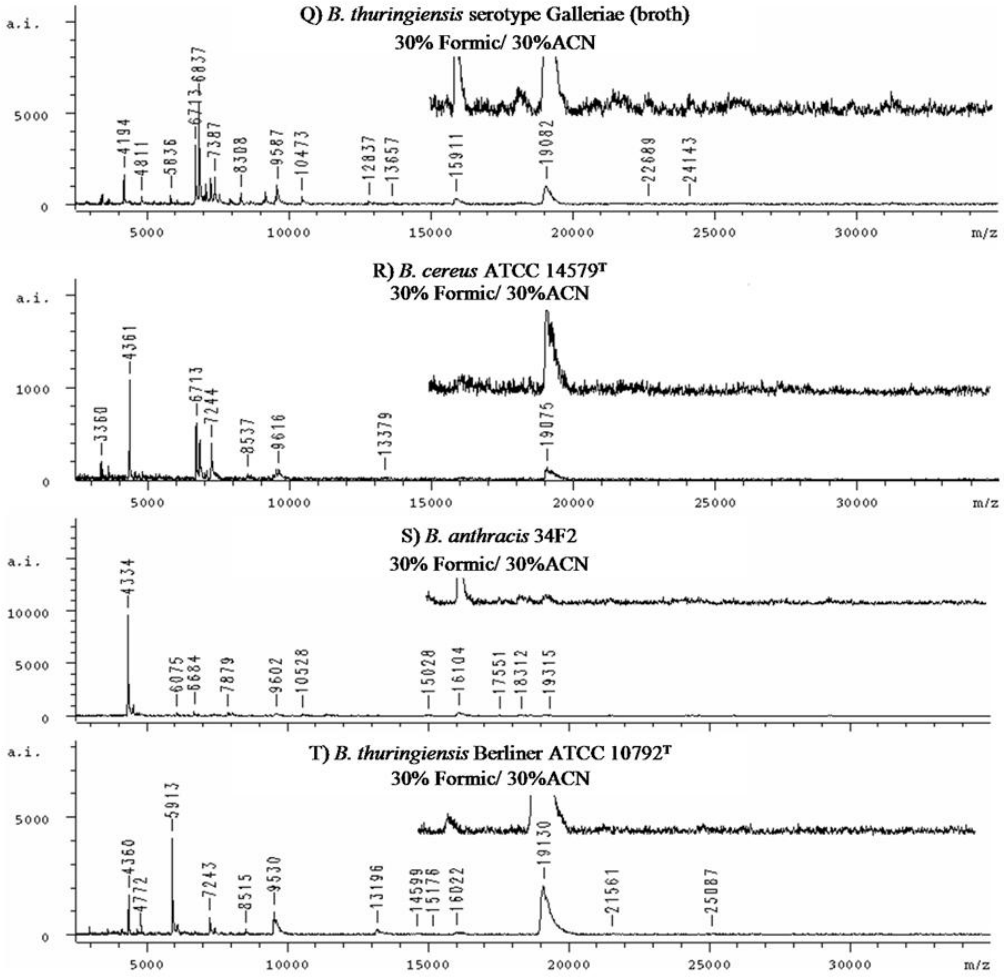

Figure 5-1. Continued.

The profiles from the BACT group of spores were remarkably different from those seen with other spore species.[101] All of the BACT spores had negligible correlation values when compared with other *Bacillus* type species (Table 5-3), confirming this low degree of spectral similarity. *B. anthracis* 34F2, FO-11, H7, H17, H2, H8, and H12 were all found to have r values from 0.71-0.97 with the *B. anthracis* 34F2 strain in the library. H3, H11, and the *B. cereus* type strain had correlation values of 0.78, 0.72, and 0.92 respectively with the *B. cereus* type strain in the library. *B. thuringiensis* Galleriae had a correlation value of 0.56 with a delta of 0.2 with the *B. cereus* type strain and therefore was considered as a possible *B. cereus* type organism. The *B. thuringiensis* type strain had an r value of 0.80 with the library reference spectra of the same strain. The remaining species H5, H10, H14, H15, and H18, *B. thuringiensis* Aizawai, Kurstaki, and Israeliensis did not have significant correlation values or delta values with any of the species in the type strain library.

Closer analysis of the relationship between the profiles is accomplished by comparing the profiles within the BACT library using linear correlation. The correlation values reveal a fairly well defined *B. anthracis* group and a poorly defined *B. cereus*/*B. thuringiensis* group (Table 5-4). This was supported by the clustering and visualization of the spectra as seen in Figure 5-2. H14 and B. thuringiensis Israeliensis are considered outliers in the cluster analysis and have r values less than 0.36 with all of the strains in the study. Within the *B. anthracis* cluster are FO-11, H8, H12, H7, H17, and H2. All of these species contain a peak at *m/z* 4, 335 and the r values for organisms within this cluster are all above 0.71. The next cluster is very diverse and contains both *B. cereus* and *B. thuringiensis* strains. The *B. thuringiensis* type strain grown in broth, and on TSA

and NA plates, formed its own resolved group within this cluster branch and has its next highest, well separated, correlation value with the *B. cereus* types strain (r = 0.44). The *B. cereus*/*B. thuringiensis* cluster can further be split into 2 subclusters, one containing H10, H18, H15, H5, and *B. thuringiensis* Aizawai and Kurstaki HD-1, of which all but the H5 and H15 strains contain a peak at 3,805 Da. H5 and H15 share a biomarker peak at *m/z* 3,751. H5 and H15 also have peaks around 10kDa which are not seen in the other strains in this subcluster. This is highlighted in the correlation values as well with H5 and H15 having r of 0.57 with each other and H10 but less than 0.28 with the *B. thuringiensis* Aizawai and Kurstaki strains. The other subcluster contains H11, H3, the *B. cereus* type strain, and *B. thuringiensis* Galleriae. These spectra all contain a peak at 4,360 Da and have r values of >0.56 with the *B. cereus* type strain. These subclusters should not be considered as true groupings. Overall, the *B. cereus*/*B. thuringiensis* cluster is poorly defined and profiles contain a great deal of variation making it very difficult to outline clear group boundaries.

Protein profiles produced after treatment with 5% TFA and 70% ACN produced more prominent peaks in the 6-9 kDa range and fewer peaks larger than 10 kDa (Figure 5-3). The peaks in this range are believed to be SASP associated proteins based on their identification in other studies.[38,99,100] The biomarkers in the 3-5 kDa range seen with the formic acid treatment are suppressed in many of the TFA treated samples. These profiles were not compared with the type strain library since it is made with 30% formic acid which produces very different spectral profiles. Comparison of the strains treated with TFA using linear correlation produced Table 5-5 and the visualization and clustering is shown in Figure 5-4. Two clusters are formed with the TFA treatment, one coherent

cluster containing the *B. thuringiensis* Aizawai and Galleriae serotypes and H10 and H15. The other cluster is very diverse and contains the remaining strains and splitting of this cluster into any subgroups was not clear-cut.

Examination of the spectral profiles, particularly of the SASP associated proteins in the *m/z* range 6,500-7,500 highlights some differences in the spectra. None of the SASP protein extracted from the other strains matches those in *B. anthracis* 34F2 which has an intense SASP at 6,684 Da. The remaining strains can be classified into two groups based on the presence of either a ~6,700 Da peak (H2, H3, H7, H8, H12, H17, H18 and the *B. thuringiensis* type strain) or a ~6,720 Da peak (*B. cereus* type strain, H5, H10, H11, H14, H15, FO-11, and *B. thuringiensis* Aizawai, Galleriae, Kurstaki, and Israeliensis). These grouping based on the SASPs are not clearly defined in the HCA.

Table 5-3. MALDI-TOFMS correlation values for BACT spore strains versus the type strain reference library using 30% formic acid as a solvent

30% Formic/ 30% ACN	34F2 *B. anthracis*	ATCC 14579 ^{T} *B. cereus*	ATCC 10792 ^{T} *B. thuringiensis*	ATCC 9372 ^{T} *B. atrophaeus*	KL-196 *B. licheniformis*	ATCC 14580 ^{T} *B. licheniformis*	ATCC 14581 ^{T} *B. megaterium*	ATCC 51516 ^{T} *B. mojavensis*	51-8C *B. niacini*	PTA-4399 ^{T} *B. odysseyi*	VSE1-06 *B. psychrodurans*	ATCC 7061 ^{T} *B. pumilus*	168 *B. subtilis*	ATCC 6051 ^{T} *B. subtilis*
B. anthracis 34F2	0.97	0.22	0.17	0.02	0.02	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01
FO11	0.79	0.31	0.10	0.02	0.05	0.05	0.02	0.03	0.03	0.01	0.04	0.01	0.00	0.01
B. cereus serotypes:														
H5	0.03	0.28	0.06	0.13	0.10	0.26	0.20	0.32	0.08	0.19	0.10	0.33	0.04	0.04
H7	0.92	0.25	0.19	0.05	0.03	0.06	0.05	0.07	0.02	0.04	0.02	0.06	0.01	0.03
H17	0.94	0.28	0.20	0.07	0.05	0.09	0.07	0.07	0.04	0.05	0.05	0.06	0.02	0.04
H2	0.71	0.27	0.21	0.02	0.05	0.07	0.03	0.07	0.04	0.03	0.06	0.02	0.01	0.01
H3	0.30	0.78	0.44	0.10	0.05	0.08	0.06	0.07	0.11	0.06	0.05	0.05	0.03	0.05
H8	0.84	0.26	0.18	0.06	0.06	0.09	0.08	0.12	0.04	0.07	0.06	0.06	0.05	0.05
H12	0.95	0.23	0.18	0.05	0.03	0.04	0.03	0.05	0.06	0.02	0.01	0.03	0.02	0.03
H10	0.02	0.09	0.00	0.03	0.02	0.04	0.01	0.03	0.02	0.02	0.02	0.01	0.02	0.02
H11	0.23	0.72	0.31	0.04	0.03	0.04	0.02	0.02	0.08	0.02	0.03	0.01	0.01	0.02
H14	0.06	0.17	0.33	0.27	0.00	0.03	0.08	0.06	0.04	0.12	0.05	0.09	0.02	0.11
H15	0.02	0.10	0.01	0.03	0.04	0.06	0.03	0.06	0.11	0.03	0.02	0.02	0.01	0.01
H18	0.01	0.05	0.01	0.06	0.03	0.04	0.03	0.02	0.04	0.03	0.03	0.01	0.03	0.02
B. cereus ATCC 14579 ^{T}	0.20	0.92	0.45	0.30	0.06	0.18	0.17	0.28	0.05	0.18	0.07	0.28	0.05	0.12
B. thuringiensis serotypes:														
Aizawai	0.02	0.07	0.02	0.06	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.01	0.02	0.02
Galleriae	0.06	0.56	0.33	0.36	0.07	0.15	0.17	0.17	0.05	0.18	0.04	0.26	0.03	0.17
Kurstaki	0.02	0.08	0.24	0.29	0.03	0.05	0.06	0.06	0.06	0.14	0.03	0.06	0.07	0.15
Israeliensis	0.07	0.22	0.10	0.09	0.02	0.05	0.03	0.04	0.06	0.04	0.04	0.05	0.01	0.04
Berliner ATCC 10792[1]	0.09	0.44	0.80	-0.01	0.10	0.09	0.03	0.05	0.04	-0.01	0.01	0.06	0.00	0.01

Table 5-4. MALDI-TOFMS correlation values for BACT spore strain library using 30% formic acid as a solvent

BACT group Spores 30% Formic/30% ACN	B. anthracis 34F2	FO11	H5	H7	H17	H2	H3	H8	H12	H10	H11	H14	H15	H18	B. cereus JCM 2152T	Aizawai	Galleriae	Kurstaki	Israeliensis	Berliner IAM 12077T
B. anthracis 34F2	-																			
FO11	0.79	-																		
B. cereus serotypes:																				
H5	0.03	0.24	-																	
H7	0.92	0.91	0.09	-																
H17	0.94	0.88	0.17	0.95	-															
H2	0.71	0.89	0.18	0.81	0.80	-														
H3	0.30	0.39	0.15	0.30	0.34	0.42	-													
H8	0.84	0.95	0.20	0.94	0.91	0.87	0.34	-												
H12	0.95	0.86	0.07	0.96	0.96	0.78	0.31	0.90	-											
H10	0.02	0.09	0.11	0.03	0.08	0.19	0.09	0.06	0.05	-										
H11	0.23	0.24	0.07	0.20	0.23	0.30	0.88	0.23	0.22	0.14	-									
H14	0.06	0.11	0.12	0.06	0.07	0.10	0.16	0.08	0.07	0.09	0.15	-								
H15	0.02	0.15	0.57	0.03	0.10	0.23	0.11	0.09	0.06	0.50	0.15	0.10	-							
H18	0.01	0.10	0.13	0.02	0.10	0.06	0.07	0.08	0.03	0.60	0.04	0.03	0.12	-						
B. cereus ATCC 14579 T	0.20	0.31	0.28	0.25	0.28	0.27	0.78	0.26	0.23	0.09	0.72	0.17	0.10	0.05	-					
B. thuringiensis serotypes:																				
Aizawai	0.02	0.09	0.12	0.03	0.10	0.11	0.07	0.07	0.04	0.76	0.08	0.05	0.28	0.93	0.07	-				
Galleriae	0.06	0.14	0.20	0.14	0.15	0.06	0.22	0.09	0.11	0.07	0.10	0.11	0.08	0.03	0.56	0.07	-			
Kurstaki	0.02	0.10	0.13	0.03	0.10	0.08	0.08	0.08	0.04	0.67	0.06	0.04	0.17	0.98	0.08	0.97	0.08	-		
Israeliensis	0.07	0.13	0.10	0.09	0.11	0.21	0.19	0.10	0.11	0.35	0.18	0.07	0.30	0.07	0.22	0.21	0.26	0.12	-	
Berliner ATCC 10792[1]	0.09	0.11	0.05	0.09	0.13	0.12	0.36	0.10	0.08	0.03	0.25	0.05	0.04	0.02	0.44	0.06	0.33	0.03	0.06	-

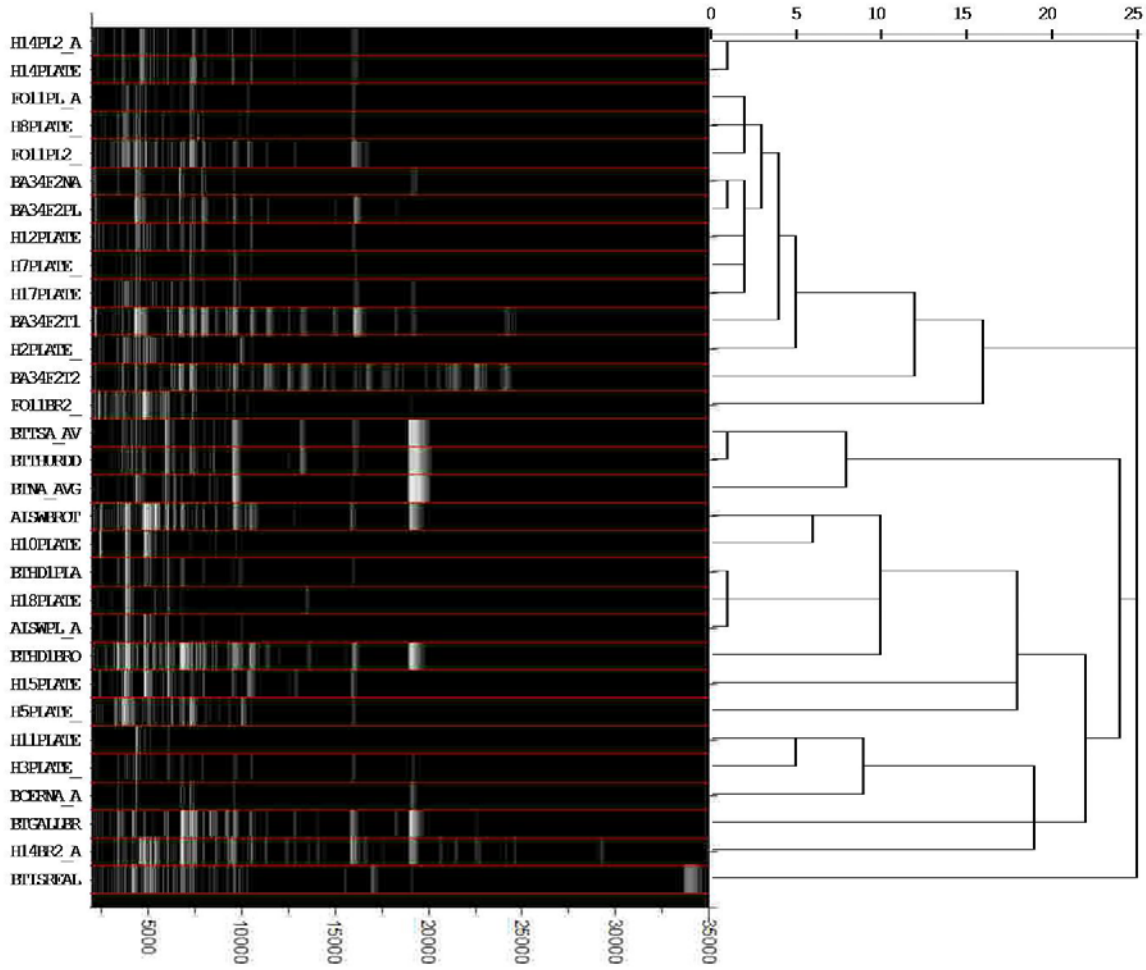

Figure 5-2. Clustering and visualization of the BACT spore protein profiles obtained using 30% formic acid as a solvent.

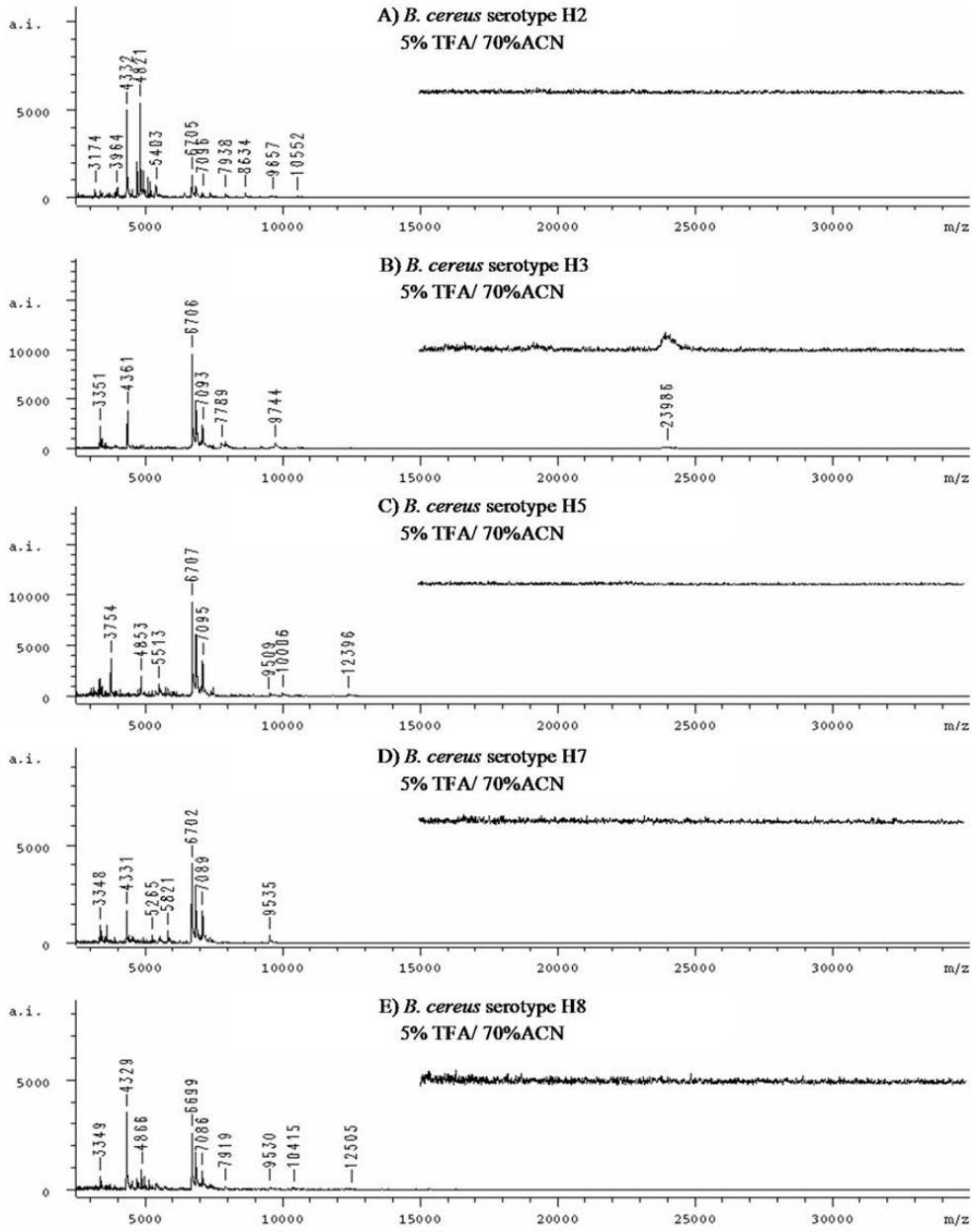

Figure 5-3. Average spectra from the BACT spores using 5% TFA as a solvent. The mass range depicted is from *m/z* 2,500-35,000. The higher molecular weight region from *m/z* 13,000- 35,000 is amplified 6x in the inset of each spectrum to help visualize the higher molecular weight proteins. A) *B. cereus* serotype H2. B) *B. cereus* serotype H3. C) *B. cereus* serotype H5. D) *B. cereus* serotype H7. E) *B. cereus* serotype H8. F) *B. cereus* serotype H10. G) *B. cereus* serotype H11. H) *B. cereus* serotype H12. I) *B. cereus* serotype H14. J) *B. cereus* serotype H15. K) *B. cereus* serotype H17. L) *B. cereus* serotype H18. M) FO-11. N) *B. thuringiensis* serotype Israeliensis. O) *B. thuringiensis* serotype Aizawai. P) *B. thuringiensis* serotype Kurstaki HD-1. Q) *B. thuringiensis* serotype Galleriae. R) *B. cereus* ATCC 14579^{T}. S) *B. anthracis* 34F2. T) *B. thuringiensis* Berliner ATCC 14579^{T}.

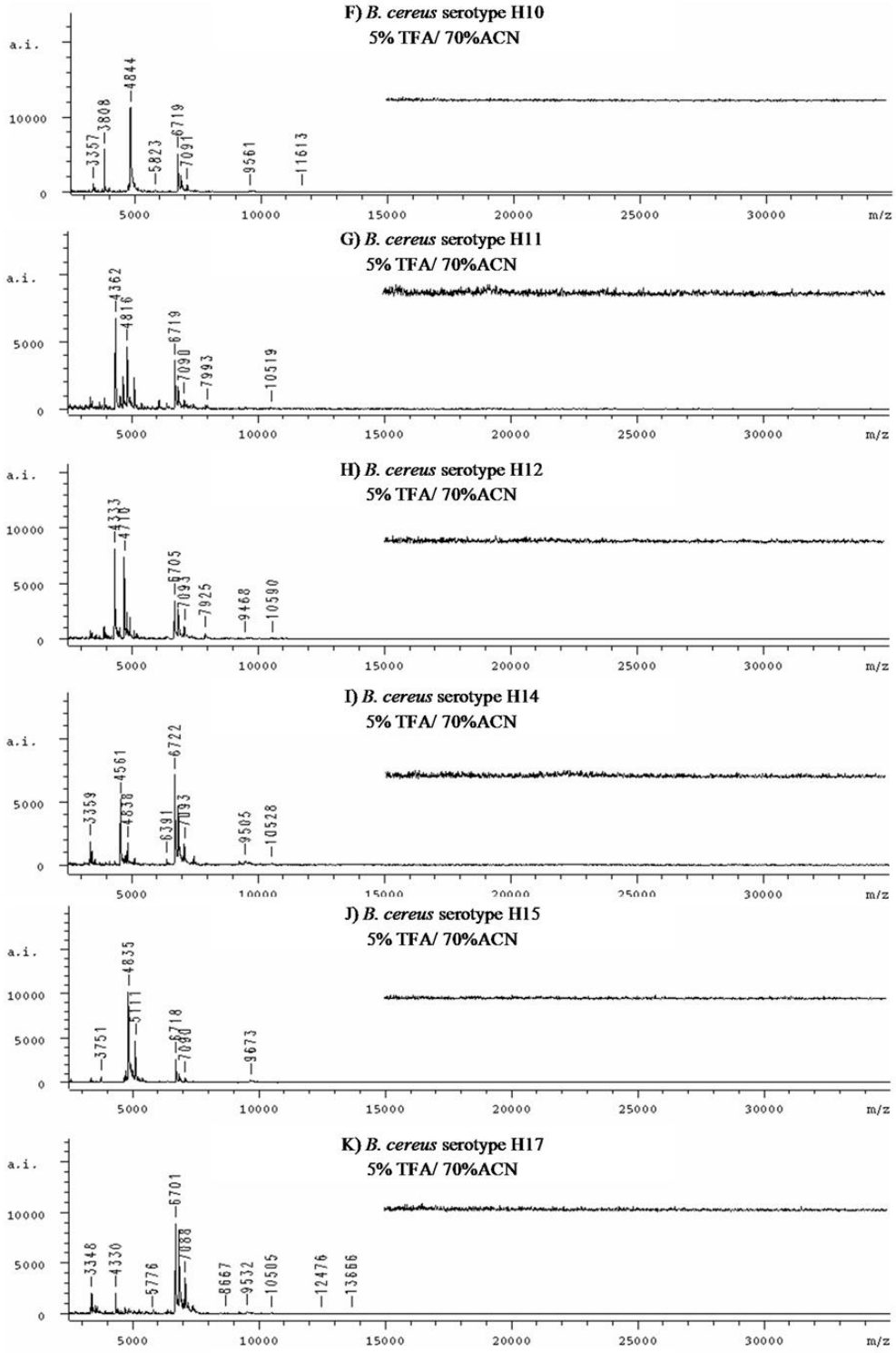

Figure 5-3. Continued.

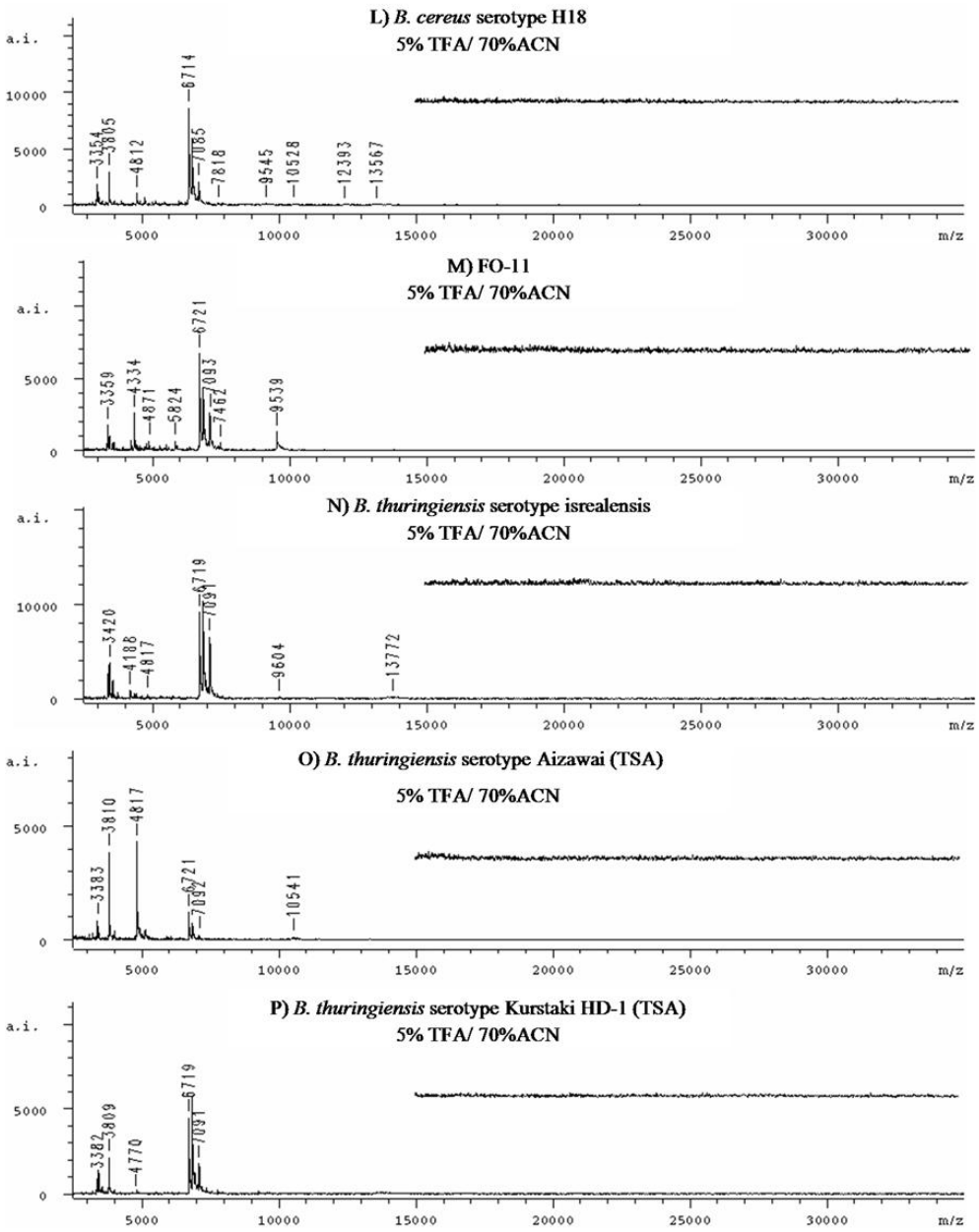

Figure 5-3. Continued.

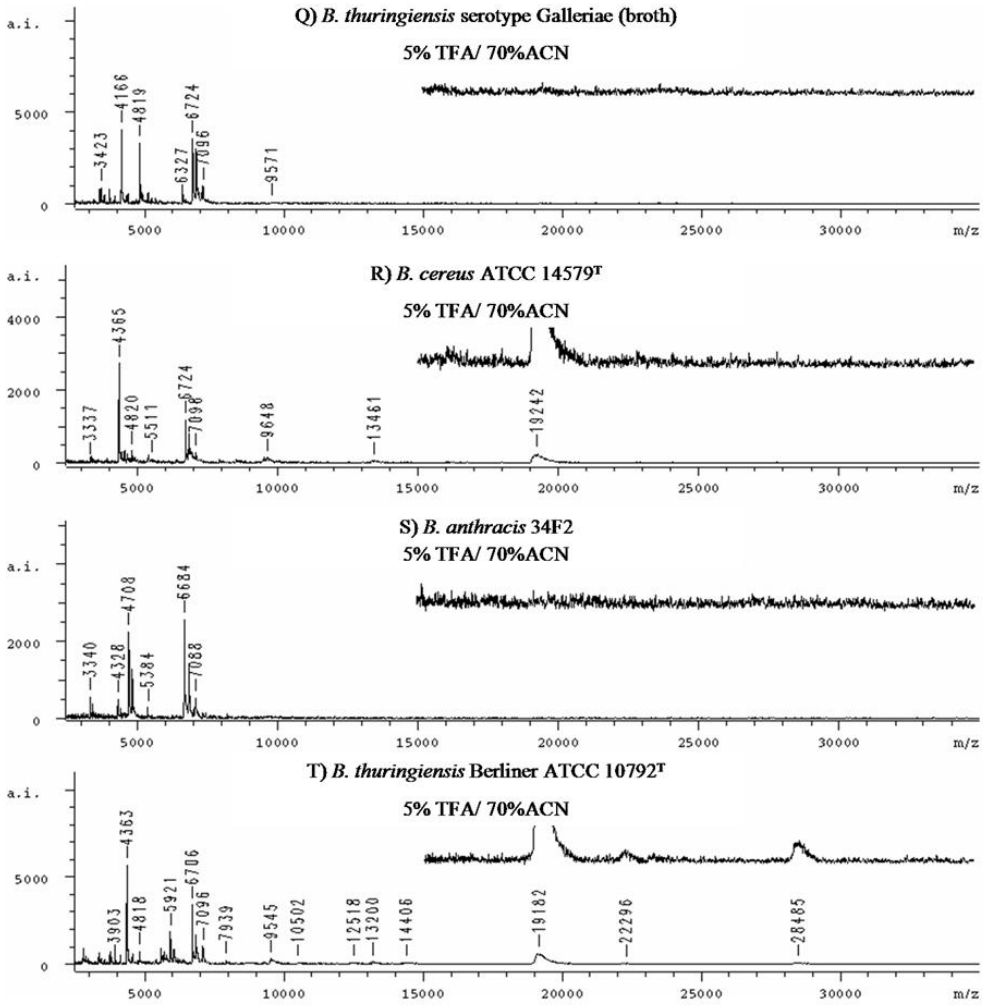

Figure 5-3. Continued.

Table 5-5. MALDI-TOFMS correlation values for BACT spore strain library using TFA as a solvent

BACT group Spores 5% TFA/70% ACN	B. anthracis 34F2	FO11	H5	H7	H17	H2	H3	H8	H12	H10	H11	H14	H15	H18	B. cereus JCM 2152T	Aizawai	Galleriae	Kurstaki	Israeliensis	Berliner IAM 12077T
B. anthracis 34F2	-																			
FO11	0.66	-																		
B. cereus serotypes:																				
H5	0.60	0.69	-																	
H7	0.76	0.72	0.81	-																
H17	0.77	0.73	0.83	0.93	-															
H2	0.41	0.38	0.29	0.46	0.33	-														
H3	0.63	0.73	0.87	0.82	0.84	0.29	-													
H8	0.79	0.69	0.68	0.90	0.83	0.60	0.70	-												
H12	0.57	0.54	0.41	0.62	0.50	0.77	0.46	0.75	-											
H10	0.21	0.37	0.32	0.22	0.23	0.49	0.25	0.23	0.25	-										
H11	0.32	0.46	0.35	0.33	0.31	0.51	0.50	0.36	0.37	0.58	-									
H14	0.54	0.85	0.66	0.61	0.65	0.23	0.68	0.51	0.35	0.42	0.51	-								
H15	0.15	0.25	0.24	0.16	0.15	0.50	0.17	0.18	0.25	0.82	0.62	0.35	-							
H18	0.60	0.81	0.79	0.70	0.77	0.25	0.82	0.60	0.37	0.41	0.41	0.78	0.26	-						
B. cereus ATCC 14579^{T}	0.37	0.51	0.38	0.38	0.38	0.32	0.56	0.39	0.32	0.27	0.76	0.50	0.22	0.40	-					
B. thuringiensis serotypes:																				
Aizawai	0.17	0.28	0.25	0.19	0.19	0.49	0.20	0.19	0.22	0.83	0.48	0.30	0.57	0.39	0.24	-				
Galleriae	0.44	0.66	0.53	0.47	0.51	0.47	0.52	0.43	0.35	0.57	0.58	0.64	0.49	0.57	0.48	0.52	-			
Kurstaki	0.57	0.84	0.68	0.64	0.75	0.17	0.69	0.52	0.31	0.39	0.37	0.81	0.19	0.86	0.42	0.40	0.60	-		
Israeliensis	0.61	0.88	0.72	0.69	0.80	0.20	0.72	0.57	0.35	0.30	0.38	0.83	0.19	0.83	0.45	0.24	0.64	0.95	-	
Berliner ATCC 10792^{T}	0.29	0.32	0.42	0.37	0.34	0.47	0.42	0.37	0.31	0.43	0.61	0.32	0.41	0.35	0.49	0.45	0.48	0.27	0.29	-

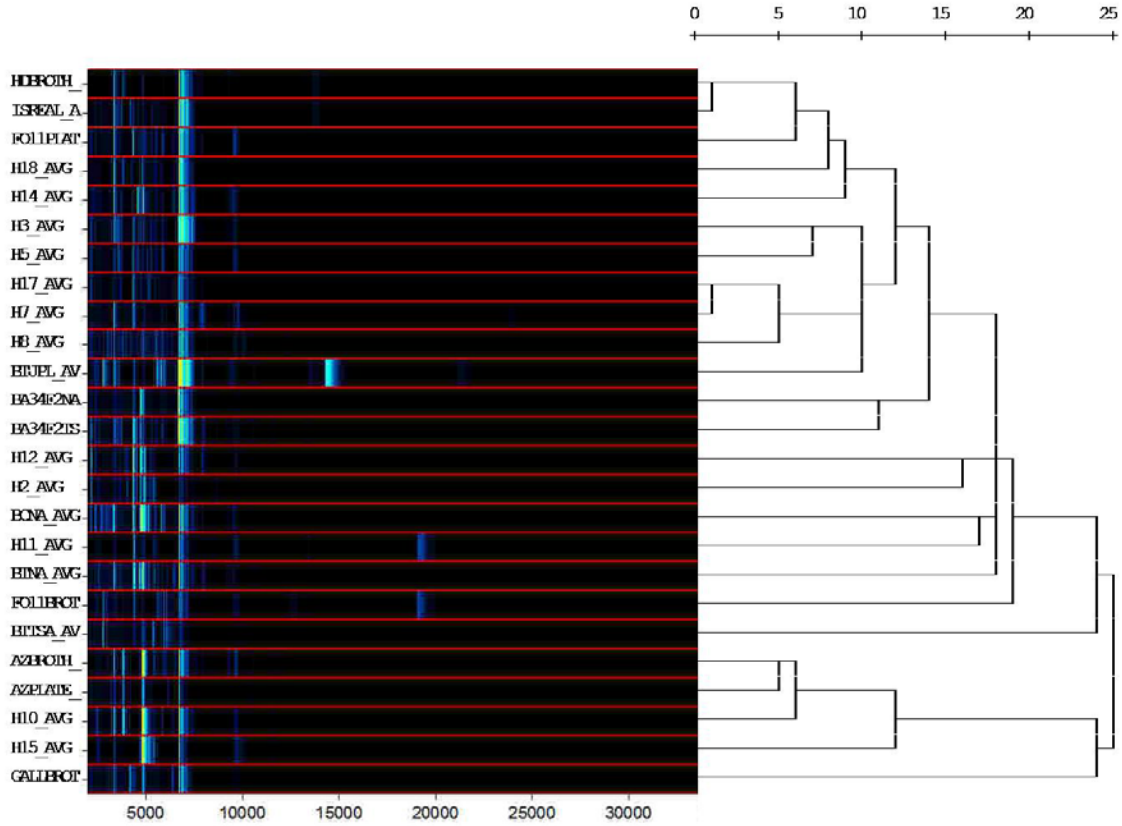

Figure 5-4. Clustering and visualization of the BACT spores protein profiles obtained from using TFA as a solvent.

MALDI-TOFMS Protein Profiling of BACT Vegetative Cells

Vegetative cells were extracted with 30% formic acid/30% ACN and were subjected to MALDI-TOFMS protein profiling. Spectra obtained for the vegetative cells were significantly different from their sporulated counterparts. Vegetative cell spectra are characterized by a larger number of biomarkers as well as the presence of higher molecular weight protein peaks (Figure 5-5). All of the vegetative strains in this study were found to have a biomarker peak at 6,425 Da, a triplet of peaks centered at 9,600 Da, and a peak at ~19.1 kDa.

The results for the BACT vegetative cells with the type strain library for vegetative cells using linear correlation analysis are shown in Table 5-6. Highlighted in black are those values that make both the r>0.75 and >0.1 delta values and highlighted in gray are those strains where the r>0.50 and the delta value is >0.1. H15, H18, the *B. thuringiensis* type strain and the Kurstaki serotype, and *B. mycoides* do not have correlation values with any of the reference strains that meet either criterion. Of the strains that meet the criteria, all of them except the *B. cereus* type strain and *B. thuringiensis* Kurstaki had the closest match being the *B. anthracis* type strain. The results of the correlation analysis of these strains with themselves are in Table 5-7 and the clustering and visualization is shown in Figure 5-6. The cluster analysis does little to differentiate this group into distinct clusters and does not mimic the correlation analysis very effectively. This is likely due to the greater complexity of peaks in the vegetative spectra.

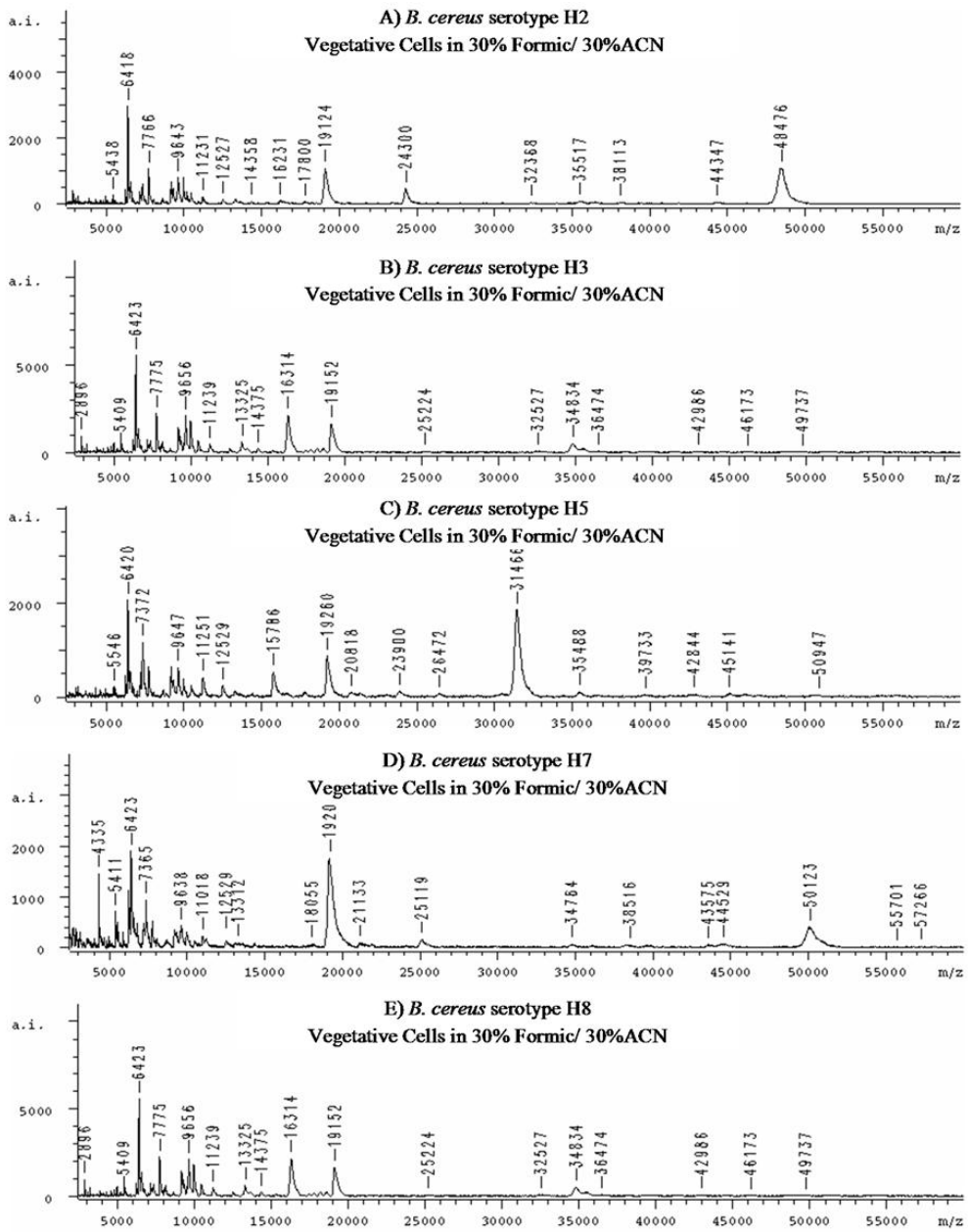

Figure 5-5. Average spectra from the BACT vegetative cells using 30% formic acid as a solvent. The mass range depicted is from *m/z* 2,500-60,000. A) *B. cereus* serotype H2. B) *B. cereus* serotype H3. C) *B. cereus* serotype H5. D) *B. cereus* serotype H7. E) *B. cereus* serotype H8. F) *B. cereus* serotype H10. G) *B. cereus* serotype H11. H) *B. cereus* serotype H14. I) *B. cereus* serotype H15. J) *B. cereus* serotype H17. K) *B. cereus* serotype H18. L) FO-11. M) *B. thuringiensis* serotype Israeliensis. N) *B. thuringiensis* serotype Aizawai. O) *B. thuringiensis* serotype Kurstaki HD-1. P) *B. cereus* ATCC 14579^{T}. Q) *B. anthracis* 34F2. R) *B. thuringiensis* Berliner ATCC 14579^{T}. S) *B. mycoides* ATCC 6462^{T}. T) *B. cereus* serotype H16. U) *B. cereus* serotype H6. V) *B. cereus* serotype H9.

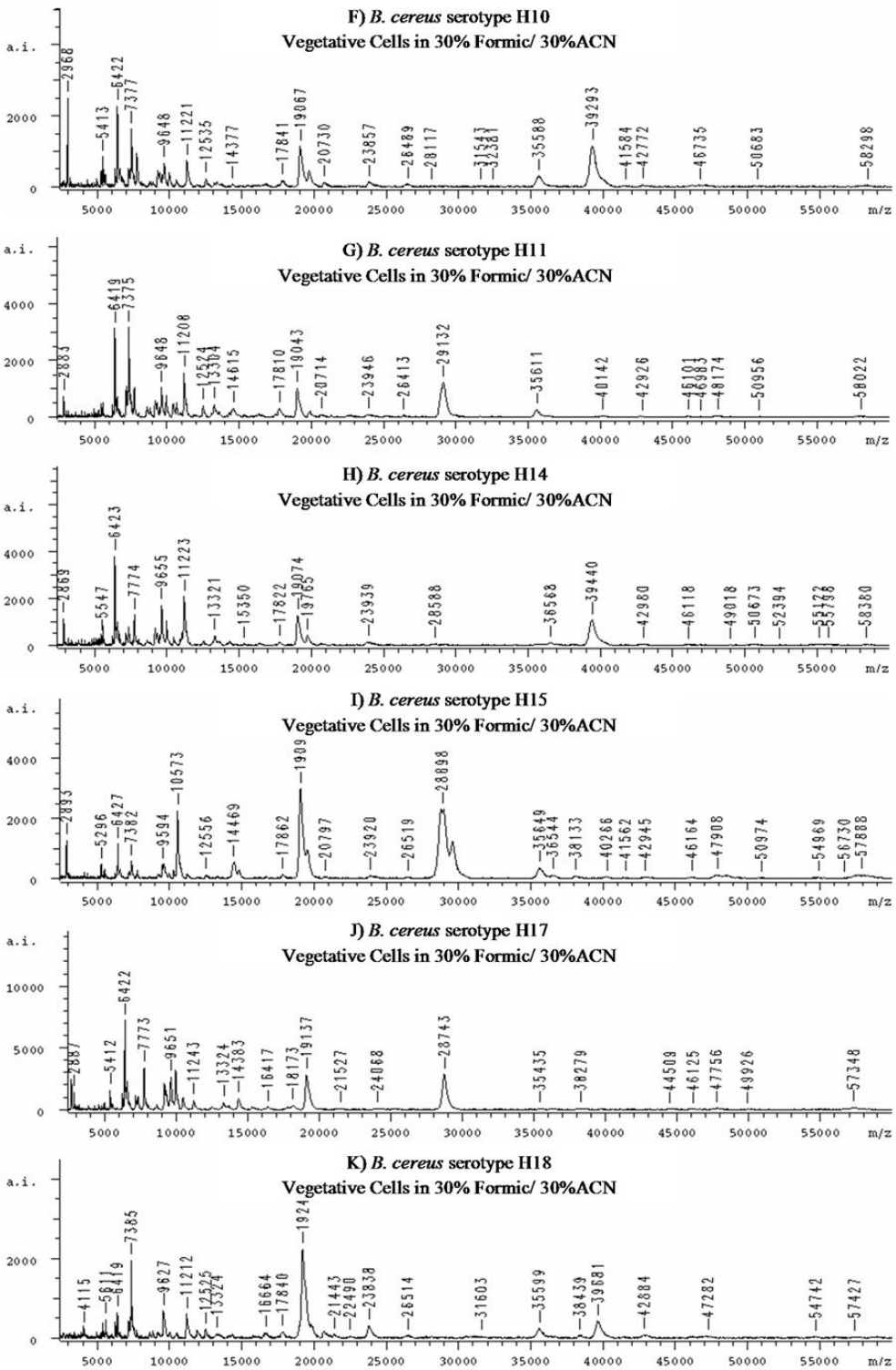

Figure 5-5. Continued.

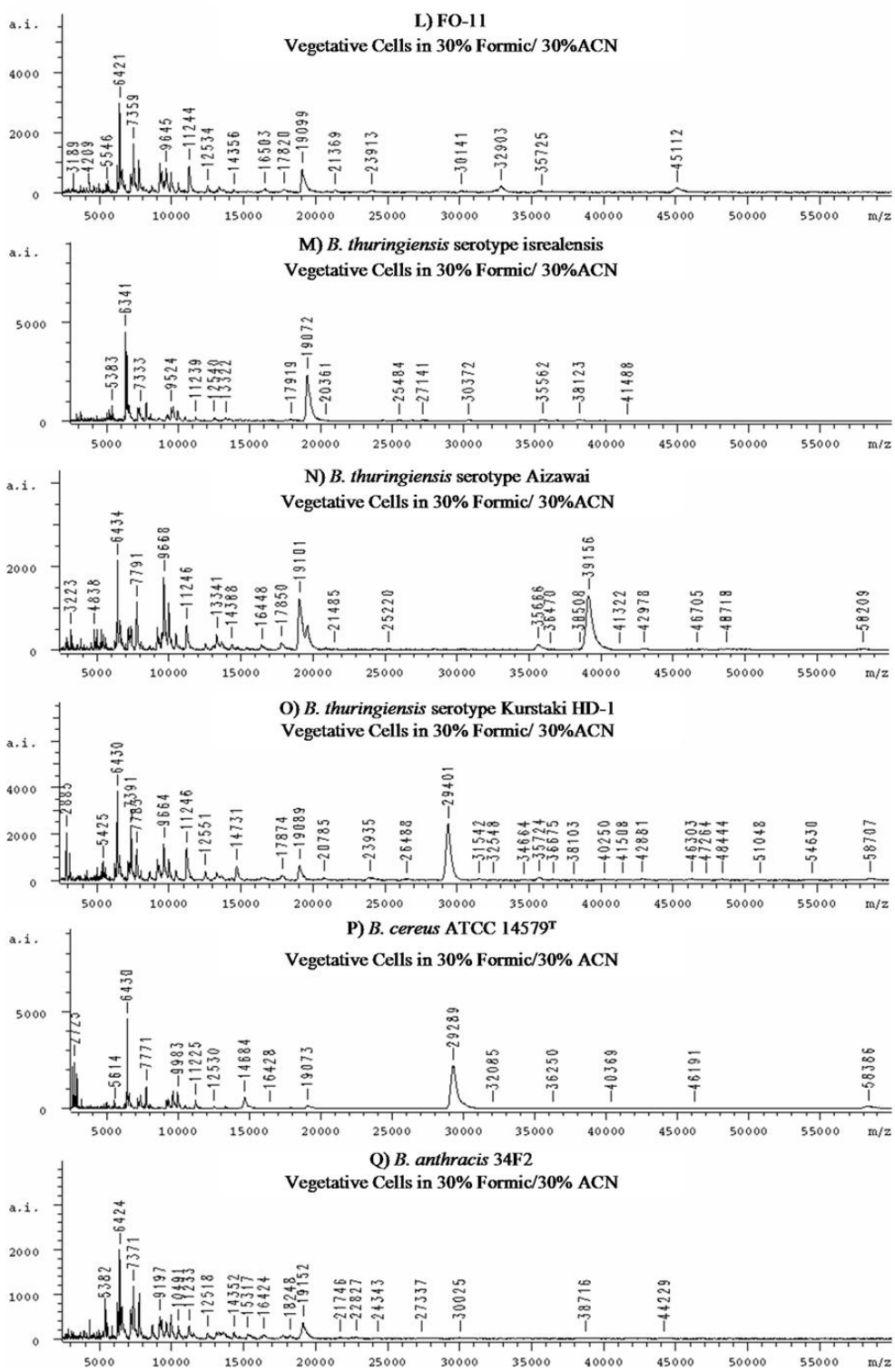

Figure 5-5. Continued.

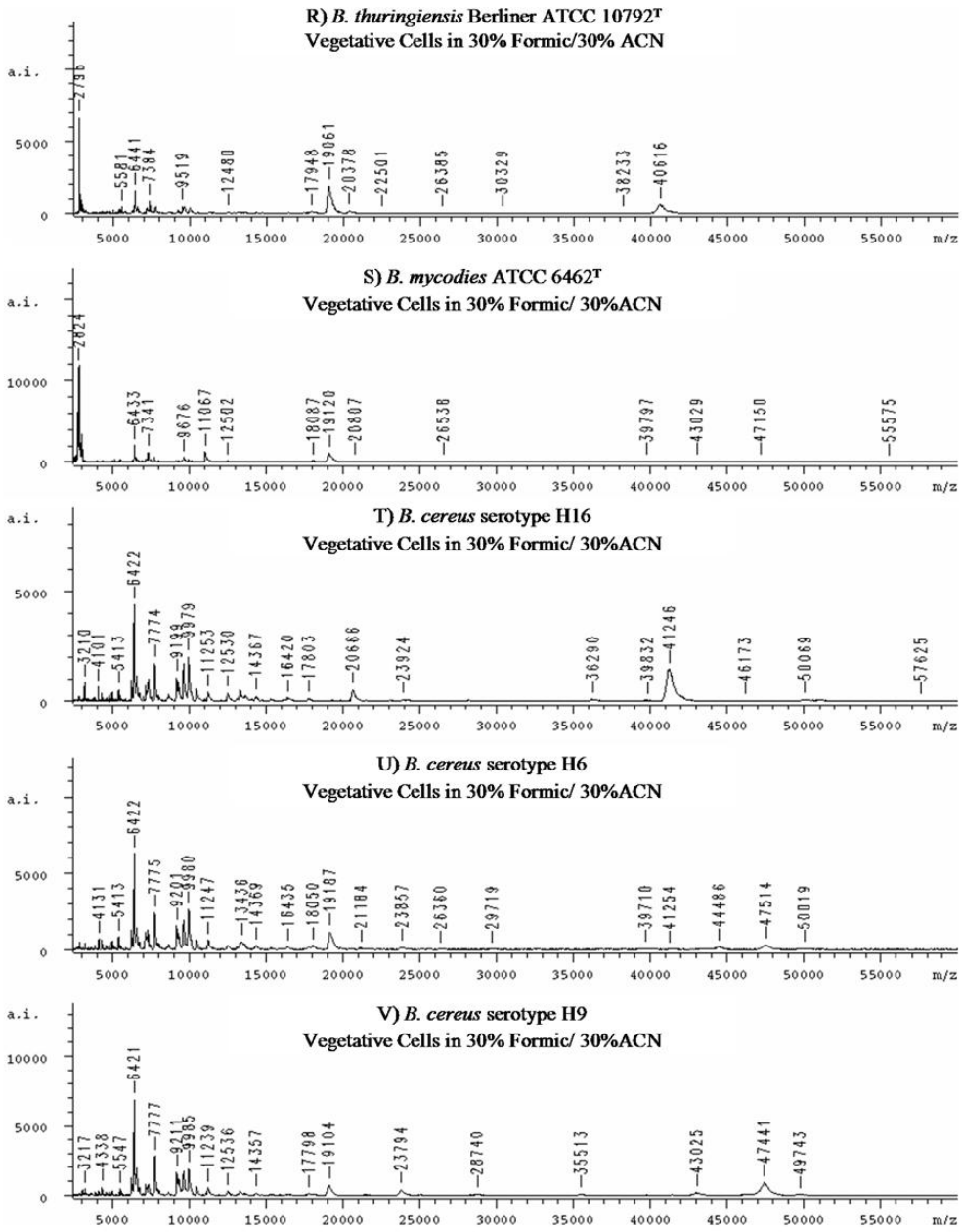

Figure 5-5. Continued.

Table 5-6. MALDI-TOFMS correlation values for BACT vegetative cells versus the vegetative type strain reference library using30% formic acid as a solvent

30% Formic/ 30% ACN Vegetative Cells	**34F2** ***B. anthracis***	**ATCC 14579 T** ***B. cereus***	**ATCC 10792 T** ***B. thuringiensis***	**ATCC 9372 T** ***B. atrophaeus***	**KL-196** ***B. licheniformis***	**ATCC 14580 T** ***B. licheniformis***	**ATCC 14581 T** ***B. megaterium***	**ATCC 51516 T** ***B. mojavensis***	**51-8C** ***B. niacini***	**PTA-4399 T** ***B. odysseyi***	**VSE1-06** ***B. psychrodurans***	**ATCC 7061 T** ***B. pumilus***	**168** ***B. subtilis***	**ATCC 6051 T** ***B. subtilis***
B. anthracis 34F2	0.83	0.14	0.21	0.21	0.10	0.18	0.43	0.18	0.13	0.36	0.38	0.38	0.20	0.18
FO-11	0.76	0.13	0.33	0.16	0.13	0.20	0.36	0.14	0.13	0.32	0.33	0.35	0.17	0.18
B. cereus serotypes:														
H5	0.47	0.07	0.16	0.09	0.05	0.09	0.22	0.06	0.05	0.19	0.20	0.23	0.08	0.08
H6	0.76	0.15	0.26	0.16	0.10	0.16	0.30	0.16	0.13	0.22	0.31	0.18	0.17	0.18
H7	0.68	0.13	0.46	0.11	0.04	0.07	0.19	0.07	0.06	0.17	0.20	0.18	0.10	0.08
H17	0.73	0.28	0.32	0.11	0.06	0.10	0.16	0.10	0.08	0.15	0.21	0.12	0.10	0.12
H2	0.63	0.10	0.39	0.09	0.02	0.07	0.17	0.06	0.04	0.14	0.22	0.13	0.07	0.08
H3	0.63	0.11	0.30	0.09	0.06	0.11	0.21	0.09	0.15	0.16	0.26	0.12	0.13	0.20
H9	0.76	0.15	0.22	0.13	0.07	0.13	0.20	0.10	0.08	0.17	0.24	0.14	0.11	0.13
H16	0.62	0.12	0.22	0.14	0.08	0.13	0.29	0.14	0.10	0.23	0.29	0.20	0.14	0.15
H10	0.63	0.10	0.36	0.12	0.09	0.14	0.29	0.09	0.05	0.26	0.26	0.29	0.11	0.08
H11	0.65	0.42	0.24	0.16	0.21	0.26	0.42	0.15	0.11	0.41	0.34	0.45	0.19	0.15
H14	0.58	0.13	0.33	0.08	0.25	0.24	0.22	0.10	0.09	0.18	0.23	0.15	0.11	0.12
H15	0.27	0.33	0.39	0.00	-0.03	-0.02	0.03	0.02	-0.01	0.09	0.05	0.10	0.00	-0.01
H18	0.47	0.03	0.39	0.10	0.08	0.12	0.28	0.07	0.04	0.26	0.25	0.30	0.11	0.06
B. cereus ATCC 14579 T	0.36	0.79	0.11	0.07	0.02	0.05	0.11	0.08	0.02	0.12	0.15	0.10	0.04	0.04
B. thuringiensis serotypes:														
Aizawai	0.53	0.08	0.37	0.09	0.07	0.10	0.19	0.07	0.15	0.17	0.25	0.13	0.08	0.15
Galleriae	0.45	0.05	0.50	0.06	0.00	0.04	0.11	0.04	0.11	0.15	0.14	0.12	0.06	0.07
Kurstaki	0.57	0.43	0.20	0.12	0.12	0.16	0.30	0.12	0.29	0.28	0.28	0.26	0.12	0.16
Israeliensis	0.56	0.07	0.54	0.09	0.02	0.05	0.16	0.04	0.13	0.17	0.17	0.09	0.05	0.09
Berliner ATCC 10792T	0.39	0.10	0.82	0.05	0.00	0.03	0.11	0.02	0.12	0.14	0.13	0.11	0.03	0.05
B. mycoides ATCC 6462T	0.19	0.16	0.22	0.03	0.04	0.03	0.06	0.01	0.05	0.06	0.05	0.04	0.01	0.02

Table 5-7. MALDI-TOFMS correlation values for BACT vegetative cells using 30% formic acid as a solvent

Vegetative Cells in 30%Formic/30%ACN	B. anthracis 34F2	FO-11	B. cereus serotypes: H5	H6	H7	H17	H2	H3	H9	H16	H10	H11	H14	H15	H18	B. cereus ATCC 14579 T	B. thuringiensis serotypes: Aizawai	Galleriae	Kurstaki	Israeliensis	Berliner ATCC 10792T	B. mycoides ATCC 6462T
B. anthracis 34F2	-																					
FO-11	0.87	-																				
B. cereus serotypes:																						
H5	0.52	0.57	-																			
H6	0.84	0.85	0.51	-																		
H7	0.67	0.73	0.48	0.67	-																	
H17	0.71	0.73	0.42	0.84	0.65	-																
H2	0.65	0.75	0.44	0.75	0.70	0.69	-															
H3	0.70	0.78	0.46	0.83	0.64	0.73	0.72	-														
H9	0.81	0.81	0.48	0.93	0.60	0.79	0.71	0.77	-													
H16	0.74	0.71	0.42	0.83	0.43	0.66	0.57	0.67	0.79	-												
H10	0.68	0.73	0.43	0.61	0.64	0.56	0.59	0.56	0.59	0.48	-											
H11	0.69	0.76	0.45	0.60	0.51	0.58	0.53	0.52	0.56	0.52	0.65	-										
H14	0.65	0.78	0.43	0.73	0.59	0.65	0.65	0.68	0.70	0.60	0.78	0.66	-									
H15	0.22	0.34	0.18	0.26	0.48	0.54	0.41	0.31	0.23	0.11	0.36	0.48	0.34	-								
H18	0.45	0.57	0.47	0.41	0.74	0.39	0.51	0.43	0.30	0.22	0.59	0.53	0.49	0.43	-							
B. cereus ATCC 14579 T	0.44	0.42	0.24	0.44	0.30	0.43	0.33	0.37	0.46	0.40	0.31	0.58	0.40	0.37	0.14	-						
B. thuringiensis serotypes:																						
Aizawai	0.61	0.68	0.39	0.70	0.61	0.62	0.63	0.68	0.66	0.54	0.80	0.50	0.82	0.37	0.48	0.37	-					
Galleriae	0.45	0.52	0.29	0.42	0.66	0.47	0.53	0.45	0.39	0.25	0.52	0.39	0.46	0.46	0.52	0.20	0.52	-				
Kurstaki	0.68	0.69	0.41	0.64	0.49	0.57	0.53	0.56	0.62	0.57	0.56	0.76	0.63	0.38	0.37	0.79	0.56	0.36	-			
Israeliensis	0.51	0.60	0.34	0.56	0.73	0.58	0.63	0.56	0.54	0.35	0.57	0.43	0.56	0.50	0.52	0.26	0.62	0.69	0.42	-		
Berliner ATCC 10792T	0.45	0.54	0.32	0.39	0.59	0.42	0.50	0.41	0.33	0.25	0.56	0.50	0.45	0.44	0.59	0.22	0.48	0.58	0.40	0.60	-	
B. mycoides ATCC 6462T	0.10	0.13	0.07	0.13	0.20	0.17	0.12	0.10	0.11	0.09	0.13	0.14	0.18	0.09	0.08	0.16	0.11	0.13	0.14	0.12	0.45	-

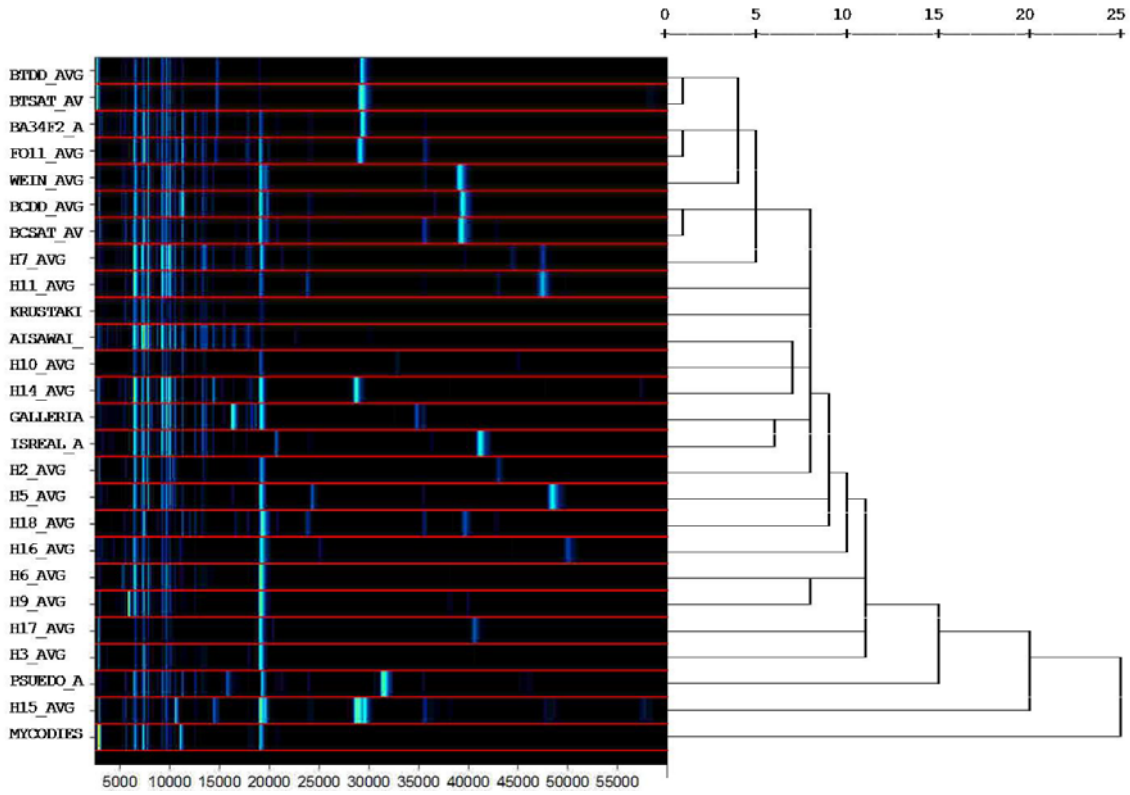

Figure 5-6. Clustering and visualization of the BACT vegetative cell protein profiles obtained from using formic acid as a solvent.

Discussion

All of the BACT strains examined using 16S rDNA analysis had greater than 97.5% sequence similarity with both of the type strains, *B. cereus* and *B. anthracis* 34F2 (Table 5-1). Therefore, 16S rDNA did not allow for the differentiation of these isolates. The DNA:DNA hybridization data is reproduced in part for the strains analyzed by MALDI-TOFMS in Table 5-8. Clustering based on DNA:DNA hybridization resulted in

Table 5-8. DNA:DNA Hybridization values of the BACT strains examined in this study. Adapted from LaDuc, *et al.* 2004 and M. Satomi personal communication

			B. cereus serotypes																B. thuringiensis serotypes					
	B. anthracis Sterne 7702	FO-11	H5	H6	H7	H17	H2	H3	H8	H9	H12	H16	H10	H11	H14	H15	H18	B. cereus JCM 2152T	Aizawai	Galleriae	Kurstaki	Israeliensis	Berliner IAM 12077T	B. mycoides ATCC 6462
B. anthracis Sterne 7702	-																							
FO-11	71																							
B. cereus serotypes:		-																						
H5	60	na	-																					
H6	78	na	97	-																				
H7	75	na	82	81	-																			
H17	75	na	86	82	82	-																		
H2	67	na	72	74	70	66	-																	
H3	69	na	65	69	61	60	70	-																
H8	71	na	66	69	66	62	74	68	-															
H9	66	na	72	69	76	57	77	67	83	-														
H12	76	na	71	72	66	70	78	80	77	80	-													
H16	84	na	79	75	77	80	58	63	62	69	67	-												
H10	50	na	54	59	55	63	50	47	54	58	51	61	-											
H11	48	na	43	51	42	46	40	47	54	48	35	54	80	-										
H14	56	na	58	56	63	44	54	49	46	56	54	59	76	78	-									
H15	56	na	56	47	60	66	55	49	53	53	57	57	78	83	84	-								
H18	40	na	45	42	41	41	41	42	37	45	35	44	59	66	60	63	-							
B. cereus JCM 2152T	56	53	55	53	53	56	59	49	52	55	49	52	72	77	88	81	58	-						
B. thuringiensis serotypes:																								
Aizawai	51	na	57	60	60	58	57	54	51	59	60	55	82	75	72	82	62	73	-					
Galleriae	51	na	58	48	56	61	58	44	55	58	66	53	80	74	63	67	56	69	69	-				
Kurstaki	47	na	54	54	52	54	55	46	49	57	50	55	64	60	63	71	73	58	64	58	-			
Israeliensis	37	na	48	49	51	47	52	38	36	44	48	49	47	39	56	58	40	55	50	64	52	-		
Berliner IAM 12077T	45	58	51	40	37	52	35	46	44	40	40	43	63	52	56	46	51	54	56	57	52	43	-	
B. mycoides ATCC 6462T	48	55	53	40	56	52	47	46	55	48	49	49	50	45	47	40	43	52	54	58	43	41	51	-

the organization of these strains into 4 groups. The *B. anthracis* group consisted of the *B. anthracis* strain and the *B. cereus* serotypes H02, H09, H08, H01, H03, H12, H05, H06, H07, H16, and H17 and the laboratory strain FO-11. The second group consisted of the *B. cereus* type strain, the serotypes H14, H11, H04, H15, H10, H18, and H13 and the *B. thuringiensis* serotypes Kurstaki HD-1, Galleriae, and Aizawai. The *B. thuringiensis*

serotypes Kurstaki HD-1, Galleriae, and Aizawai had DNA reassociation values of 58%, 69%, and 73% with the *B. cereus* type strain. The third grouping consisted of the *B. thuringiensis* serotype Israeliensis and the type strain, Berliner which had less than 60% hybridization with all of the other strains. The sole member of the fourth group was *B. mycoides* (personal communication, M. Satomi and K. Venkateswaran).[88] Since DNA:DNA hybridization is considered the gold-standard for bacterial speciation it is used as the benchmark for comparison of the analyses performed in this study.

FAME analysis was only able, at a high confidence level, to assign 4 of the strains in this study to a reference strain. Based on the DNA:DNA hybridization studies the assignment of the 3 of strains (*B. anthracis* 34F2, *B. thuringiensis* Kurstaki, and the *B. cereus* type strain) is correct, however H7 is misidentified as *B. thuringiensis*. The FAME analysis was not useful for identifying the other 6 species analyzed as they had high similarity indices with 2 or more of the reference strains. The cluster analysis of this data also resulted in several misclassifications, including H16, H3, H7 to the *B. cereus*/*B. thuringiensis* cluster and H18 to the *B. anthracis* cluster.

MALDI-TOFMS protein profiling with a 30% formic acid/30% ACN solvent split the spores into several groupings which somewhat mimicked the DNA-based molecular characterization of these strains. The *B. anthracis* group is well defined and contains H02, H07, H08, H12, H17, FO-11, and the *B. anthracis* 34F2 strain which all agree with the hybridization assignment. The other groupings based on protein profiling are not as clearly defined as the *B. anthracis* group and are characterized by having strains within them that are well below the 0.75 cutoff typically used in our analyses. The second grouping contains the *B. cereus* type strain, H3, H10, H11, H18, H15, H5, and *B.*

thuringiensis Aizawai and Kurstaki HD-1. H18, Aizawai , Galleriae , and Kurstaki HD-1 have r > 0.93 with each other; H11, H3 have correlation values >0.70 with the *B. cereus* type strain while the Galleriae strain has a lower r value of 0.56. H5 and H15 have r of 0.57 with each other but overall very low correlation values with the other strains. The third group contains what can be considered outliers as the *B. thuringiensis* type strain forms its own well defined cluster and *B. thuringiensis* Israeliensis and H14 do not fall within any of the clusters in this study. This is expected for the *B. thuringiensis* type strain and Israeliensis as hybridization studies showed less than 60% hybridization with the other BACT strains.

H3 is identified as a *B. cereus* type organism based on the linear correlation analysis (r=0.78) and according to the hybridization data it should fall out closer to the *B. anthracis* group (69% hybridization with *B. anthracis* 7702). Galleriae, H5, and H15 are in the same misclassification predicament; however, the linear correlation values for these strains are below 0.56, 0.28, and 0.10 respectively with the *B. cereus* type species and most of the other strains in this cluster. So their inclusion in the cluster would need to be justified by the presence of additional species-specific biomarker peaks, and neither have the 3,805 Da or 4,360 Da peaks that are characteristic in this cluster. H10 and H18 also have low correlation values with the *B. cereus* type strain but have r > 0.76 with Aizawai and contain the biomarker peak at 3,805 Da. Therefore it is proposed that Galleriae, H5, H15, and H14 are considered outliers as there is not sufficient evidence to include them in either cluster. The inclusion of the H10 and H18 are justified by additional biomarker evidence.

Because of the misclassification of H3 and the lack of classifications for H5, H15, and H14, additional MALDI-TOFMS profiles were collected. These profiles were collected on spores that were treated with 5%TFA/70% ACN in lieu of the formic acid solvent to determine if better classification could be obtained using a different solvent system. Using the TFA treatment, even less classification was obtained than with the formic acid system as evidenced by the poor HCA clustering and the higher correlation values that were found between most of the strains. This can be attributed to the similarities in the SASP proteins that are released using the TFA treatment. Interestingly, the *B. anthracis* strain had a unique set of SASPs while the rest of the strains were divided into two SASP groups, one containing a 6,700 Da SASP (*B. thuringiensis* type strain group) and the other a 6,720 Da SASP (*B. cereus* type strain group). Whereas in the formic acid study, there was ambiguity between the *B. thuringiensis* and *B. cereus* group, in the TFA analysis, strains classified by hybridization as *B. anthracis* and *B. thuringiensis* showed more overlap and higher correlation values than with the *B. cereus* strains.

To further characterize the BACT strains in this study, MALDI-TOFMS protein profiling was performed on the vegetative cells using formic acid as a solvent. The vegetative cells proved to be the least discriminating of the MALDI-TOFMS studies. Several group specific biomarkers were identified at 6,425 Da and 19.1 kDa, and a triplet of peaks was found in all the strains around 9,600 Da. Most of the variation in the vegetative profiles was found above 20 kDa. In looking strictly at the classification of these strains as *B. anthracis*, *B. cereus*, or *B. thuringiensis* in the type strain library, 14 of 26 strains met one of the two criteria to be included as *B. anthracis*. These included

several of the *B. thuringiensis* (Aizawai and Kurstaki) and *B. cereus* serotypes (H10, H11, and H14) that were not classified as *B. anthracis* based on the hybridization study.

The proper "classifications" of the strains described thus far have been based on DNA hybridization analysis. Although this is the gold standard for species differentiation, it does not effectively identify differences in the phenotypes or toxins that may exist in these isolates which are plasmid encoded. This direction leaves the classification of these organisms based on evolutionary relationships and goes for the more applied identification of these strains. In looking at these properties, *B. cereus* H3 (reassociation values of 49% and 69% with *B. cereus* and *B. anthracis* respectively) was the only serotype which produced the cereuride toxin characteristic of *B. cereus*. *B. cereus* H17 (reassociation values of 56% and 75% with *B. cereus* and *B. anthracis* respectively) and H6 (reassociation values of 53% and 78% with *B. cereus* and *B. anthracis* respectively) tested positive for the genes encoding the protective antigen (*pag*) from *B. anthracis*.[88] H17 was classified as *B. anthracis* and H3 was classified as *B. cereus* using the formic acid treatment and linear correlation analysis. We were unable to obtain spores for H6 but the vegetative cells had one of the highest correlation values with the *B. anthracis* reference strain ($r = 0.84$). Based on toxigenic properties, the identification of these strains would be correct; however the converse would also be true. Strains that lack the toxins should also be distinguishable, which with the current analysis they clearly are not.

Conclusion

The MALDI-TOFMS-based protein profiling method using formic acid as a solvent proved to be the most successful for the differentiation and classification of the BACT group spores in this study. MALDI-TOFMS protein profiles using TFA were less

discriminatory than those with formic acid solvent, instead causing an improper grouping of the *B. anthracis* strains with *B. thuringiensis*. Analysis of vegetative cells was the least discriminatory and most of the strains in the study would be assigned to the *B. anthracis* group. MALDI-TOFMS protein profiling also proved to be more discriminating than both FAME and 16S rDNA analysis. This is the first investigation of this many BACT group bacteria with MALDI-TOFMS protein profiling where the results are compared directly with genetic analyses.

Using formic acid as a solvent, only 1 "misclassification" was made according to DNA hybridization results (H3) and this organism could arguable be identified properly based on pathogenic properties. There were several species that could not with certainty be assigned to any of the species in the library. These missed assignments, or false negatives, may highlight the fact that the small number of type strains or reference strains that are used are not representative of the microbial diversity within this group.[2] Additionally, to improve the analysis of this group of spores it would be beneficial to test to determine if better biomarker extraction could be facilitated by higher formic acid concentrations. Initial optimization experiments were carried out on *B. subtilis* 168, which does not contain an exosporium, and is less hydrophobic than the BACT group bacteria.[2] The increased hydrophobicity and exosporium layer may make the protein extraction step more challenging and a different solvent system may prove better for this group of spores. It may also be feasible and necessary to target the expressed toxin proteins for effective differentiation of this group.

The use of MALDI-TOFMS protein profiling as a true diagnostic tool for the BACT group organisms has not been solidified and its effectiveness would be based on

the goal of the classification or identification. The MALDI analysis tracks closely with the DNA hybridization results but has not been shown to be an effective discriminator of pathogen versus nonpathogen strains. Put simply, the MALDI analysis lends itself more towards the taxonomists goals of classification and perhaps slightly away from a pragmatic approach based on pathogenic properties. To ascertain the ability of the technique to do both, the profile library should be expanded to include more strains with and without toxin proteins. The ultimate goal could then be shifted to identify both species- and pathogen-specific biomarkers in each strain.

CHAPTER 6
PEPTIDE PROFILING AND BIOMARKER IDENTIFICATION FOR SELECTED *Bacillus* SPECIES

Introduction

MALDI-TOFMS protein profiling of proteins extracted from whole spores and cells has proven useful for the differentiation and identification of microorganisms. The profiles obtained from such analyses can be used in several ways. Profiles can be analyzed by statistical methods for comparing them with standardized libraries of microbial profiles in order to identify the organism at the species and strain level.[20,71,72,101,103] In another approach, peak masses extracted from the profiles can be matched to protein masses predicted from the genome.[104-106] In both cases, the analyses are hindered by the inherent variation among microbial species and by the low mass accuracy, lower resolution and sensitivity, and variability of protein profiles obtained with MALDI-TOFMS. To overcome these limitations, several groups have explored using bioinformatics-based approaches, using proteolytic peptides generated by tryptic digestion of proteins from intact microorganisms for identification purposes.[29,107-111] Using peptide fragments allows for higher sensitivity, mass resolution, and mass accuracy than can be obtained for the precursor proteins. Tandem mass spectrometry of peptides to obtain peptide mass tags has also been performed using collision induced dissociation (CID) or post source decay (PSD).

Recent studies of the genus *Bacillus* have been focused on the release and digestion of small acid soluble proteins (SASPs).[99,100] Fenselau, *et al*. have shown that treatment

with 10% TFA followed by digestion using trypsin immobilized on agarose beads results in the production of tryptic peptides from SASPs in under 30 minutes.[99,100] Using this approach, combined with a specially designed database of *in silico* produced tryptic peptides from the sequences of all SASP proteins available in the NCBI database, they were able to differentiate *B. anthracis* Sterne from closely related species. However, they were unable to differentiate *B. cereus* T from *B. thuringiensis* Kurstaki HD-1 based on the tryptic peptides generated from the SASPs.

While SASPs do allow for some species differentiation, they do not account for all the peaks observed nor do they allow for differentiation at the species level as in the case of *B. thuringiensis* and *B. cereus.*[32,38,40,99,100] Therefore, the ability to visualize and identify coat proteins as well as small acid soluble proteins in the spore is critical for complete and effective species differentiation of the *Bacillus* genus. A solvent extraction system for protein profiling of whole spores has been developed that targets the more hydrophobic coat protein constituents instead of the SASPs. Using this system, the differentiation of over 50 strains encompassing 15 species of *Bacillus* species has been accomplished. Additionally, strain variation in the protein profiles from a group of *B. pumilus* and *B. cereus* and *B. thuringiensis* serotypes has been evaluated. These studies, which use linear correlation analysis to compare profiles, have not been directed at assigning identities to the peaks in the profile spectra. The goal of this chapter of the research is to identify the proteins which produce peaks in the MALDI profile to confirm the presence of spore coat proteins in the spectra and to begin to eliminate proteins that are expressed only in response to external stimuli. Armed with this additional information, a subset of peaks, which is species specific, can be identified as solid

biomarkers for analysis, allowing for identification regardless of experimental and environmental factors. The identification of the proteins in the spectra is the backbone needed to bring this technique into fruition as a viable microbial analysis tool.

Spores were extracted using the MALDI protocol described previously.[101] MALDI extracts from several species of *Bacillus,* including four putative *B. pumilus* isolates, were digested with trypsin. Profiles of the tryptic peptides were obtained using MALDI-TOFMS and were compared using linear correlation analysis. Peptide profiles obtained for the different species analyzed were distinguishable from each other but had higher correlation values than their corresponding protein profiles. Identification of the peaks in the MALDI profiles was pursued by using capillary liquid chromatography-tandem mass spectrometry (CLC-MS^2) to obtain peptide fragment information from the digested proteins in the MALDI extracts. The fully sequenced, *B. subtilis* 168 had the greatest number of coat proteins identified with 4 coat associated proteins identified in the MALDI extract. Other species examined resulted in the identification of mainly SASP associated proteins due to their high sequence conservation. Additional proteins identified include proteases, hydrolases, transport and membrane associated proteins, and several hypothetical proteins with no known function. A PE-PGRS family protein from *Mycobacterium* was also found in all of the *B. pumilus* group organisms except the type strain ATCC 7061. The only additional coat proteins identified include the spoIVA protein in *B. pumilus* 7061 and a cotT homologue in FO-36b. The ability to rectify the identified protein masses from the CLC-MS^2 analysis with the MALDI protein profiles proved to be difficult. This is most likely due to post translational modifications of

proteins prior to deposition in the spore coat and lack of complete sequence information for the species studied.

Materials and Methods

Bacterial Strains

The type strains of *B. licheniformis, B. mojavensis, B. thuringiensis,* and *B. pumilus* were procured from the American Type Culture Collection (ATCC, Manassas, Va). *B. subtilis* 168 was a gift from Dr. Wayne Nicholson at the University of Florida and *B. subtilis* NB200 and JH642 were a gift from Dr. Arnold Aronson at Purdue University. *B. odysseyi* PTA-3499*, B. niacini* 51-8C, FO-11, FO-36b, SAFN-029, SAFN-036, SAFR-032, and *B. psycrodurans* VSE1-06, were isolated from several NASA spacecraft and assembly facilities surfaces. Identity of the test organisms was determined based on 16S rDNA sequencing for the environmental isolates; whereas for the ATCC strains, those sequences available in the GenBank database were used.[61] Preparation of spores followed standard protocols described previously.[7,73]

Protein Extraction and Digestion

A 10 µL aliquot of each spore suspension (0.6 OD_{660}) was diluted in 100 µL of 30% formic acid/30% acetonitrile. This sample was vortexed briefly and then centrifuged for 5 minutes at 9600 x g. The supernatant (MALDI extract) was removed and placed in a clean microfuge vial. The solvent was removed with a speed vac and the sample was reconstituted in 50 µL of 50 mM ammonium bicarbonate with 0.1 µg/µL sequencing grade trypsin (Promega). This concentration was found to be optimal for the efficient digestion of the extracts as less trypsin resulted in incomplete digestion. The proteins were digested overnight at 37°C.

Peptide profiling

To perform the peptide profiling, 2.5 μL of the digested sample from each spore species was diluted into 10 μL of a MALDI matrix solution containing 10 mg/mL HCCA matrix in 50% ACN/ 0.1% TFA. A 1 μL aliquot of this solution was then spotted onto a MALDI plate for analysis. MALDI-TOFMS analysis was performed on a Bruker Daltonics Reflex II Mass Spectrometer (Bruker Daltonics, Billerica, Ma) retrofitted with delayed extraction. The instrument was operated in the linear mode. A nitrogen laser (337 nm) pulsed at a frequency of 5 Hz irradiated the sample. Spectra were obtained in the positive ion mode with a delay time of 50 ns. The acceleration voltage was 20 kV. External calibration was performed using a mixture of angiotensin I and II and ACTH I and II.

All spectra represent the accumulation of 50 laser shots. Ten spectra were collected from each spot on the MALDI plate. A total of 20 spectra were collected per sample. Each spectrum was baseline corrected and smoothed using a ten-point Savitzky-Golay smoothing algorithm prior to statistical analysis. Linear correlation analysis was performed on software developed in-house with Visual Basic 6.0 as described previously.[73] A library of the peptide spectra was complied by averaging the 20 spectra collected from each digest sample. Average spectra obtained from the peptide profiling of strains in this study were compared to MALDI-TOFMS profiles stored in this library. Higher correlation coefficients are indicative of spectral similarity.

1-D Gel Electrophoresis

1-D gel electrophoresis was performed on both a whole spore coat extract (SDS) and a MALDI extract. To prepare the extract, approximately 500 μL of spores (0.6 OD_{660}) was centrifuged and the supernatant was removed. The whole spore extract was

prepared by extracting the spores with 50 µL of a tricine sample buffer (Biorad, Hercules, CA) containing 200 mM Tris-HCl at pH 6.8, 2% SDS, 40% glycerol, and 0.04% Coomassie Brilliant Blue G-250. For the MALDI extract, 1 mL of MALDI solvent (30% formic acid/30% acetonitrile) was first added to the spores. The spores were vortexed for 30 seconds and centrifuged in order to remove the supernatant. The spore-free supernatant was then evaporated in a speed vac, leaving behind only those proteins extracted by the MALDI solvent. These proteins were reconstituted in 50 µL of tricine sample buffer. Both samples were then boiled for 10 minutes, with a 1 minute vortexing step in the middle. For the whole coat extract, the solution was centrifuged and the supernatant removed for loading onto a gel; the MALDI extract could be loaded directly. Samples were loaded onto a Criteron precast 16.5% Tris-tricine gel (Biorad, Hercules, CA) and were run at 150 V for 1.5 hours. The gel was subsequently stained with Coomassie R-250.

Proteomic Analysis

The MALDI protein extracts from the spores of *B. subtilis* 168, *B. pumilus* 7061, FO-11 (wild-type *B. cereus*), *B. thuringiensis* 10792, FO-36b, SAFN-036, SAFN-029, SAFR-032 and *B. licheniformis* 14580 were subjected to proteomic analysis. Tryptic peptides were analyzed by capillary liquid chromatography-tandem mass spectrometry ($CLC\text{-}MS^2$) using a system similar to that described elsewhere.[112] Sequence information was obtained for tryptic peptides via collision-induced dissociation. The mass-to-charge (*m/z*) ratio of the precursor ion and product ions for each tryptic peptide were searched against the NCBI (National Center for Biotechnology Information) protein database using the Sequest[113] and Mascot[114] algorithms for protein identification. All searches used a peptide mass tolerance of 1.8 Da and a fragment mass tolerance of 1 Da. Only the

oxidation of methionine was considered as a modification. Initial database searches were performed against the entire database (all organisms) using trypsin for enzyme specificity, allowing for 1 missed cleavage. Because of the unique structure of many of the proteins identified in this study, it was necessary in subsequent searches to use no enzyme specificity. In these searches, the database search was limited to the gram positive bacteria in order to decrease the search time.

Results and Discussion

Peptide Profiling

Peptide profiling was performed on the 10 *Bacillus* species in this study. The peptide profiling was performed in linear mode on the MALDI-TOFMS. The correlation results of the peptide profiling are shown in Table 6-1. Correlation values between the different species were higher than those seen with whole-cell protein profiling studies[73] and several were above the 0.75 correlation value used for species differentiation in protein profiling studies. However, using the Student's t-test, complete differentiation of the peptide profiles was possible. For all species, the difference (delta value) between the first and second hit was >0.1. The applicability of the same criteria used for protein profiling ($r > 0.75$) needs to be further evaluated using peptide profiling.

Average spectra from each sample are shown in Figure 6-1. *B. pumilus*, *B. thuringiensis*, and *B. odysseyi* resulted in spectra that were very similar from *m/z* 600-2000 but had differences in the higher *m/z* range (*m/z* 2000-5000) and had correlation values ranging from 0.81-0.85 with each other. *B. licheniformis* and *B. psychrodurans*, with a correlation value of 0.85, are characterized by an absence of high intensity peptide peaks that were seen in the spectra of the other species. A notable similarity between all the spectra is the presence of biomarker peaks at *m/z* 1330 and 1185. Though the identity

of these peptides is presently unknown, they could be genus-specific peptide biomarkers as they are not known contaminants resulting from keratin or trypsin autolysis peaks.

Table 6-1. Correlation values based on the peptide profiles

Trypsin Digest Fingerprint From		*B. subtilis*	*B. cereus*	*B. licheniformis*	*B. mojavensis*	*B. niacini*	*B. odysseyi*	*B. psychrodurans*	*B. thuringiensis*	*B. pumulis*	SAFN-029	SAFR-032	FO36b	SAFN-036
B. subtilis	**168**	1.00												
B. cereus	**FO-11**	0.38	1.00											
B. licheniformis	**ATCC 14580**	0.45	0.77	1.00										
B. mojavensis	**ATCC 51516**	0.33	0.44	0.53	1.00									
B. niacini	**51-8C**	0.12	0.23	0.25	0.19	1.00								
B. odysseyi	**34hs1**	0.30	0.74	0.60	0.36	0.19	1.00							
B. psychrodurans	**VSE1-06**	0.43	0.75	0.85	0.52	0.25	0.59	1.00						
B. thuringiensis	**ATCC 10792**	0.20	0.67	0.46	0.28	0.14	0.81	0.42	1.00					
B. pumulis	**ATCC 7061**	0.27	0.68	0.53	0.31	0.16	0.73	0.50	0.85	1.00				
B. pumulis	**SAFN-029**	0.19	0.38	0.48	0.25	0.11	0.31	0.42	0.27	0.32	1.00			
B. pumulis	**SAFR-032**	0.23	0.47	0.54	0.29	0.14	0.38	0.50	0.33	0.40	0.95	1.00		
***B. pumulis*/FO group**	**FO-36b**	0.25	0.64	0.67	0.33	0.15	0.41	0.56	0.36	0.37	0.43	0.44	1.00	
***B. pumulis*/FO group**	**SAFN-036**	0.37	0.71	0.87	0.44	0.19	0.50	0.70	0.39	0.49	0.37	0.41	0.64	1.00

1-D Gel Electrophoresis

1-D gel electrophoresis was used to ascertain the effectiveness of the MALDI extraction when compared to a whole spore coat (SDS) extract. Figure 6-2 shows the results of this experiment for the type strains of *B. subtilis*, *B. pumilus*, *B. licheniformis*, and *B. atrophaeus*, *B. megaterium*, and *B. subtilis* 168. The lanes alternate across the gel showing the whole coat extract next to the MALDI extract (30% formic acid/30% acetonitrile) for each species. Clearly, the MALDI solvent is only able to extract a subset of proteins from the spores when contrasted with the SDS extract. The higher concentration of formic acid used in this experiment did allow us to extract and subsequently visualize, by MALDI-TOFMS, a more diverse set of proteins from the spores. However, the resolution obtained in the gel based method is far less than that obtained using the MALDI-TOFMS protein profiling. Many of the proteins visualized in the protein profile are not resolved in the gel analysis.

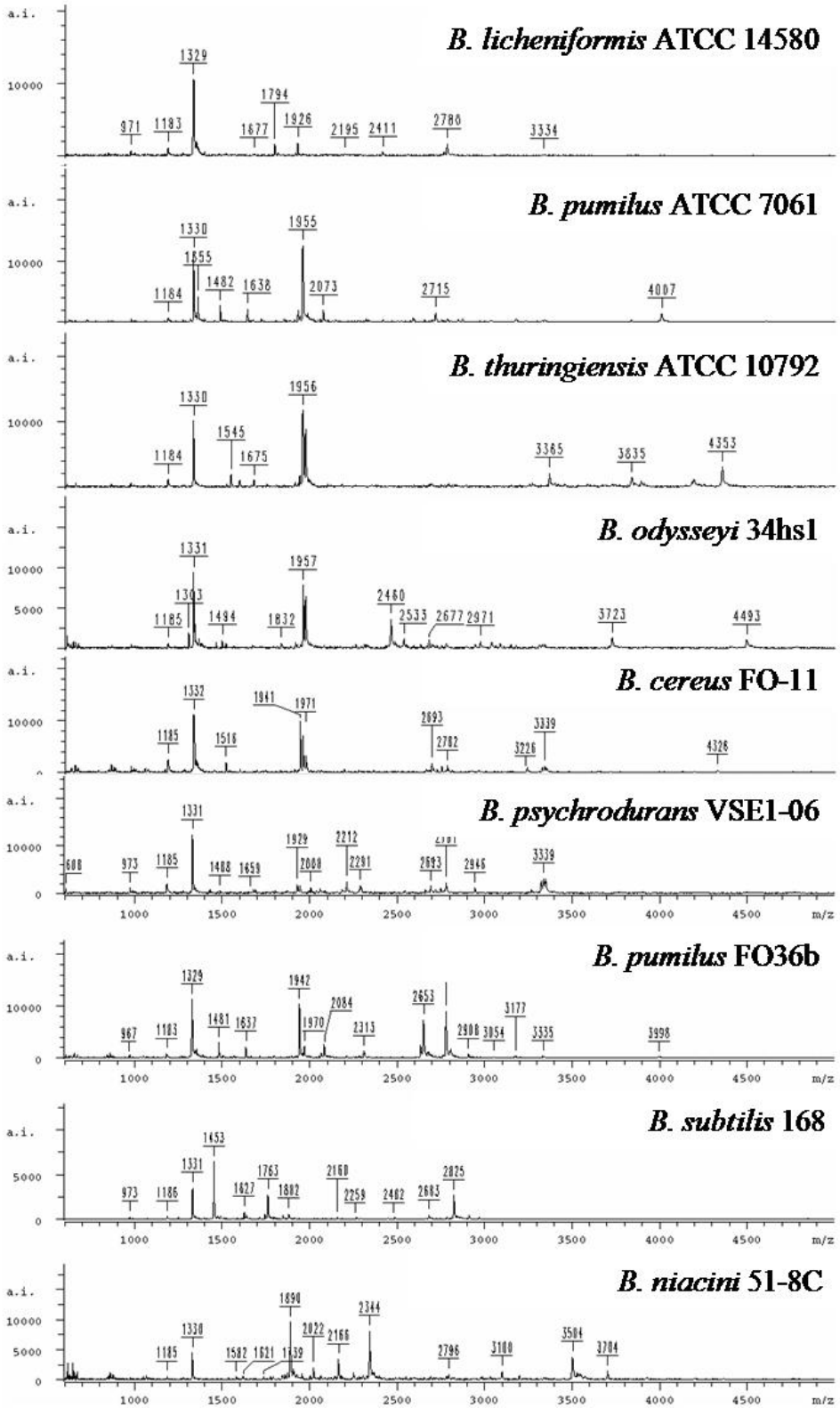

Figure 6-1. Peptide profiles obtained for the different species examined in this study. The scale is from *m/z* 500-5,000.

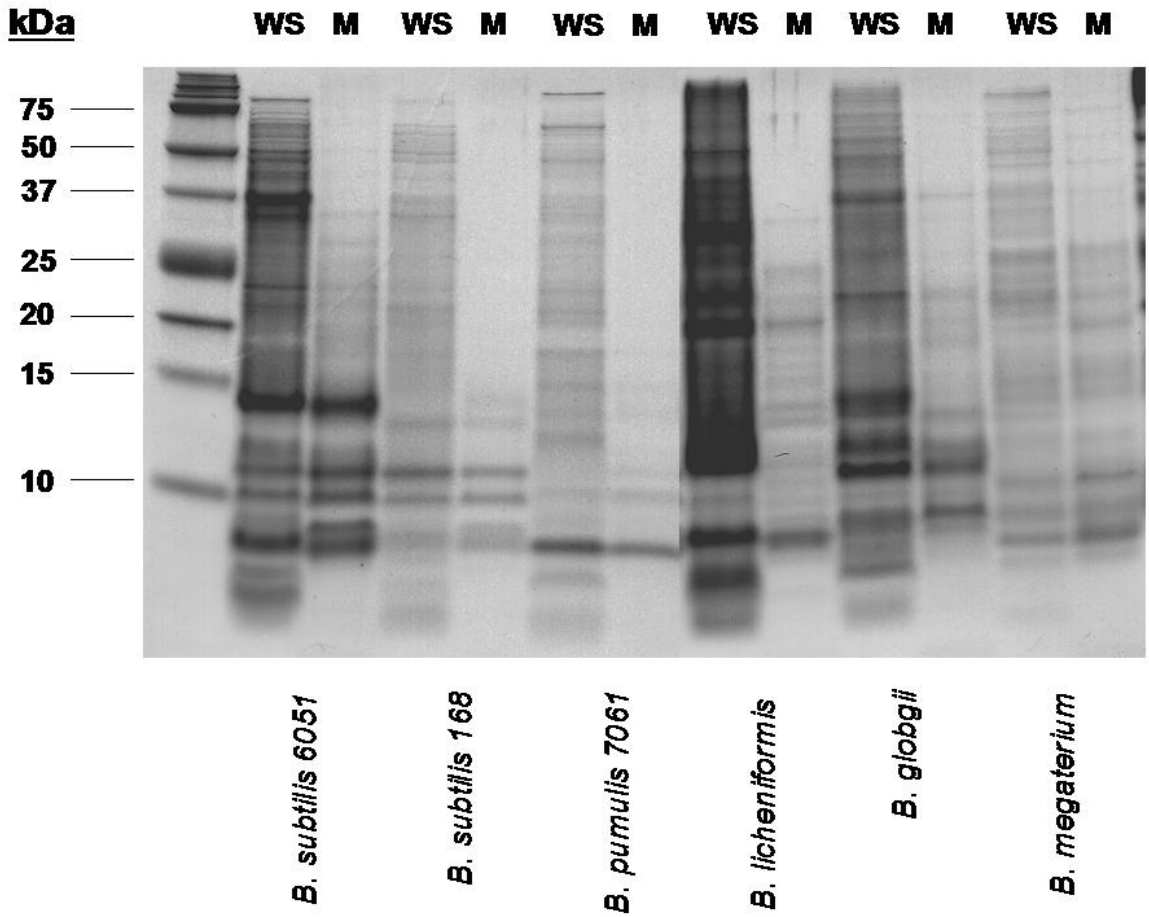

Figure 6-2. 1-D gel electrophoresis showing the whole spore coat extract (WS) and MALDI extract (M) for 6 *Bacillus* strains

In examining the gel, it is also interesting to note the difference in the number and amount of proteins that are extracted from each species using both the SDS and formic acid treatment. The extraction of protein from the spore coat can be affected by the number of coat proteins, by cross linking of the proteins, and by the presence or absence of an exosporium. These effects were species specific, as can be seen in the gel, where species such as *B. subtilis* 6051, *B. globgii*, and *B. licheniformis* appeared to have more coat proteins than the other species. *B. pumilus* 7061 is characterized the lowest number of coat proteins extracted using the formic acid treatment.

Proteomic Analysis for Biomarker Identification

Peptide profiling, while useful as an alternate method for the characterization and classification of spore species, still does not explain the source of the peptides and thus the source of the biomarkers in the MALDI spectra from the whole spores. To accomplish this goal, CLC-MS^2 has been employed to obtain peptide mass tag on the tryptic peptides from the MALDI extracts.

Since the genome of *B. subtilis* 168 is completely sequenced, this organism was selected first for further proteomic studies directed at identifying which proteins are represented by the biomarker peaks observed by MALDI-TOFMS. Fourteen proteins were identified in the extract of the *B. subtilis* 168 spore sample using CLC-MS^2. The protein description, sequence coverage, molecular weight, and database searching scores are shown in Table 6-2. Four small acid soluble proteins, A, B, C, and D, one DNA binding protein, hypothetical protein ymfJ, ribosomal protein L12, a phosphocarrier protein, a protein similar to 1-pyrroline-5-carboxylate dehydrogenase, and 4 proteins

associated with the spore coat and spore coat formation, coat JB, coat F, coat T, and spoIVA were identified.

Table 6-2. Proteins identified from B. subtilis 168 using CLC-MS2

B. subtilis 168				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
16080009	small acid-soluble spore protein (major alpha-type SASP) sspA - *B. subtilis*	7066	375	68.7
16078040	small acid-soluble spore protein (major beta-type SASP) sspB - *B. subtilis*	6975	287	61.2
16081105	spore coat protein cotF - *B. subtilis*	18714	176	37.5
16078751	conserved hypothetical protein ymfJ - *B. subtilis*	9642	114	56.5
16077757	spore coat peptide assembly protein cotJB - *B. subtilis*	11745	76	26
230576	Histidine-Containing Phosphocarrier Protein Hpr Mutant With Met 51 Replaced By Val And Ser 83 Repl	9037	88	32.6
32468826	small acid-soluble spore protein [*B. subtilis*]	5159	77	4.6
16077390	similar to 1-pyrroline-5-carboxylate dehydrogenase [*B. subtilis*]	56454	71	51.2
1075916	Spore Coat Protein Precursor- cotT *B. subtilis*	10125	69	59.7
16078411	small acid-soluble spore protein (minor alpha/beta-type SASP) sspD - *B.*	6800	67	43.7
16077173	ribosomal protein L12 (BL9) [*B. subtilis*]	12743	67	16.3
9630297	small acid-soluble spore protein C - *B. subtilis*	7753	65	27.8
16079336	DNA-binding protein HU - *B. subtilis*	9878	58	41.3
16079337	coat morphogenesis sporulation protein spoIVA - *B. subtilis*	55140	50	4.3

Relating these proteins to the peaks observed in the MALDI-TOFMS profile for *B. subtilis* 168 was challenging. It is important to note that the masses contained within the database include the methionine residue as the start of the protein, which is in most cases removed during processing. Therefore, when looking for intact proteins in the protein profiles, 131 Da should be subtracted from the molecular weight given in the databases.

Proteins in the lower molecular mass region (under *m/z* 10,000) include the small acid soluble proteins, the DNA binding protein, the ribosomal protein, and smaller spore coat polypeptides processed from larger precursors. We postulate that the higher *m/z* peaks represent other processed and intact spore coat proteins such as cotJB at *m/z* 11,638 Da. The large peak at *m/z* 7,758 is the processed form of cotT, which starts as a 10kDa protein in which the first 19 residues (termed the propeptide) are removed to leave behind a ~7,800 Da spore coat protein.[115] To confirm this identification, two additional *B. subtilis* strains, JH 642 and NB 200, were examined. NB 200 is a cotT knockout mutant

derived from the parent strain JH 642. Figure 6-3 shows the average spectrum of *B. subtilis* 168 compared with JH 642 and NB 200. As expected, the large 7,758 Da peak is present in both *B. subtilis* 168 and *B. subtilis* JH 642 but is missing from the cotT mutant NB200. This evidence confirms the assignment of the 7,758 Da peak to the processed from of the cotT protein.

The molecular mass listed in the protein database for the remaining proteins identified, including spoIVA, ymfJ, cotF, and sspB, did not directly match with the *m/z* of singly charged ions observed in the MALDI spectra. Separations of the proteins prior to proteomic analysis is required in order to reduce the complexity of the biomarker protein extract and confidently assign protein identifications made by CLC-MS^2 to peaks in the MALDI spectra. This is the first time that proteins associated with the spore coat have been identified from direct spore analysis using a MALDI extract, as previous studies have only identified small-acid soluble proteins found in the spore cortex as the source of the biomarker peaks.

Rectifying proteins identified in the CLC-MS^2 study for *B. subtilis* 168, a fully sequenced and highly annotated species, with peaks in the MALDI-TOFMS protein profile proved to be very difficult. This difficulty resulted from the post-translational processing and modifications that occur in proteins. Therefore, the success of doing this for species that do not have full sequence information available will be limited. The only currently available, fully sequenced genomes for *Bacillus* include *B. subtilis* 168 and *B. anthracis* Ames. Limited database entries exist for *B. megaterium*, *B. halodurans*, *B. stearothermophilus,* and *B. cereus*. Further proteomic analysis of several other species was pursued in order to ascertain what kind of proteins would be identified using the

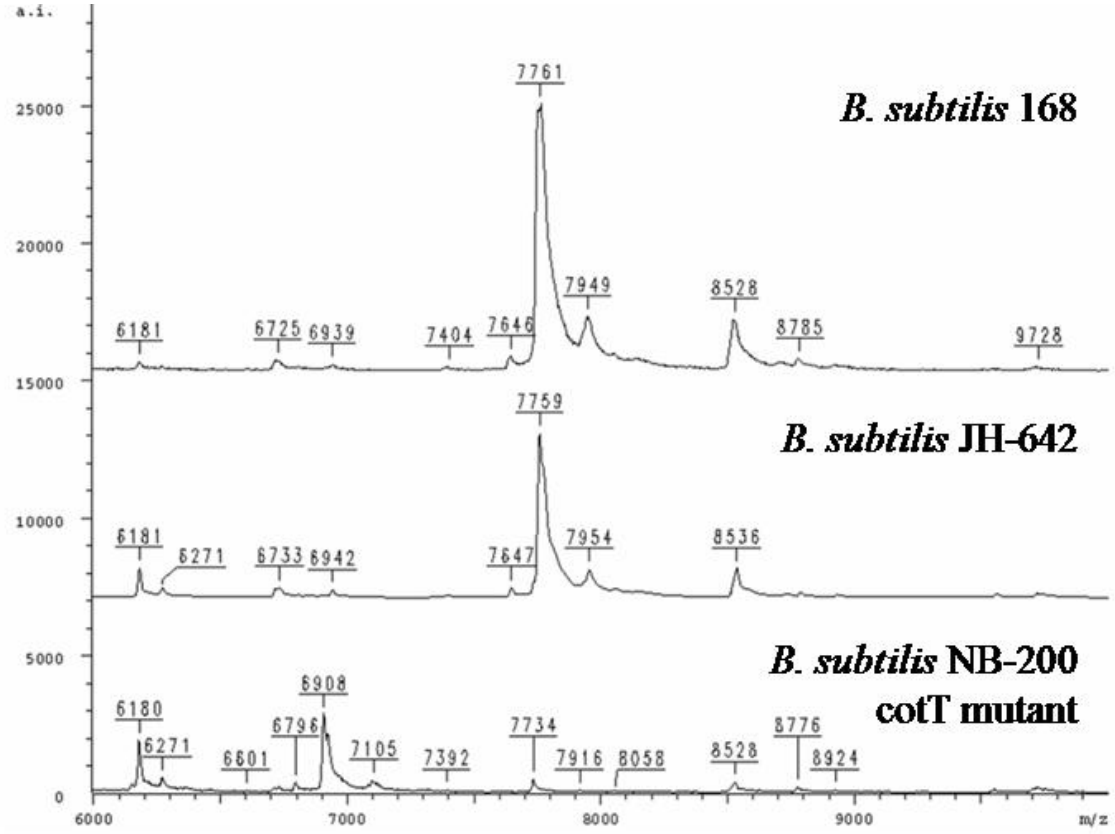

Figure 6-3. Comparison of the protein profile of the cotT protein using a cotT mutant. On the top is the *B. subtilis* 168 genetic strain, the middle is the JH-642 strain which is the parent strain, and NB-200 on the bottom which is the mutant with the cotT protein deleted from its sequence. Note the absence of the large 7,760 Da biomarker in the NB-200 mutant strain.

CLC-MS^2 approach. The proteins identified from the analysis of *B. licheniformis*, FO-11, *B. thuringiensis*, and several of the *B. pumilus* group bacteria including the type strain 7061 and FO36b, SAFN-029, SAFN-036, and SAFR-032 are given in Table 6-3 through 6-10.

The majority of the proteins identified from the unsequenced species were SASP associated proteins. This is expected due to the high sequence conservation among the SASP proteins. Spore coat proteins, on the other hand, have high sequence divergence, and a significant overlap with other species is not expected. We believe this is why no coat proteins are identified. Even though no spore coat proteins were identified, the analysis did result in the identification of other functional categories of proteins in these spores which have been observed in other proteomic analyses.[37,41,42] These categories of proteins included enzymes responsible for transport, translation, and various metabolic processes, including proteases, hydrolases, and synthases. The presence of these types of proteins in spores is not surprising as they function in protein synthesis and degradation and other cellular processes that allow the spore to adapt to atypical conditions. Many of these proteins are believed to be active in the signaling of the spore to break dormancy and return to a vegetative state. In *B. thuringiensis*, a hypothetical protein identified (NCBI accession number 21401766) shows high homology to the ymfJ protein predicted and detected in *B. subtilis* 168. Unfortunately, none of the adjusted molecular weights from the databases matched the *m/z* of peaks in the spectra for any of these other species. This is likely due to the fact that although there are peptides identified from these proteins, the coverage is by no means complete. The variable regions of protein in these species likely remain unidentified in the database search. In addition to the precursor

Table 6-3. Proteins identified from *B. licheniformis* using CLC-MS^2

B. licheniformis				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
13676638	small acid-soluble protein gamma-type [Bacillus subtilis]	9339	162	47.1
16079336	non-specific DNA-binding protein HBsu [Bacillus subtilis]	9878	151	30.4
80156	DNA-binding protein HB - Bacillus sp	9891	150	30.4
30022721	Small acid-soluble spore protein [Bacillus cereus ATCC 14579]	6837	113	52.3
134246	small acid-soluble protein gamma-type SASP	9015	110	61
21399863	Small, acid-soluble spore proteins, alpha/beta type [Bacillus anthracis A2012]	7208	89	23.5
30020123	Small acid-soluble spore protein [Bacillus cereus ATCC 14579]	7438	89	23.5
23019957	COG0168: Trk-type K+ transport systems, membrane components [Clostridium thermocellum ATCC 27405]	37836	76	6.7
3015572	gamma-type small, acid-soluble spore protein [Bacillus aminovorans]	15351	71	38.6
16078040	small acid-soluble spore protein (beta-type SASP) [Bacillus subtilis]	6975	70	29.4

Table 6-4. Proteins identified from *B. thuringiensis* using CLC-MS^2

B. Thuringiensis				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
21402693	Small, acid-soluble spore proteins, alpha/beta type [*B. anthracis* A2012]	6805	422	70.8
21399863	Small, acid-soluble spore proteins, alpha/beta type [*B. anthracis* A2012]	7208	309	60.3
30020123	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	7438	308	60.3
21398813	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	6961	259	53.7
1710854	small acid-soluble spore protein 2 (SASP) *Sporosarcina Ureae*	7172	229	28.3
134241	small acid-soluble spore protein C-5 *Bacillus Mmgaterium*	7675	229	26
134242	small acid-soluble spore protein A *Bacillus megaterium*	6387	220	41.9
21401766	hypothetical protein predicted by GeneMark [*Bacillus anthracis* A2012]	9333	115	30.5
30021283	IG hypothetical 17224 [*Bacillus cereus* ATCC 14579]	17520	51	6.7
21397483	hypothetical protein predicted by GeneMark [*Bacillus anthracis* A2012]	7008	50	21.3

Table 6-5. Proteins identified from FO-11 using CLC-MS2

FO11				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
30022721	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	6837	185	60
21398813	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	6961	180	73.5
21399863	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	7208	170	39.7
30020123	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	7438	170	
21401004	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	7290	111	31.4
32266031	conserved hypothetical protein [*Helicobacter hepaticus* ATCC 51449]	76791	66	1.9
30021283	Uncharacterized conserved protein, YLXR *B.subtilis* homolog [*Clostridium acetobutylicum*]	10276	41	5.4

Table 6-6. Proteins identified from FO-36b using CLC-MS^2

FO36b				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
134224	small acid-soluble spore protein 1 *Bacillus Stearothermophilu* s	7223	270	60
21402693	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	6805	264	49.2
21399863	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	7208	253	39.7
30020123	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	7438	250	39.7
134243	small acid-soluble spore protein C *Bacillus megaterium*	7550	242	38.8
3688809	sspA *Bacillus firmus*	7152	219	34.3
16078040	small acid-soluble spore protein (beta-type SASP) [*Bacillus subtilis*]	6975	209	35.8
16080009	small acid-soluble spore protein (alpha-type SASP) [*Bacillus subtilis*]	7066	208	35.8
21398813	SASP, Small, acid-soluble spore proteins, alpha/beta type [*B. anthracis* A2012]	6961	200	44.8
1710854	small acid-soluble spore protein 2 (SASP) *Sporosarcina ureae*	7172	198	26.8
134241	small acid-soluble spore protein C-5 *Bacillus megaterium*	7675	229	24.6
134246	small acid-soluble spore protein, gamma type *B. Stearothermophilus*	9015	135	64.6
21633218	HlbB [*Lactobacillus delbrueckii* subsp. *bulgaricus*]	9740	110	24.4
31794684	PE-PGRS Family Protein [*Mycobacterium bovis* subsp. *bovis* AF2122/97]	113121	94	11.8
3015572	gamma-type small, acid-soluble spore protein [*Bacillus aminovorans*]	15351	93	19.4
13676638	small acid-soluble protein gamma-type [*Bacillus subtilis*]	9330	87	54.1
1075916	Spore Coat Protein Precursor- cotT *Bacillus subtilis*	10125	78	36.6
37522475	ATP synthase delta chain of CF(1) [*Gloeobacter violaceu* s]	21005	77	7.6
37522856	photosystem I subunit VII [*Gloeobacter violaceus*]	8792	75	13.6
16077202	ribosomal protein L30 (BL27) [*Bacillus subtilis*]	6634	72	25.4
37521939	protochlorophyllide reductase iron-sulfur ATP-binding protein [Gloeobacter	29977	69	4.3
31793669	PE-PGRS Family Protein [Second Part] [*Mycobacterium bovis* subsp. *bovis* AF2122/97]	91009	68	25.6
30248366	hypothetical protein [*Nitrosomonas europaea* ATCC 19718]	37139	68	10.8

Table 6-7. Proteins identified from SAFN-036 using CLC-MS^2

SAFN-036				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
21402693	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	6805	185	61.5
21398813	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	6961	143	79.1
134224	small acid-soluble spore protein 1	7223	108	34.3
134246	small acid-soluble spore protein, gamma type	9015	79	31.7
134246	small acid-soluble spore protein 1 [*Oceanobacillus iheyensis* HTE831]	6883	73	41.7
134229	small acid-soluble spore protein 1	7785	67	41.7
21399863	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	7208	70	44.1
30020123	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	7438	70	44.1
15923506	2-amino-4-hydroxy-6-hydroxymethyldihydropteridin e pyrophosphokinase *Staphylococcus aureus* subsp.	17990	69	6.3
160800009	small acid-soluble spore protein (alpha-type SASP) [*Bacillus subtilis*]	7066	64	28.9
28868478	conserved hypothetical protein [*Pseudomonas syringae* pv. tomato str. DC3000]	37335	63	3.8

Table 6-8. Proteins identified from SAFN-029 using CLC-MS^2

SAFN-029				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
30022721	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	6837	225	86
21402693	SASP, Small, acid-soluble spore proteins, alpha/beta type [*B. anthracis* A2012]	6805	217	86
21399863	SASP, Small, acid-soluble spore proteins, alpha/beta type [*B. anthracis* A2012]	7208	210	41
30020123	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	7438	209	
134243	small acid-soluble spore protein C	7550	178	55
16078040	small acid-soluble spore protein (beta-type SASP) [*Bacillus subtilis*]	6975	145	36
160800009	small acid-soluble spore protein (alpha-type SASP) [*Bacillus subtilis*]	7066	142	
1710854	small acid-soluble spore protein 2 (SASP)	7172	145	43
21398813	Small, acid-soluble spore proteins, alpha/beta type [*B. anthracis* A2012]	6961	138	46
143652	Hbsu Protein	2660	109	16
16079336	non-specific DNA-binding protein HBsu [*Bacillus subtilis*]	9878	108	
80156	DNA-binding protein HB - *Bacillus* sp	9891	106	
22537119	NADH oxidase [*Streptococcus agalactiae* 2603V/R]	49823	85	3.9
37519783	unknown protein [*Gloeobacter violaceus*]	78729	81	1.8
15608383	PE_PGRS [*Mycobacterium tuberculosis* H37Rv]	47258	70	23
15608210	echA8 [*Mycobacterium tuberculosis* H37Rv]	27256	68	5.4
538879	Ig light chain-binding protein precursor - *Peptostreptococcus magnus*	78935	65	3.3
17986643	ABC Transporter ATP-Binding Protein/ ABC Transporter Permease Protein [*Brucella melitensis*]	40568	62	3.9
27365336	Acetyl-CoA carboxylase beta subunit [*Vibrio vulnificus* CMCP6]	33994	59	7.4
31792838	PE-PGRS Family Protein [*Mycobacterium bovis* subsp. *bovis* AF2122/97]	88941	59	6.8
23113959	COG0015: Adenylosuccinate lyase [*Desulfitobacterium hafniense*]	49281	57	5.6
28897257	glutamate synthase, small subunit [*Vibrio parahaemolyticus* RIMD 2210633]	53348	57	11
15614888	BH2325~unknown conserved protein in B. subtilis [*Bacillus halodurans*]	19773	57	12.3
15805864	hypothetical protein [*Deinococcus radiodurans*]	23998	56	11.1

Table 6-9. Proteins identified from SAFR-032 using CLC-MS^2

SAFR-032				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
16078040	small acid-soluble spore protein (beta-type SASP) [*Bacillus subtilis*]	6975	135	55.2
160800009	small acid-soluble spore protein (alpha-type SASP) [*Bacillus subtilis*]	7066	133	55.2
134243	small acid-soluble spore protein C	7550	120	51.4
29567098	39K protein [*Adoxophyes honmai nucleopolyhedrovirus*]	30218	86	4.2
32474990	hypothetical protein-putative transmembrane protein [*l* .]	30698	75	4.3
15615456	BH2893~unknown conserved protein in bacilli [*Bacillus halodurans*]	6819	71	30.5
21400449	Peptidase_M4, Thermolysin metallopeptidase, catalytic domain [*B. anthracis*	40054	68	3.6
31794687	PE-PGRS FAMILY PROTEIN [*Mycobacterium bovis* subsp. *bovis* AF2122/97]	151940	68	6.75
15610447	hypothetical protein Rv3311 [*Mycobacterium tuberculosis* H37Rv]	45704	67	2.9
22991783	COG0636: F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K [*Enterococcus sp* .]	7295	67	31
32265695	hypothetical protein [*Helicobacter hepaticus* ATCC 51449]	25991	64	3.4
15925035	conserved hypothetical protein [*Staphylococcus aureus* subsp. *aureus* Mu50]	39273	63	2.2
20808467	hypothetical protein [*Thermoanaerobacter tengcongensis*]	33693	60	3.9
16081074	yydB [*Bacillus subtilis*]	56425	59	1.6

Table 6-10. Proteins identified from *B. pumilus* 7061 using CLC-MS[2]

B. pumilus 7061				
Accession Number	Protein Description	Molecular Weight	Mascot Score	% Coverage
30022721	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	6837	248	61.5
21398813	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	6961	248	77.6
21399863	Small, acid-soluble spore proteins, alpha/beta type [*Bacillus anthracis* A2012]	7208	193	57.3
30020123	Small acid-soluble spore protein [*Bacillus cereus* ATCC 14579]	7438	192	57.3
80126	small acid-soluble spore protein C-1 - *Bacillus megaterium*	7201	105	32.3
134223	small acid-soluble spore protein C1 - *Bacillus megaterium*	7329	105	32.3
134239	small acid-soluble spore protein C4- *Bacillus megaterium*	7343	105	32.3
80127	small acid-soluble spore protein C-2 - *Bacillus megaterium*	7598	104	32.3
134232	small acid-soluble spore protein C-2 - *Bacillus megaterium*	7726	104	32.3
21401004	SASP, Small, acid-soluble spore proteins, alpha/beta type [*B. anthracis* A2012]	7290	91	30
9630126	site-specific recombinase [Bacteriophage SPBc2]	63099	70	3.3
30021283	hypothetical protein [*Bacillus cereus* ATCC 14579]	17520	51	6.8
16079337	coat morphogenesis sporulation protein spoIVA - *Bacillus subtilis*	55140	50	4.3

protein masses not matching with peaks in the protein profile, very few of the peptides identified in the tandem mass spectrometry experiment match to peaks in the peptide profiles obtained by MALDI-TOFMS (Figure 6-1). This difference in the peptides identified can be attributed to differences in the ionization mechanisms between MALDI and electrospray and/or to the fact that the most intense peptide peaks in the MALDI profiles do not match to sequence data in the databases.

Of particular interest are the differences noticed in the *B. pumilus* strains analyzed. Based on the DNA-DNA hybridization data and the MALDI-TOFMS protein profile analysis, the *B. pumilus* group is split into two groups, the FO group and the type strain group. Spectra of the peptides from the FO group strains were very similar to one another, with >0.90 correlation values among them. The type strain group strains had more diversity in their spectral profiles. SAFN-036 and FO36b are members of the FO group and SAFN-029, SAFR-032, and ATCC 7061 belong to the type strain group. When looking at the peptide profiles in Table 6-1, the *B. pumilus* type strain and FO-36b have very low correlation values (<0.50) with all of the other strains in this study. The

highest correlation value (0.95) is found between SAFN-029 and SAFR-032, which are both in the type strain group.

The proteins identified from each of the *B. pumilus* strains are shown in Table 6-3. The type strain ATCC 7061 had the fewest number of proteins identified (eight). Five of these proteins were SASPs and one was the coat protein, spoIVA. FO36b had the largest number of unique proteins identified with 21 followed by SAFN-029 with 19, SAFR-032 with 13, and SAFN-036 with nine. All five of the *B. pumilus* group strains had high homology with the *B. megaterium* SASP C protein. All of the species except SAFR-032 also had peptides matching to SASPs associated with *B. anthracis* and *B. cereus* (21399863, 30020123, and 21398813). FO36b, SAFR-032, and SAFN-029 all had peptides identified from PE-PGRS family proteins. These are surface associated antigens with high glycine content found in *Mycobacteria.* Since these are surface associated antigens in another organism, it is possible these have similar functions and loci in the spores as well. It is interesting to note that although the SAFN-029 and SAFR-032 had very similar peptide profiles via MALDI-TOFMS, they resulted in the identification of many different proteins when compared to each other. The only overlaps between the strains were for SASPs and the PE-PGRS proteins described above.

FO36b had peptides which were identified as arising from the cotT protein from *B. subtilis*. This is the first report of the cotT protein being present in another organism. Other species that have been examined have not had a cotT-like protein and it was thought to be unique to *B. subtilis*.[116] The FO36b strain also had high homology with 13 SASPs including gamma type SASPs and SASPs from the extremophile *B. stearothermophilus*. This is interesting in light of the increased resistance of FO-36b in

comparison with other *Bacillus* strains. Further evaluation of the unique resistances and properties associated with the FO group strains and with SAFR-032 and SAFN-029 is necessary to correlate protein identification with the differing properties.

Conclusions

Profiling of the tryptic peptides obtained from the species in this study produced unique peptide spectra for each species. Possible genus specific peptides were observed at *m/z* 1330 and 1185. Correlation values between peptide profiles for the different species were above 0.75 in some cases. However, complete differentiation using the Student's t test was possible. The applicability of the same criterion used for protein profiling ($r > 0.75$) needs to be further evaluated for use with peptide profiling. The peptide profiling experiment requires additional sample processing and time. Little additional information was obtained when compared to the protein profiling study. The better resolution and sensitivity that can be achieved for peptide fragments may decrease the absolute limit of detection and the differentiation of closely related strains may be feasible using this approach.

Ultimately the goal of this study was to identify the biomarker peaks in the MALDI-TOFMS protein profiles. One dimensional gel electrophoresis of the MALDI extracts was attempted and abandoned based on the decreased resolution obtained on gels, strongly indicating the co-elution of multiple proteins in a single band. Instead, the MALDI extract containing all the extracted proteins was digested with trypsin and analyzed using CLC-MS^2. Using this proteomics approach, we were able to identify a number of proteins in the different species. However, correlation of this information to peaks in the MALDI profiles proved challenging due to post translational processing of proteins and due to the high sequence divergence expected for spore coat proteins.

Proteins identified in this study included 4 coat proteins in *B. subtilis* and a cotT type protein in FO-36b. This is the first time coat proteins have been identified as a source of the peaks observed in the MALDI profiles of whole spores. The cotT protein was also observed in an organism other than *B. subtilis* for the first time. The majority of proteins identified across the different species and strains were SASPs. The identification of SASPs and not coat proteins in the unsequenced organisms was expected due to the high sequence conservation among the SASPs. Little else can be said about the protein identifications that is not speculation at this point. Additional research on the resistances and unique properties of these organisms is needed to gain an understanding of the presence and absence of proteins in certain strains.

CHAPTER 7
IMPACT OF ENVIRONMENTAL FACTORS AND STERILIZATION ON THE MALDI-TOFMS PROTEIN PROFILE OF SPORE SPECIES

Introduction

Spores are highly resistant to many sterilization treatments and can withstand a wide range of environmental conditions. These conditions include desiccation, heat, radiation, and harsh chemical treatments. Although the exact mechanism for these resistances is not completely understood, many of them are protein-based. An example of this is the small acid soluble proteins (SASPs), which are known to play a role in DNA repair and radiation survival. Spore coat proteins, which provide structural integrity to the spore, can also serve as an additional barrier between the spore and challenges present in the outside world. If disruption, alteration, and/or removal of these proteins results from environmental exposure, there may be a detectable change in the protein biomarkers found in the spore species.

MALDI-TOFMS protein profiling has been established as a rapid diagnostic tool for bacterial differentiation of laboratory cultured isolates. In the field, real-time spore samples will be exposed to a wide variety of both controlled and uncontrolled environmental conditions. To translate this technology into a field-based analysis technique, the impact of these environmental conditions on the protein profiles should be evaluated. A comprehensive study on the impact of environmental factors and sterilization on the MALDI-TOFMS protein profile of spores could have several consequences. Spectral differences resulting from environmental conditions may prove

useful in providing forensic and epidemiology information for source tracking. Profile differences may also help to explain elevated resistance levels and may be able to provide viability information for an unknown sample. Additionally, having the option to first inactivate a pathogenic strain prior to transport, analysis, and cataloging is significant from a personnel safety standpoint.

The study reported in this chapter focused on a subset of environmental conditions including differences in the growth/sporulation state of the organism, initial sporulation conditions, storage conditions, aging of the spore crop, and exposure to radiation, heat, and chemical agents. Sporulation mutants were used as a model for differences in the sporulation state and incomplete spore formation that could be encountered in the environment. Lyophilized and desiccated spores were compared to spores stored in water. The aging effect of spores stored in water was also considered. Gamma and UV radiation, liquid and vapor hydrogen peroxide treatments, and autoclaving were used to mimic possible environmental exposures and to ascertain the effect of sterilization treatment on protein profiles. Spore strains that were resistant to select treatments are included to see if the profiles would provide information on viability. Spores grown on agar plates were compared with spores prepared with the standard preparation to determine if initial sporulation conditions might impact protein profiles.

Using these different factors with the optimized extraction conditions established for the analysis of spores, several questions will be asked and answered. Primarily, we sought to determine whether species-specific biomarkers were still present under the exposure conditions such that differentiation of the spore strains would still be possible. We then asked whether additional biomarkers provide evidence that could be used for

source tracking, forensic investigations, or epidemiology studies. We also evaluated whether differences observed in the protein spectra could be translated into viability information on the spore samples, as well as whether there are unique biomarkers present or variable alterations in resistant strains that may help to explain elevated resistances. The answers to these questions, discussed in this chapter, will provide a basis for the use of MALDI-TOFMS protein profiling as a rapid diagnostic tool for real-time bacterial identification.

Materials and Methods

MALDI-TOFMS Protein Profiling and Statistical Analysis

Samples were prepared and analyzed by MALDI-TOFMS using the optimized matrix solution and instrumental parameters as described previously.73 Spore samples were combined with the matrix solution consisting of ferulic acid in 30% acetonitrile, 30% formic acid. Sample preparation was done in duplicate. All spectra represented the accumulation of 50 laser shots. Ten spectra were collected across each spot for a total of 20 spectra per sample. Protocols for spectral processing, library spectra creation and statistical analysis have also been described in a previous chapters. The average spectrum for each sample was put into a library and compared using linear correlation. The treated samples were compared with previously created reference libraries containing 14 species to determine if the different exposures affected the identification. Visual observation was necessary to understand the quantitative and qualitative changes in the protein profiles caused by the various treatments and exposures.

Bacterial Strains

Table 7-1 contains a list of the bacteria, source, and description of strains and known resistances used in this study. B. subtilis 168 was obtained from Dr. Wayne

Nicholson at the University of Florida. All the sporulation mutants were provided by Dr. Adam Driks at Loyola University.

Table 7-1. List and description of strains used in this study.

Species	Strain #	Source[a]	Comments
B. pumilus	ATCC 7061^T	ATCC	Type strain
B. pumilus	FO-036b	JPL-SAF	Clean room air particulate, H2O2 resistant
B. subtilis	168	Nicholson	Genome fully sequenced
B. subtilis	PY79 (ADL18)	Driks	"wild type" parent strain for mutants
B. subtilis	AGS232 (ADL392)	Driks	blocks sporulation very early around stage I
B. subtilis	PM806 (ADL201)	Driks	blocks sporulation at stageII
B. subtilis	SC500 (ADL40)	Driks	blocks sporulation at stage III
B. subtilis	SAB50 (ADL58)	Driks	blocks sporulation at stage IV
B. subtilis	PE241 (ADL956)	Driks	blocks cortex synthesis
B. subtilis	AD17 (ADL56)	Driks	prevents some mother cell gene expression and formation of the inner coat to a large degree
B. subtilis	AD28 (ADL77)	Driks	prevents outer coat assembly
B. subtilis	AD142 (ADL57)	Driks	prevents core assembly and some mother cell gene expression but cortex is largely normal

[a]Abbreviations: JPL, Jet Propulsion Laboratory, SAF, Spacecraft Assembly Facility, ATCC, American Type Culture Collection

Standard Sporulation in Liquid Media

The standard protocol for the production of spores in this study followed the procedure outlined by Nicholson and Setlow7 and has been described previously.73 Briefly, a single purified colony was inoculated into nutrient broth sporulation medium (NSM), and incubated at 32oC with shaking for ca. 2-4 days until the cultures reached >99% spores.

Sporulation on Solid Media

A second protocol utilized in this study involved streaking a tryptic soy agar plate with purified cells of the strain to be sporulated. The plate was incubated at 32oC for 12 hours and was then kept at room temperature. The center portions of the colonies on the plate were periodically checked with phase contrast microscopy for the appearance of

spores. After approximately 2 weeks of storage at room temperature, the colonies reached >90% spores. A sterile loop was used to scrape and remove the middle portion of the colonies on the plate. The scrapings were resuspended in 5 mL of sterile water until a milky solution was obtained.

Spore Purification

Spore cultures were harvested by centrifugation and purified to remove remnant vegetative cells and cellular debris using either the water washing or the lysozyme treatment with salt and detergent washing.7 Purified spores were adjusted to an optical density of 0.6 at 600 nm and were suspended in sterile deionized water and stored at 4oC in glass vials until exposure and/or analysis.

Storage Conditions and Aging

Freshly prepared B. subtilis 168, B. pumilus 7061, and FO-36b spores were compared with spores that had been stored in sterile water at 4°C for 1 year. Additionally 25 µL from the stock solution of B. subtilis 168, B. pumilus 7061, and FO-36b was centrifuged down at 9,600 x g for 10 minutes and the supernatant solution removed. The remaining spore pellets were lyophilized, desiccated, or resuspended in sterile water and stored for 1 week under these conditions. The lyophilized and desiccated samples were reconstituted in sterile water and the resulting solutions were diluted 1:10 with the matrix solution for MALDI analysis.

Radiation Exposure

Radiation dosimetry was performed using a cobalt 60 source in an ion chamber. B. subtilis 168, B. pumilus 7061, and FO-36b spores were exposed to gamma radiation at 1 MRad (50 rad/sec for 330 minutes) and 0.5 MRad (25 rad/sec for 330 minutes). Prior to gamma exposure, the concentration of the spores was adjusted to 108 cells/mL and 1mL

of this dilution was placed in a 1.5 mL glass vial for exposure. Controls (unexposed spores) were also prepared and stored under the same conditions. For the MALDI analysis, the solutions were removed and spun down at 9,600 x g for 10 minutes. The supernatant was discarded and a 100 µL aliquot of the MALDI matrix solution was used to resuspend the pellet. Survival of the spores after the radiation treatment was quantitatively determined by growing the gamma irradiated samples in tryptic soy broth at 32°C.

B. subtilis 168, B. pumilus 7061, and FO-36b spores were exposed to UVC irradiation at 254nm. Prior to exposure, the concentration of the spores was adjusted to 106 cells/mL using phosphate buffered saline. Spores were placed into an uncovered Petri dish and were exposed to UVC radiation at 254 nm using a low pressure hand held mercury arc UV lamp (UV Products Inc., model #UGV-11, Upland, CA). Spores were exposed for 169.8 seconds, the time needed to produce 1 kJ of energy at the sample surface based on measurements using a radiometer. Control (unexposed) spores were also diluted and stored under the same conditions. One mL of the exposed and control solutions were centrifuged at 9,600 x g for 10 minutes. The supernatant was discarded. No pellet was observed in the bottom of the microfuge tubes, however 10 µL of the MALDI matrix solution was added to the tubes and was used for analysis.

Hydrogen Peroxide Exposure

B. subtilis 168, B. pumilus 7061, and FO-36b spores were exposed to hydrogen peroxide vapor sterilization. A vial containing 1mL of a desiccated solution of 107 cells/mL of each spore strain was place an appropriate sealed bag. The bags were placed inside a Sterrad 100 vapor H2O2 chamber (Advanced Sterilization Products, Irvine, CA). The vials were exposed to 2 cycles of H2O2 injections. Each injection

cycle resulted in the exposure of the spores to 3 mg/mL vapor H2O2. Controls consisted of desiccated spores in vials not exposed to the vapor treatment. To recover the spores from the vial, 1 mL of water was added and the vial was vortexed for 1 hour. After 1 hour of shaking, the solution was centrifuged at 9,600 x g for 10 minutes. The supernatant was discarded. Again, no pellet was observed but 10 µL of MALDI matrix solution was added to the tubes for analysis.

B. subtilis 168, B. pumilus 7061, and FO-36b spores were exposed to a 5% liquid hydrogen peroxide treatment. An 833 µL aliquot of a 107 cells/mL culture was combined with 167 µL of 30% hydrogen peroxide which resulted in a final concentration of 5% hydrogen peroxide. Controls were prepared by adding 167 µL of water to the spores. The solutions were incubated for 1 hour at room temperature (25°C) with gentle mixing. After 1 hour the treated solutions were centrifuged and the supernatant solution was discarded. A 10 µL aliquot of the MALDI matrix solution was added to the remaining spore pellets and the solution was deposited directly onto the MALDI plate.

Autoclave Exposure

Aliquots (100 µLs) of and both freshly prepared and 1 year old B. subtilis 168, B. pumilus 7061, and FO-36b spores were autoclaved for 30 minutes at 15 psi and 220°C. No dilutions were performed on the original OD 0.6 spore solutions prior to autoclaving. For MALDI analysis, 2.5 µL of the autoclaved sample was combined with 22.5 µL of the MALDI matrix solution.

Preparation of *B. subtilis* Sporulation Mutants

Sporulation mutants were incubated in Luria Bertani (LB) media until turbid. A 10 µL aliquot of the culture was then inoculated onto LB agar plates containing antibiotics to isolate cells with the mutation of interest. SAB50, AD142, and AD28 were

grown on LB containing 5 μg/mL choloramphenicol; AGS232 was grown with 5 μg/mL choloramphenicol and 50 μg/mL spectinomycin and PE241 was grown on 5 μg/mL tetracycline. PM806, SC500, and AD17 had no drug resistance markers and were grown on LB with no drug. Single isolated colonies from the LB plates were used to prepare glycerol stocks. The glycerol stocks were then used to streak tryptic soy agar (TSA) plates for sporulation. The sporulation was prepared as described by Driks.117-119

It was necessary to harvest the mutants with different blocks in sporulation at different times to prevent lysing of the cells in solution. The mutants were harvested shortly after the time at which the blocks occurred. AGS232 was harvested at T2 (hour 2 of sporulation), PM806 at T3, SC500 at T4, and SAB50 and PE241 were harvested at T5. The cell pellets (consisting mostly of vegetative cells and protoplasts) were collected by centrifugation and were used for MALDI analysis with no further purification. For the remaining strains, including the parent strain PY79, AD28, AD17, and AD142, a 1 mL aliquot was taken at T6 and then the culture was completely harvested at T24. The T6 samples consisting of vegetative looking cells were pelletized and were used with no further purification. The T24 samples were further purified using a Renografin gradient.7 All of the cell pellets were resuspended in 50 μL of sterile water and were diluted 1:10 with the matrix solution for MALDI analysis.

Results and Discussion

Initial Sporulation and Purification Conditions

To date, standardized protocols for the laboratory culture of vegetative cells have been necessary for obtaining reproducible MALDI-TOFMS protein profiles. However, very few studies have been done on the reproducibility of spores produced under different initial sporulation conditions. These limited studies have examined liquid

sporulation media formulations and have shown that the same protein peaks are generally present across the samples.26 No studies have examined spores grown on solid media or with media that is not chemically altered to enhance sporulation.

In the environment, spores are not likely to have the abundance of nutrients and ions available in broth cultures and will certainly not have a controlled temperature of incubation. To pseudo-mimic this type of formation, bacterial cultures were streaked on regular TSA plates. Following an initial incubation period to start colony formation, the plates were left at room temperature and checked for sporulation. After approximately 2 weeks, the center of the colonies from all the strains contained mainly sporulated cells which were removed for analysis. Spores of B. subtilis 168, B. pumilus 7061, and FO36b produced in this manner had linear correlation values of 0.96, 0.87, and 0.92, respectively, with spectra from the same species contained in the reference library.

In addition to the initial sporulation conditions, there are several methods for the purification of spore crops. The lysozyme treatment with salt and detergent washes or Renografin gradients are typically rather than plain water washing to ensure the removal of mother cell components adsorbed to the forespore. These purification methods are not likely to be available in the environment and so an examination of the protein profiles obtained under different purification methods is warranted. Correlation values among B. subtilis spores purified in water, lysozyme, or Renografin were above 0.90 indicating there was no difference in the profiles. The MALDI-TOFMS protein profiling method proved to be robust enough to handle spores produced under different sporulation conditions and purification methods.

Storage Conditions and Spore Aging

The conditions and times for which spores are stored may also impact the protein profile of the sample. Storage in water is recommended as it does not present a radical change in the spore environment. Long term storage in water can be complicated by spore germination in response to small amounts of nutrients leaked by spores and by possible hydrolysis or degradation of spore coat components.7 Therefore, lyophilization is recommended for long-term storage. In the case of a bioagent release, spores will have been desiccated to maximize dispersal and obtain the optimal particle size for inhalation.

Two spore crops were examined: one that was prepared 1 year ago and stored in water and the second which was less than 1 month old. Three storage conditions were also examined: storage in water at 4°C, desiccated spores, and lyophilized spores. Storage conditions were examined for short term effects after 1 week of storage under the different conditions.

Table 7-2 shows the correlation results for B. subtilis 168, B. pumilus 7061, and FO-36b spores under different storage conditions and different ages. The impact of the storage conditions on the MALDI protein profiles was minimal; however, the effect of aging for some of the spores was significant. For B. subtilis 168 and B. pumilus 7061 spores, aging had only a minimal effect on the protein profiles. The peaks in the range from 3-10 kDa were stable over time. For proteins above 10kDa, there was a loss in sensitivity noticed but it had minimal impact on the correlation values, with values of 0.90 and 0.84 for the different aged spores of B. subtilis and B. pumilus. FO-36b showed a significant difference in its protein profile even in the 3-10 kDa range and had a low correlation value of 0.56 with the older spores. Figure 7-1 shows the difference in the protein profiles for the 1 month old and 1 year old FO-36b spore crops. There is an

appearance of intense biomarkers at 6,824, 7,240, and 9,620 Da in spores stored in water over long periods of time. In examining spectra taken over a 1 year period, the 7,630 Da peak was initially the most intense peak in the spectra with a small 7,240 Da peak. The relative intensity of the 7,240 Da peak increased compared to the 7,630 peak over the course of the year. Although the species and strain specific biomarkers identified previously were still present, the change in the relative intensities of these peaks would impact the correlation analysis. It is unknown whether these biomarkers are a result of the hydrolysis of spore coat proteins or are indicative of a change in the permeability of the spore coat.

Table 7-2. Correlation values for aged spores and spores stored under different conditions.

Species	Age	Storage	*B. pumulis* 7061	*B. pumulis* 7061	*B. pumulis* 7061	*B. pumulis* 7061	*B. subtilis* 168	*B. subtilis* 168	*B. subtilis* 168	*B. subtilis* 168	FO36b	FO36b	FO36b	FO36b
			1 year	1 month	1 month	1 month	1 year	1 month	1 month	1 month	1 year	1 month	1 month	1 month
			control	control	lyophilized	dessicated	control	control	lyophilized	dessicated	control	control	lyophilized	dessicated
B. pumulis 7061	1 year	water	-											
B. pumulis 7061	1 month	water	0.84	-										
B. pumulis 7061	1 month	lyophilized	0.78	0.83	-									
B. pumulis 7061	1 month	dessicated	0.75	0.92	0.96	-								
B. subtilis 168	1 year	water	0.04	0.05	0.04	0.05	-							
B. subtilis 168	1 month	water	0.04	0.05	0.05	0.05	0.90	-						
B. subtilis 168	1 month	lyophilized	0.02	0.03	0.02	0.02	0.94	0.80	-					
B. subtilis 168	1 month	dessicated	0.02	0.03	0.02	0.02	0.92	0.85	0.97	-				
FO36b	1 year	water	0.09	0.08	0.07	0.06	0.14	0.15	0.15	0.16	-			
FO36b	1 month	water	0.68	0.74	0.72	0.74	0.14	0.16	0.11	0.12	0.56	-		
FO36b	1 month	lyophilized	0.70	0.71	0.65	0.70	0.16	0.17	0.12	0.14	0.62	0.98	-	
FO36b	1 month	dessicated	0.64	0.64	0.71	0.70	0.15	0.16	0.12	0.13	0.56	0.94	0.93	-

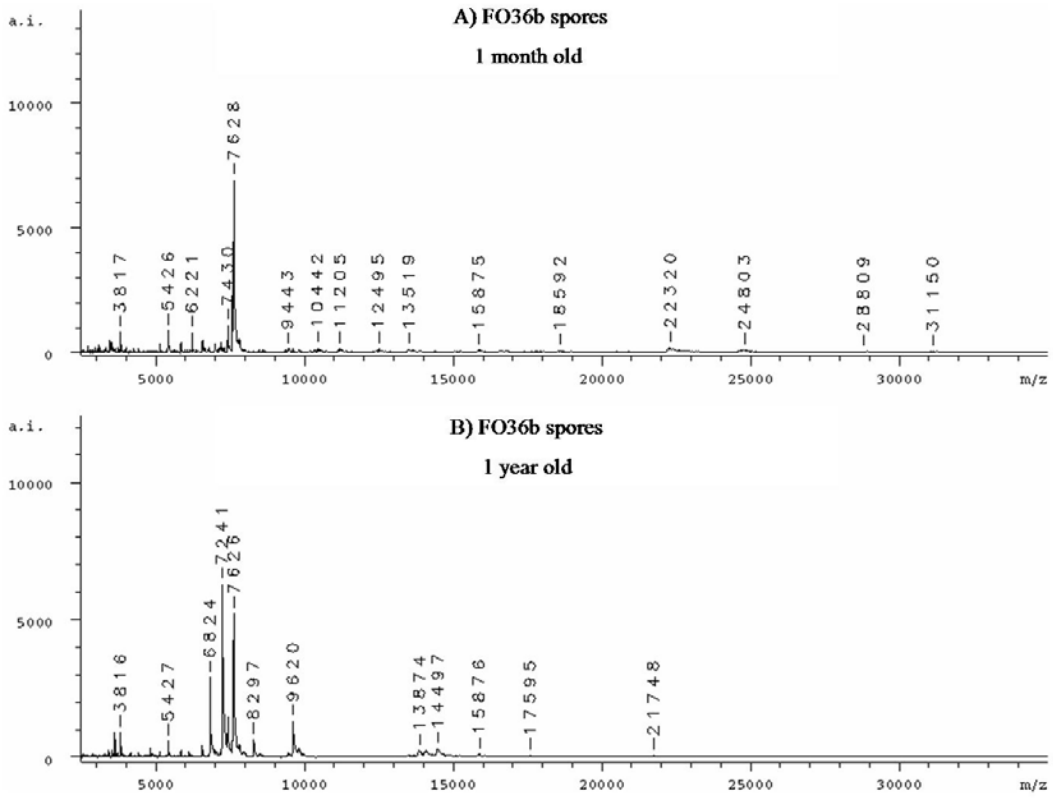

Figure 7-1. MALDI-TOFMS protein profiles of aged FO36b spores. A) FO36b spores that are stored in water for 1 month. B) FO36b spores that were stored in water for 1 year. The mass range is shown from *m/z* 2,500-40,000.

For the same age spores of each strain that had been lyophilized or desiccated, correlation values remained above 0.80, indicating a high degree of similarity. Upon closer observation, it was noted that the greatest impact of the different storage conditions were on the higher molecular weight proteins. Figure 7-2 shows the average spectra for the B. subtilis 168 spores that are 1 month old under the different storage conditions. Figure 7-3 is the average spectra for the FO36b spores that are 1 year old. The lyophilized spores overall had a lower signal to noise ratio and a loss in sensitivity was seen for the higher molecular weight proteins. This impact was more significant for lyophilized spores than desiccated spores and was more evident in freshly prepared spores than in older preparations where a loss in sensitivity for the higher molecular weight proteins was already present. Other freshly prepared spore strains from other species gave similar results for the lyophilized and desiccated spores (data not shown).

Radiation Exposure

Gamma and UV radiation are sterilants for most microbial organisms. Spores possess a unique ability to resist and/or repair the damage caused by ionizing radiation, which allows them to persist in the environment and be immune to the use of these technologies for sterilization. The level of resistance to radiation is strain-specific. All of the spore strains in this study had elevated resistances to UVC radiation. The LD90 for B. subtilis 168, B. pumilus 7061, and FO36 b are 200 J/m2, <200 J/m2, and 900 J/m2 respectively (K. Venkateswaran and D. Newcombe, personal communication). All of the spore strains were resistant to gamma irradiation at 0.5 MRad and both B. subtilis 168 and FO36b were resistant to 1MRad dose of gamma irradiation. The MALDI-TOFMS

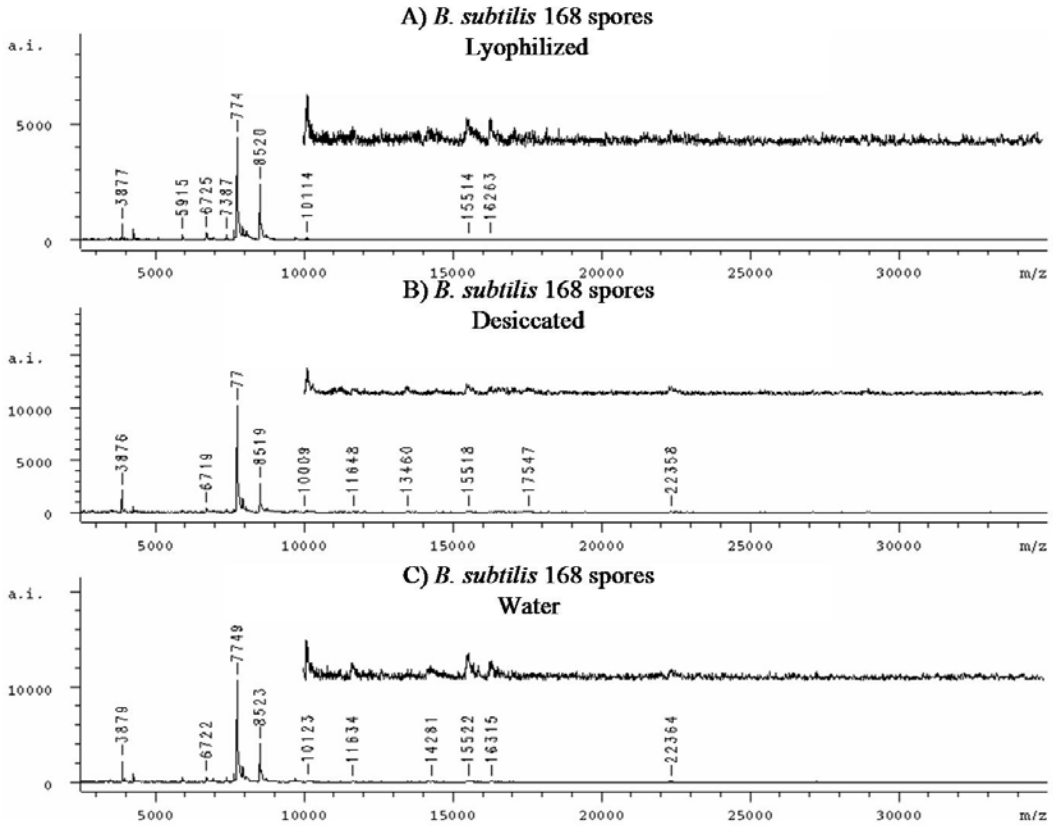

Figure 7-2. MALDI-TOFMS protein profiles of *B. subtilis* 168 spores under different storage conditions. A) Lyophilized spores. B) Desiccated spores. C) Water storage. The mass range is shown from *m/z* 2,500-40,000. The higher molecular mass region from *m/z* 10,000-40,000 is amplified 4x (see inset of each spectrum) in order to visualize the higher molecular weight peaks.

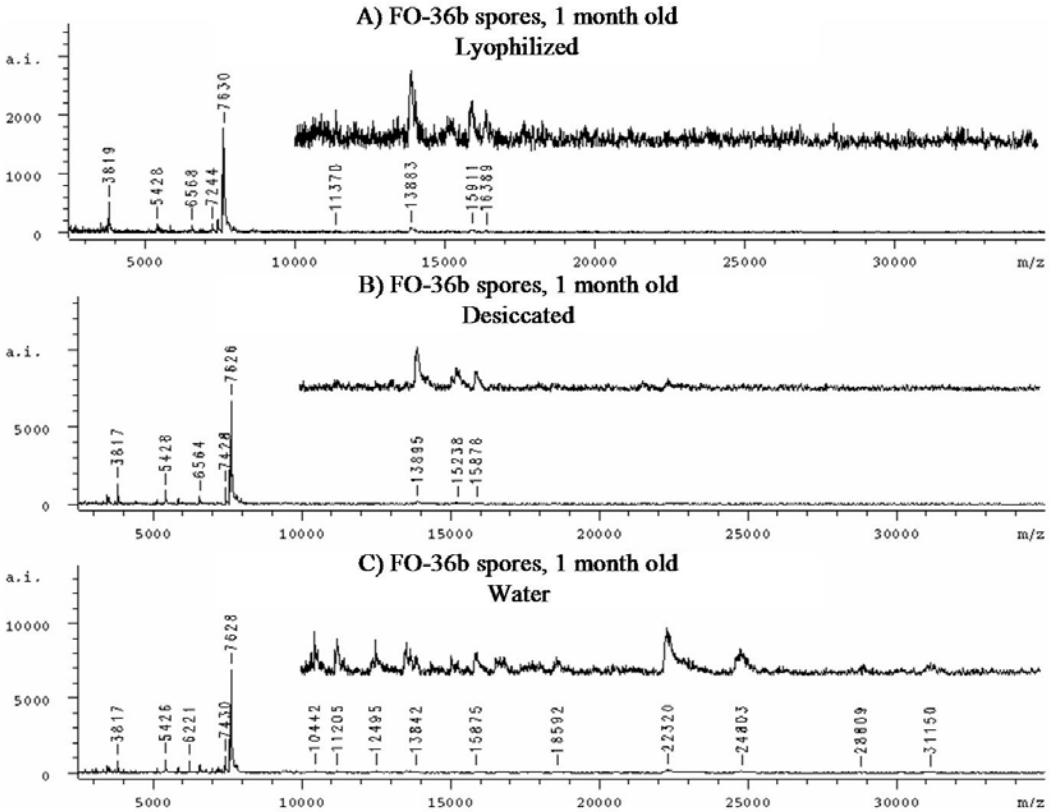

Figure 7-3. MALDI-TOFMS protein profiles of 1 month old FO36b spores under different storage conditions. A) Lyophilized spores. B) Desiccated spores. C) Water storage. The mass range is shown from *m/z* 2,500-40,000. The higher molecular mass region from *m/z* 10,000-40,000 is amplified 4x (see inset of each spectrum) in order to visualize the higher molecular weight peaks.

protein profiles of spores exposed to both irradiation sources were examined using linear correlation.

Gamma exposed spores from all three species were nearly identical to the control and reference strains. For the 0.5 and 1 MRad exposure, correlation values (r) for B. subtilis 168 were 0.96 and 0.94, for B. pumilus 7061 r = 0.98 and 0.87, and for FO36b, r = 0.94 and 0.88 with the control strains. UV exposed spores of B. pumilus 7061 and FO36b were also not different from controls with correlation values of 0.90 and 0.94 respectively; however, an increase in intensity was noted in the spectra from treated spores. Profiles of B. subtilis 168 spores were significantly altered by the UV treatment. Figure 7-4 shows both the treated and control spectra from B. subtilis 168. There was a drop in overall sensitivity and there was a large peak at 5, 387 Da which had not been previously observed in B. subtilis spores.

H_2O_2 Exposure

While intensely oxidizing conditions are not prevalent in the environment, H2O2 vapor sterilization is commonly used for the treatment of spacecraft parts and other heat sensitive medical equipment and devices. Of the three spore strains in this study, FO-36b was found to be resistant to both liquid and vapor H2O2 treatments. MALDI-TOFMS protein profiling of the vapor treated spores was hindered by the ability to efficiently recover the treated, desiccated spores from the inside of the vials used in the analysis. This combined with the molecular weight shifts caused by the oxidation of amino acids in the proteins resulted in very low correlation values when compared with reference and control strains. Visual interpretation of the spectra was necessary to interpret spectral differences.

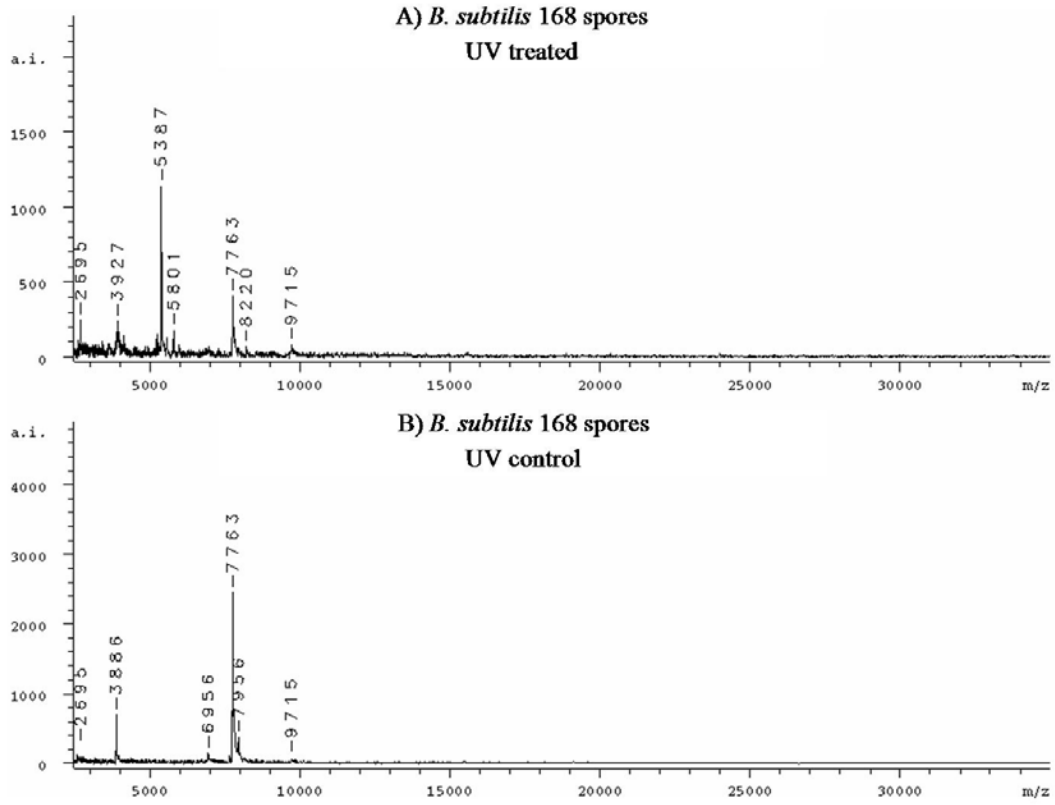

Figure 7-4. MALDI-TOFMS protein profiles of UV treated *B. subtilis* 168 spores. A) UV exposed spores. B) Unexposed (control) spores. The mass range is shown from *m/z* 2,500-40,000.

Average spectra from the control and H2O2 vapor-treated spores of B. subtilis 168, B. pumilus 7061, and FO36b are shown in Figure 7-5, 7-6, and 7-7. Oxidation of the protein peaks is readily evident in the treated spectra where shifts in molecular weight between 60-100 Da are seen for the base peaks in the spectra. These shifts are combined with a decrease in peak resolution for the oxidized samples. The B. subtilis 168 and B. pumilus 7061 spectra also have an increase in intensity (and accompanying peak shift) for the 6,940 Da peak in B. subtilis and the 6,870 Da peak in B. pumilus. In FO36b, there is a decrease in intensity of the 6,830 Da peak and the 7,630 peak is missing from the treated spectra.

Since the recovery of spores from the vial could have contributed to changes in the resulting protein profiles, a second H2O2 treatment, using a 5% liquid exposure directly on 1 month old spores was performed. Average spectra from the control and treated spectra from the liquid exposure are shown in Figure 7-8, 7-9, and 7-10. Treated spectra from B. subtilis 168 and B. pumilus 7061 again showed a shift in molecular weight of the base peak in the spectra. Interestingly, spores of FO36b showed no corresponding shift in molecular weight for the intensity 7,630 peak. This may have been a result of the protein not having oxidizable residues or the presence of a catalase protein in FO36b that quenches oxidizing species. The oxidation of residues in the proteins could alter the protein structure to the point that it would lose activity or function and inhibit germination.

Autoclave Exposure

Autoclaving at high temperatures and high pressure is the most reliable method to inactivate spore strains. The 1 month old spores from the three strains were autoclaved.

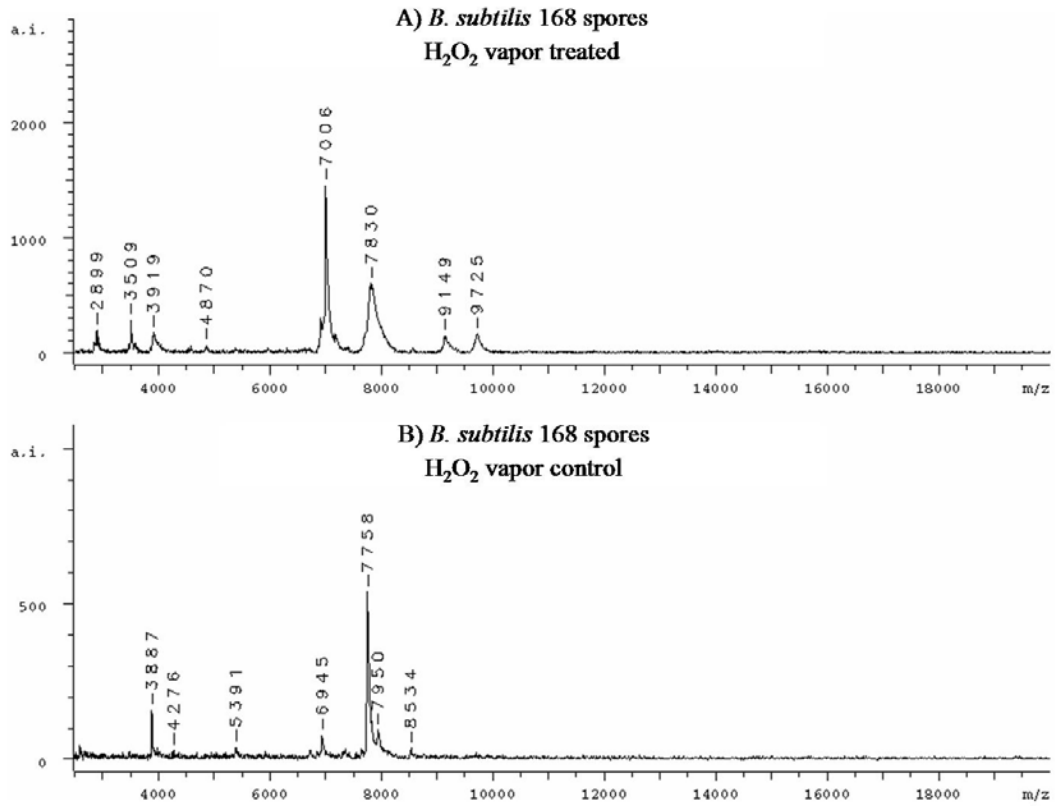

Figure 7-5. MALDI-TOFMS protein profiles of H_2O_2 vapor treated *B. subtilis* 168 spores. A) H_2O_2 exposed spores. B) Unexposed (control) spores. The mass range is shown from *m/z* 2,500-40,000.

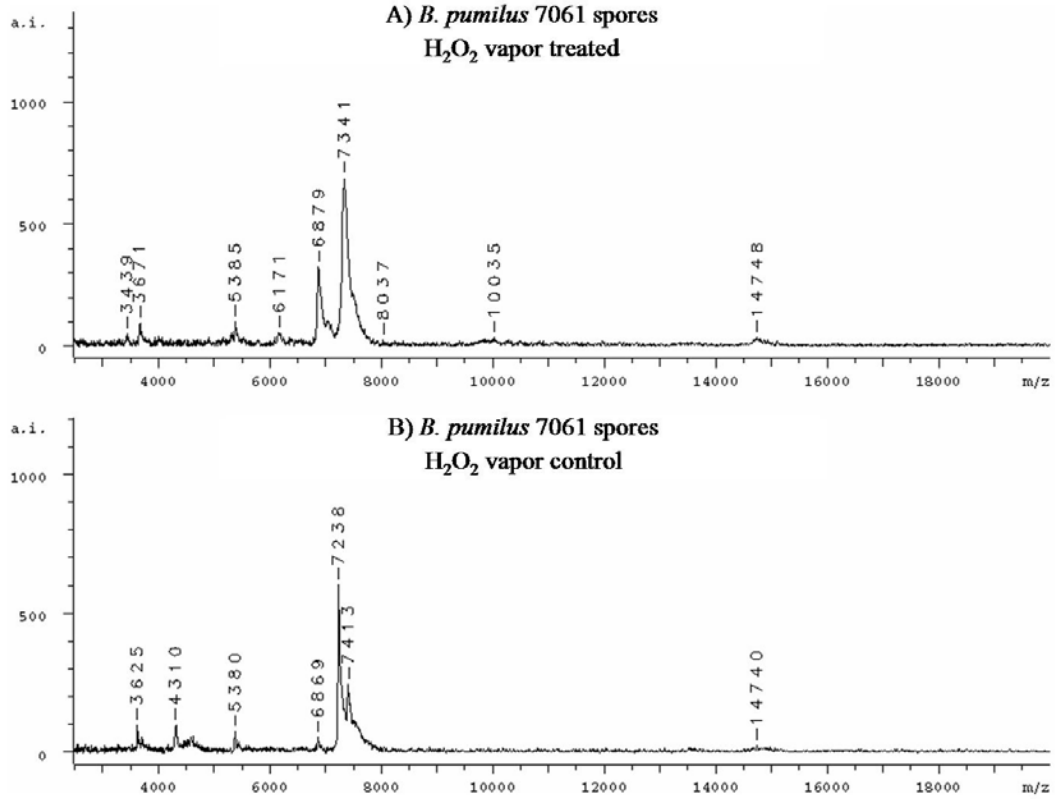

Figure 7-6. MALDI-TOFMS protein profiles of H_2O_2 vapor treated *B. pumilus* 7061 spores. A) H_2O_2 exposed spores. B) Unexposed (control) spores. The mass range is shown from *m/z* 2,500-40,000.

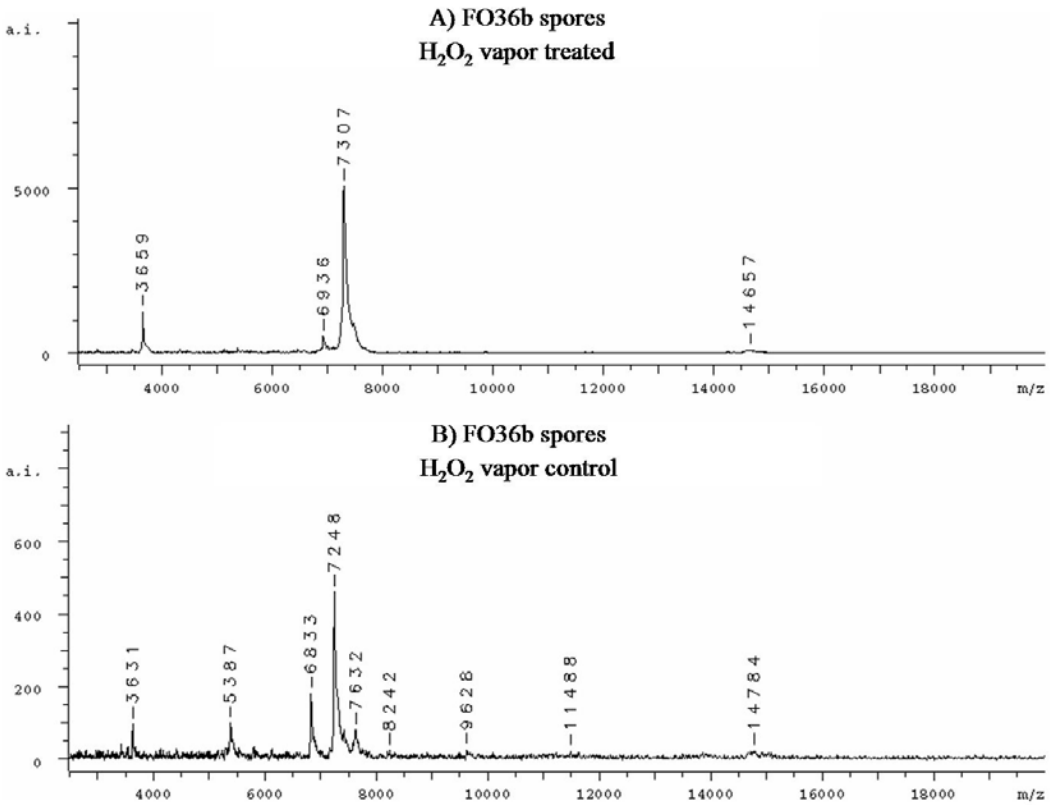

Figure 7-7. MALDI-TOFMS protein profiles of H_2O_2 vapor treated FO36b spores. A) H_2O_2 exposed spores. B) Unexposed (control) spores. The mass range is shown from *m/z* 2,500-40,000.

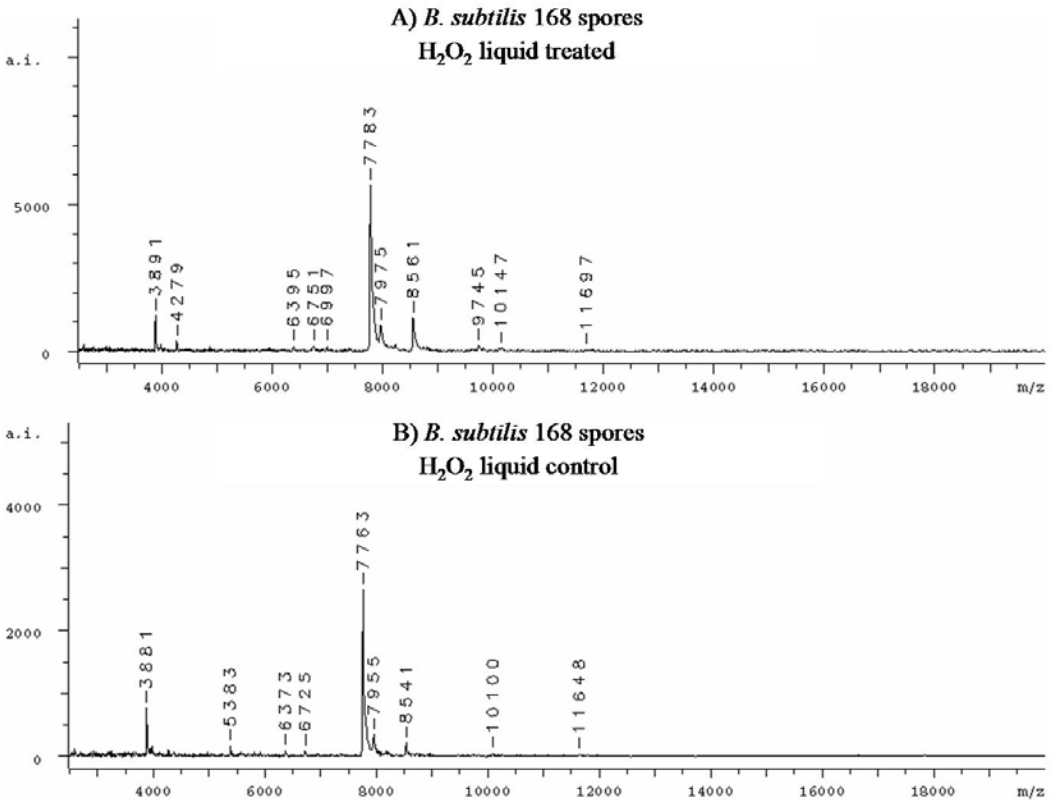

Figure 7-8. MALDI-TOFMS protein profiles of H_2O_2 liquid treated *B. subtilis* 168 spores. A) H_2O_2 exposed spores. B) Unexposed (control) spores. The mass range is shown from *m/z* 2,500-40,000.

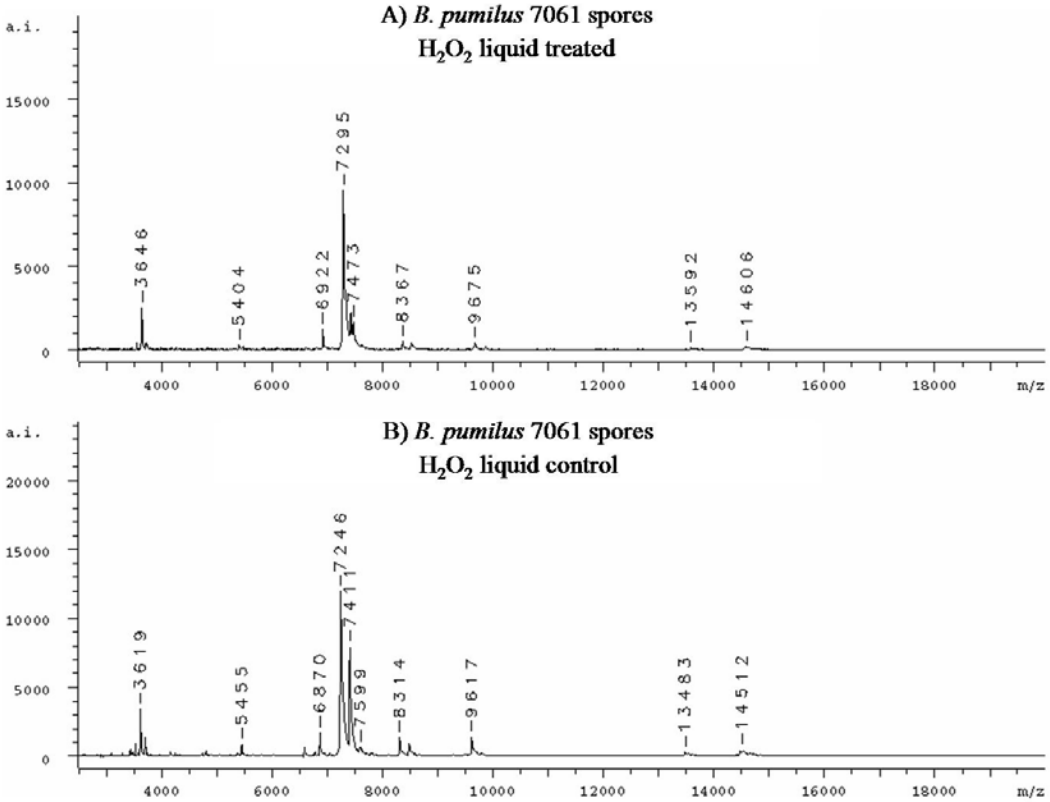

Figure 7-9. MALDI-TOFMS protein profiles of H_2O_2 liquid treated *B. pumilus* 7061 spores. A) H_2O_2 exposed spores. B) Unexposed (control) spores. The mass range is shown from *m/z* 2,500-40,000.

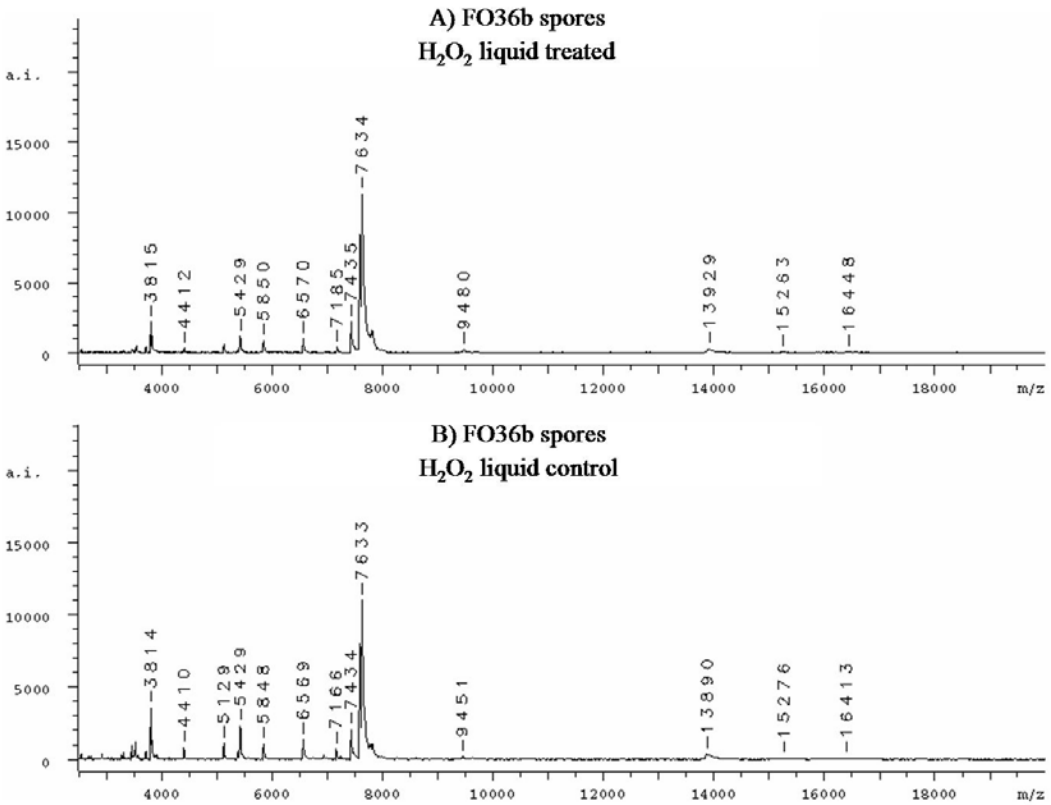

Figure 7-10. MALDI-TOFMS protein profiles of H_2O_2 liquid treated FO36b spores that are 1 month old. A) H_2O_2 exposed spores. B) Unexposed (control) spores. The mass range is shown from *m/z* 2,500-40,000.

None of the spore strains were viable post-autoclaving. The average spectra from the control and autoclaved spores are shown in Figure 7-11, 7-12, and 7-13. Surprisingly, the spectra from B. subtilis 168 and B. pumilus 7061 are very similar to the control spectra as evident by the high correlation values of 0.86 and 0.85 respectively. Change in the relative intensity of the m/z 6,940 and 9,136 peaks are the most predominant changes in the B. subtilis pair. In B. pumilus 7061, there is not an obvious change in the spectra other than a change in overall signal-to-noise and the disappearance of the higher molecular weight peaks in the autoclaved samples. As Figure 7-13 shows, the FO36b sample had significant changes that caused the correlation value between the spectra to be 0.20. The autoclaved spores showed a similar spectral pattern to the aged FO36b samples with the appearance of intense biomarkers at 6,830 and 7,240, which are not observed in the fresh FO36b control samples.

Sporulation Mutants

In the environment or in a spore sample, it is possible that the sporulation process may start but not go through to completion. Accordingly, examining sporulation mutants may give us insight into the sporulation process and the cascade of gene and protein expression that occurs during spore formation. In this study, the MALDI-TOFMS protein profiles were obtained for sporulation mutants blocked at different stages of sporulation and for mutants which resulted in incomplete formation of the spore coat. Table 4-3 contains the results from the linear correlation analysis of these strains.

Mutants which resulted in incomplete spore formation (ADL392, ADL201, ADL40, ADL58, and ADL956) resulted in spectra (Figure 7-14 A-E) which were drastically different from both the T6 (Figure 7-15 A) and T24 (Figure 7-16 A) samples from the parent strain (ADL18). Average spectra from these strains are shown in time

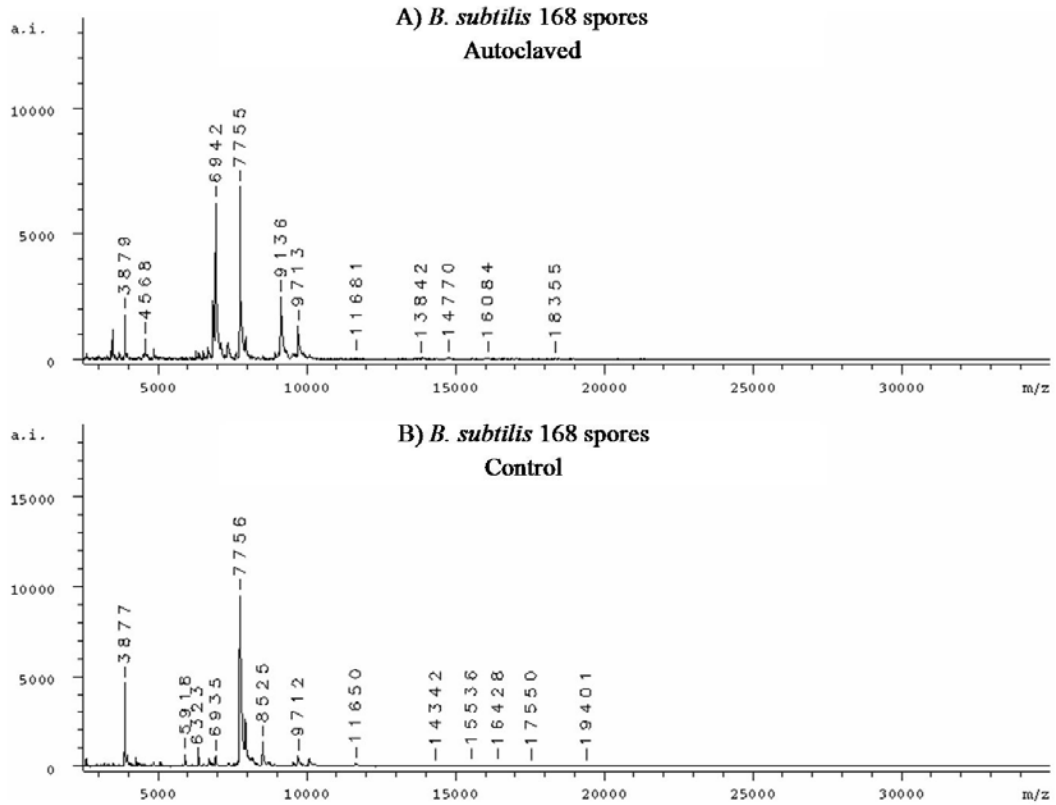

Figure 7-11. MALDI-TOFMS protein profiles of autoclaved *B. subtilis* 168 spores. A) Autoclaved spores. B) Control spores. The mass range is from *m/z* 2,500-40,000.

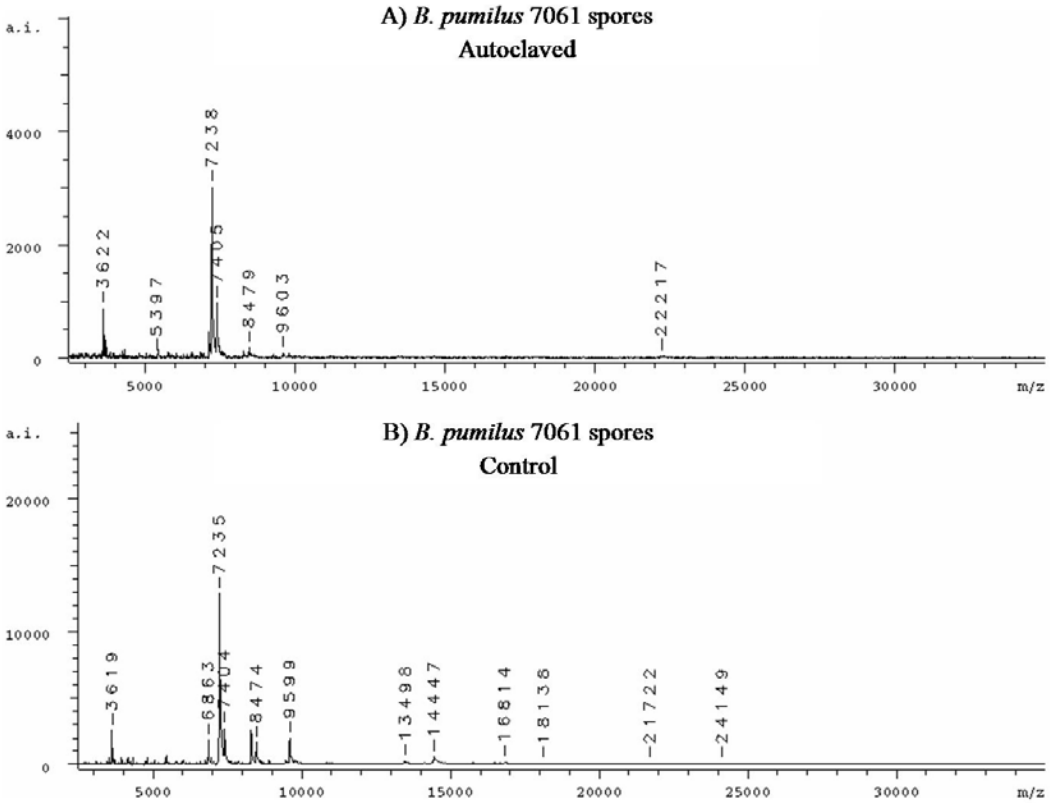

Figure 7-12. MALDI-TOFMS protein profiles of autoclaved *B. pumilus* 7061 spores. A) Autoclaved spores. B) Control spores. The mass range is from *m/z* 2,500-40,000.

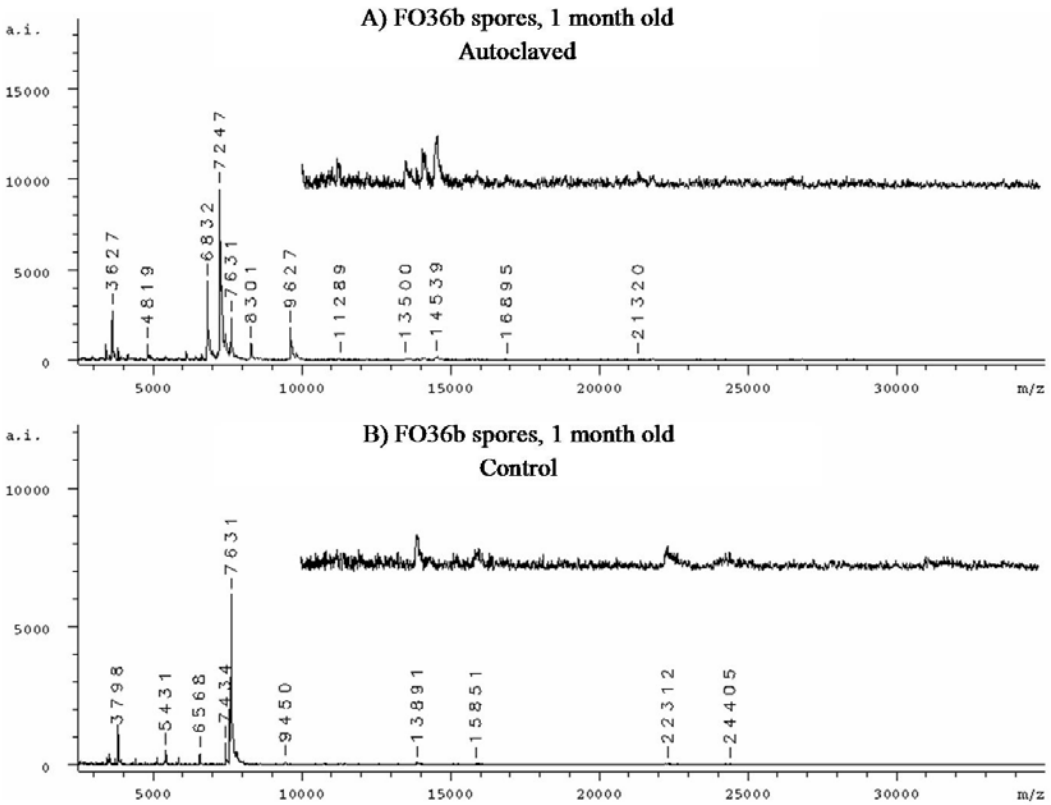

Figure 7-13. MALDI-TOFMS protein profiles of autoclaved FO36b spores. A) One month old autoclaved spores. B) One month old control spores. The mass range is from *m/z* 2,500-40,000.

sequence in Figure 7-14 A-E. The 5 mutants were markedly similar to each other even though they were sampled at different time points and sporulation was blocked at different stages. Prominent biomarker peaks for these mutants were found at m/z 7,300, 9,870, 16,400, 17,700, 22,380, and 32, 630. In the late blocks, ADL58 and ADL956, additional biomarker peaks appear at 6,930 Da and 12.6 kDa.

Table 7-3. Correlation values for mutated *B. subtilis* strains

Strain	Time	Description	AGS232 (ADL392)	PM806 (ADL201)	SC500 (ADL40)	SAB50 (ADL58)	PE241 (ADL956)	PY79 (ADL18)	AD17 (ADL56)	AD28 (ADL77)	AD142 (ADL57)	PY79 (ADL18)	AD17 (ADL56)	AD28 (ADL77)	AD142 (ADL57)
AGS232 (ADL392)	2	blocks sporulation very early around stage I	-												
PM806 (ADL201)	3	blocks sporulation at stageII	0.82	-											
SC500 (ADL40)	4	blocks sporulation at stage III	0.84	0.85	-										
SAB50 (ADL58)	5	blocks sporulation at stage IV	0.55	0.66	0.63	-									
PE241 (ADL956)	5	blocks cortex synthesis	0.77	0.94	0.84	0.76	-								
PY79 (ADL18)	6	"wild type" parent strain for mutants	0.41	0.56	0.56	0.69	0.59	-							
AD17 (ADL56)	6	prevents some mother cell gene expression and formation of the inner coat to a large degree	0.30	0.40	0.38	0.44	0.52	0.50	-						
AD28 (ADL77)	6	prevents outer coat assembly	0.20	0.27	0.29	0.64	0.40	0.60	0.78	-					
AD142 (ADL57)	6	prevents core assembly and some mother cell gene expression but cortex is largely normal	0.36	0.50	0.48	0.51	0.61	0.55	0.98	0.75	-				
PY79 (ADL18)	24	"wild type" parent strain for mutants	0.00	0.05	0.00	-0.01	0.07	0.35	0.01	0.00	0.01	-			
AD17 (ADL56)	24	prevents some mother cell gene expression and formation of the inner coat to a large degree	0.11	0.12	0.07	0.10	0.14	0.32	0.26	0.22	0.25	0.57	-		
AD28 (ADL77)	24	prevents outer coat assembly	0.06	0.08	0.03	0.21	0.13	0.36	0.29	0.35	0.26	0.41	0.72	-	
AD142 (ADL57)	24	prevents core assembly and some mother cell gene expression but cortex is largely normal	0.08	0.13	0.08	0.08	0.17	0.26	0.43	0.29	0.41	0.38	0.89	0.54	-

For the mutants which continued through the complete sporulation cycle, samples were analyzed at T6 and T24. At T6, the spores should be entering stage V where they would begin formation of the spore coat. The average spectrum for each of the strains at T6 is shown in Figure 7-15 B-D. Spectra from the mutant strains (ADL56, ADL77, and ADL57) at T6 were more similar to each other than to the parent strain (Figure 7-15 A). The mutant strains and the parent strain have several biomarkers in common with the blocked mutants including m/z 6,930 (with a variation in relative intensity between the 2

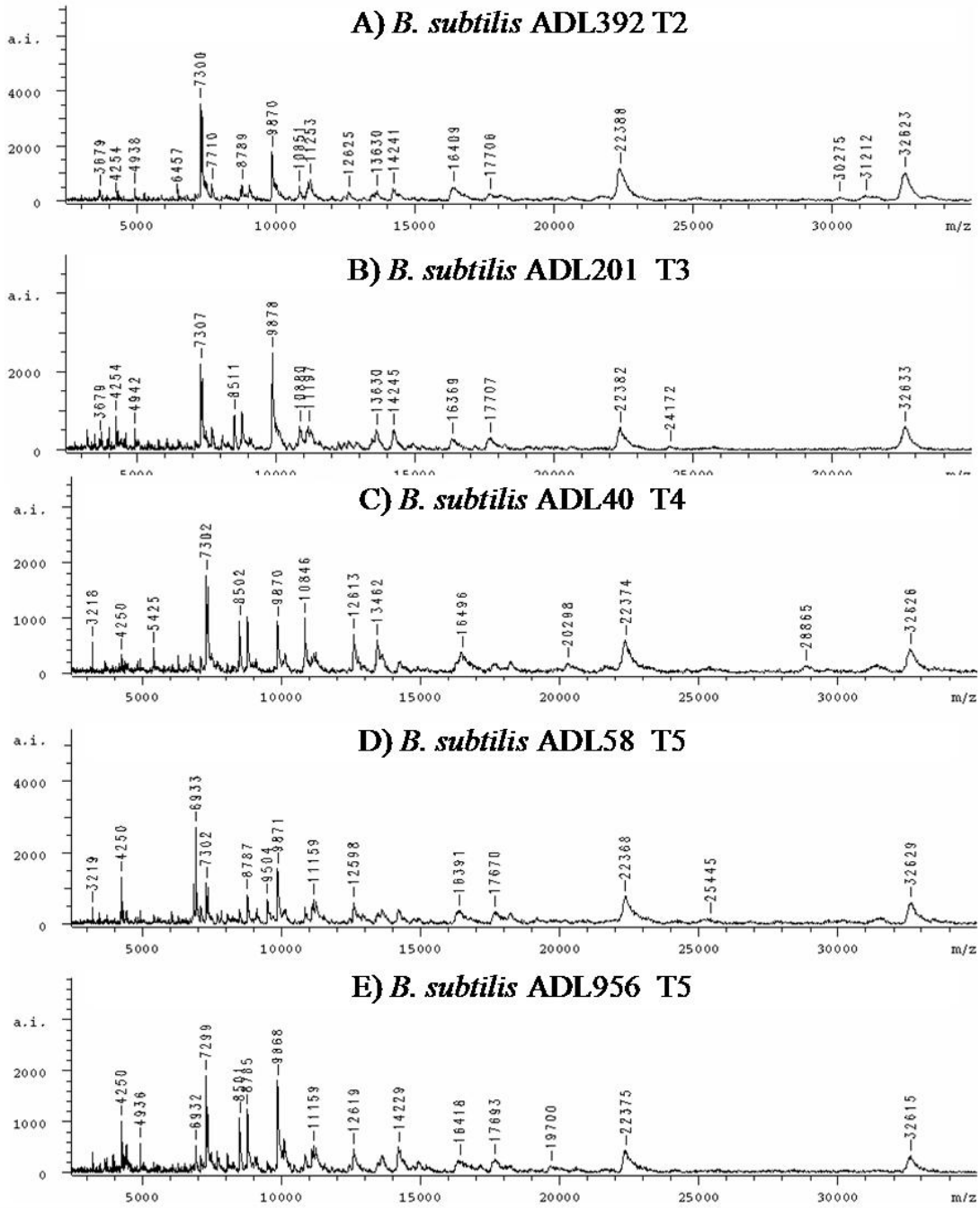

Figure 7-14. MALDI-TOFMS protein profiles of *B. subtilis* sporulation mutants. A) ADL392 at T2. B) ADL201 at T3. C) ADL40 at T4. D) ADL58 at T5. E) ADL956 at T5. The mass range is from *m/z* 2,500-40,000.

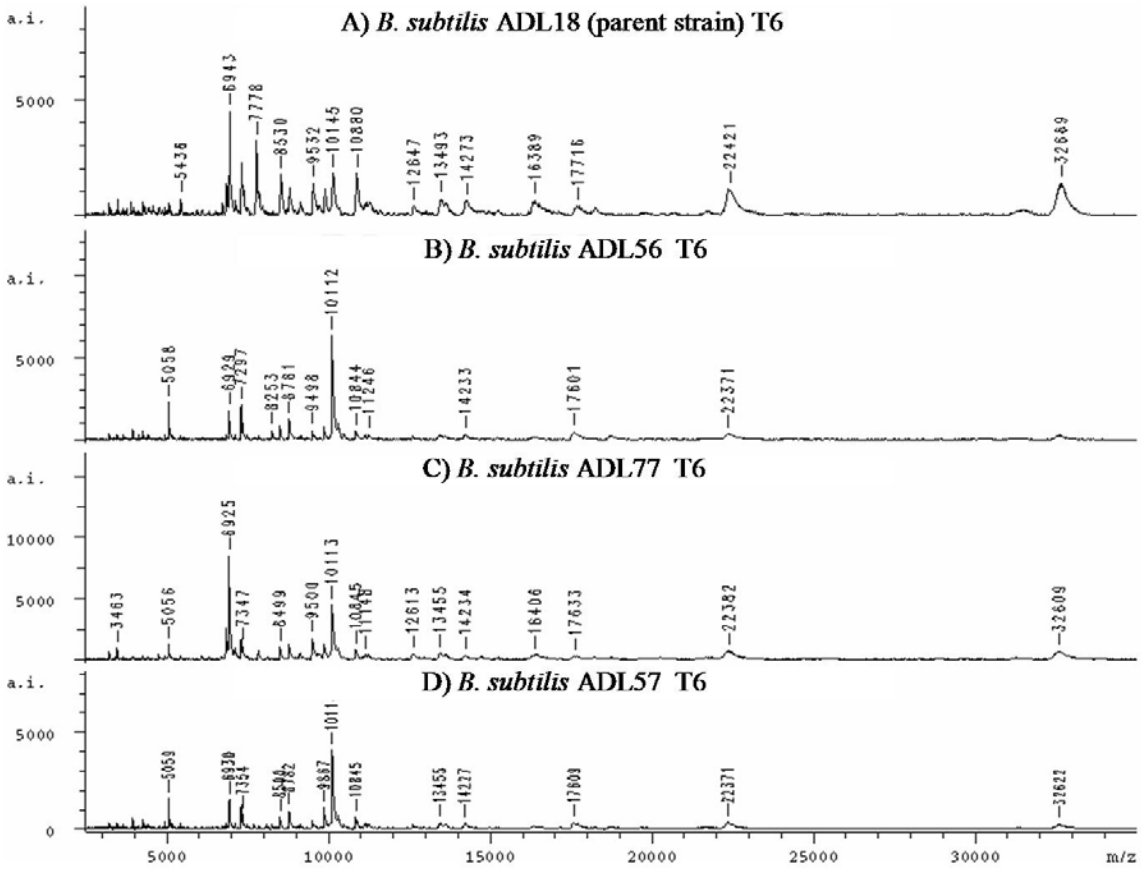

Figure 7-15. MALDI-TOFMS protein profiles of the cells of the *B. subtilis* spore coat mutants at T6. A) ADL18 (parent strain). B) ADL56. C) ADL77. D) ADL57. The mass range is from *m/z* 2,500-40,000.

groups), 12,600, 16,400, 17,700, 22,380, and 32, 630; however, this similarity did not translate into high correlation values between the two groups.

At T24, the sporulation should be complete with the phase bright spore released from the mother cell. These spectra are shown in Figure 7-16 A-D with the parent strain "normal" spore shown in spectra A and the mutated strains in B-D. The parent strain spore in this study had a high correlation value (r=0.94)with the B. subtilis 168 reference strain in the library. The mutated spores which resulted incomplete coat formation had correlation values ranging from 0.38-0.57 with the parent strain spore indicating a low degree of spectral similarity. Visual observation of these mutants did reveal that all spores had the 7,760 Da biomarker peaks in common with the parent strain. The mutated spores also had several peaks in common with each other including m/z 6,695, 8,760 and 10, 112. ADL56 and ADL57 were very similar to each other (r=0.89) with additional peaks in common at 7,915 Da and 17.6 kDa.

In general, the mutated spores had a greater number of peaks in the 3-10kDa range than "normal" B. subtilis spores. These peaks could be associated with SASPs or with polypeptide components of the spore coat that were now extracted due to the incomplete spore coat being more susceptible to the solvent extraction. Proteomic analysis and growth studies are needed to determine this. Follow-up studies on these mutants should also include analysis of vegetative cells in log phase growth (not entering the sporulation cycle) to determine if the biomarkers identified in the different mutants are specific to the sporulation process. Proteomic analysis of the samples in the different stages could then be focused on the sporulation specific proteins.

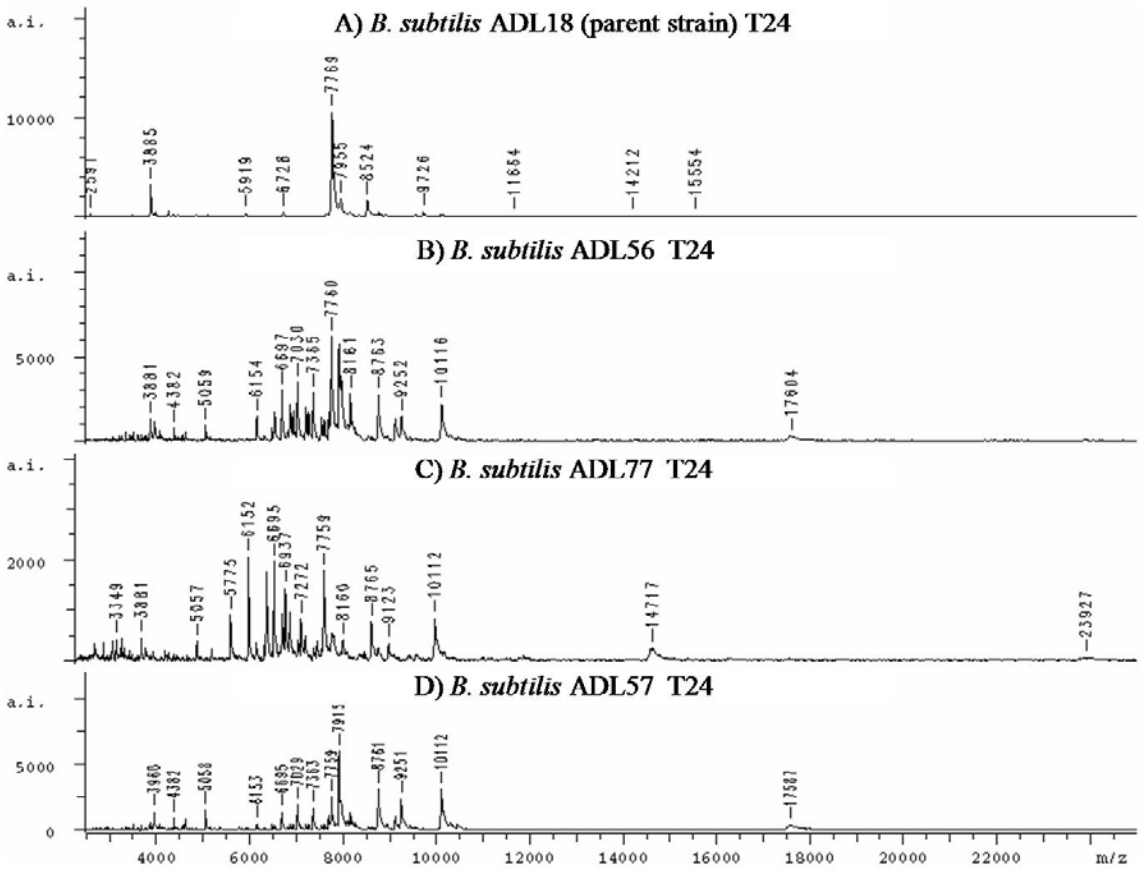

Figure 7-16. MALDI-TOFMS protein profiles of the spores of the *B. subtilis* spore coat mutants at T24. A) ADL18 (parent strain). B) ADL56. C) ADL77, D) ADL57. The mass range is from *m/z* 2,500-40,000.

Conclusion

The *B. pumilus* 7061 spores were the least impacted by the different exposure conditions. This may be due to the fact that these spores are more susceptible to the initial formic acid extraction and thus no detectable change in the protein profile occurs for spores that have been treated. The *B. subtilis* 168 spores did show increases in SASP associated proteins (6,940 and 9,136 Da) particularly for the autoclaved and H_2O_2 vapor treated samples indicating a change in spore coat permeability post treatment. Radiation exposed samples, except in the case of the UV treated *B. subtilis* 168, showed no difference that could be attributed to exposure or a change in spore coat permeability for any of the strains. The examination of the sporulation and spore coat mutants from *B. subtilis* provided insight into possible sporulation specific biomarkers that can be linked to spore formation, although further proteomic analysis is needed to identify the peaks.

The FO-36b spores proved to be the most interesting group in this study. FO-36b spores exhibited significant changes in protein profiles when autoclaved and in long term storage in water. This change involved the appearance of biomarkers at 6,820 and 7,240 Da which we believe are SASP associated proteins that are also present in *B. pumilus* 7061. The presence of the biomarker peak at 7,620 is necessary for the differentiation of FO group spores from the *B. pumilus* type strain.[96] The 7,620 Da peak also proved interesting in the liquid H_2O_2 treated spores. Unlike the other 2 species in this study, no evidence of protein oxidation was seen in the FO-36b spores. The reason for this warrants further studies but the lack of a molecular weight shift could be used as an indicator of H_2O_2 resistant strains.

This study sought to answer the following important question: are species-specific biomarkers still present under the exposure conditions such that differentiation of the

spore strains is still possible? Except in the case of the mutated strains and H_2O_2 treated spores, most of the exposures or differences in growth or storage conditions resulted in spectra that contained species-specific biomarkers. The linear correlation values obtained in some cases were low due to changes in the relative intensities of peaks in treated versus control spectra and/or the appearance of additional peaks in the treated spectra. To compensate for this, either different statistical analyses could be used which only consider peak location and not intensities, or spectral libraries could contain profile entries from treated spores. This study also indicated that it is possible to use the differences in the spectra to provide information on cell viability and explanations for increased resistances. The additional information provided by the appearance of these biomarkers can potentially provide evidence that can be used for source tracking, forensic investigations, and epidemiological studies. These results are encouraging and confirm the robustness of MALDI-TOFMS protein profiling methodology as a rapid diagnostic tool for both laboratory cultured strains and field samples.

CHAPTER 8
CONCLUSIONS AND FUTURE WORK

MALDI-TOFMS based protein profiling has been demonstrated as a rapid diagnostic tool for the analysis of *Bacillus* spores and cells. The success of the technique was realized through the combination of linear correlation and cluster analysis with an optimized one-step sample preparation. The technique was validated by its application for the analysis of over 50 different *Bacillus* strains, the largest and most thorough examination of the genus to date. Using a polyphasic approach for classification, we were able to substantiate the MALDI-TOFMS protein profiling technique with other well-established genotypic and phenotypic methods. MALDI-TOFMS protein profiling was shown to be more discriminating than other phenotypic tests, such as Biolog and FAME analysis, and was demonstrated to be complementary to genetic methods such as DNA:DNA hybridization and *gyrB* sequence analysis. The protein profiling experiment was rapid, reproducible, and sensitive, requiring less than 10,000 cells for species identification (data not shown). Using the criteria for species identification of a correlation value greater than 0.75 and a delta value of 0.1, the technique was shown to be robust and versatile for the analysis of environmentally challenged spores. Invariant and omnipresent species-specific biomarkers could be identified for almost all of the strains examined.

Because of its reliability, discriminating power, speed, and sensitivity, the protein profiling technique using MALDI-TOFMS is far better than other phenotypic typing and chemotaxonomic methods that are in use today including whole-cell protein

fingerprinting using gel electrophoresis. Since the MALDI analysis allows for better resolution of proteins, it can define more unique proteins in a sample which can allow for better differentiation. An example of this is seen in the analysis of the *B. pumilus* group isolates where the differentiation of the FO group bacteria required the resolution of the 7,640 Da biomarker peak. The rate at which analyses can be performed and the potential for automation is also a great advantage over genotypic techniques such as hybridization and sequence analysis. There is a significant initial capital cost associated with the purchase of a mass spectrometer; however, the long-term cost per sample is low, due mainly to the non-labile and inexpensive consumables. The technique also uses a reference library for comparison of new isolates, eliminating the time and effort needed to grow and maintain reference strains for each analysis as in DNA:DNA hybridization experiments.

Limitations that still exist for this technique are the size and breadth of the MALDI protein profile database; the reliance on manual interpretation of the spectra in order to deal with atypical strains; the need for pure cultures; the lack of proteome information on *Bacillus* in databases for protein identification and the lack of understanding of the post translational modifications and divergence in spore coat proteins.

Neill Logan put it best when he said "Taxonomists can only be as good as their culture collections and identification systems can only be as good as their databases." Therefore, future work in this area must first be to continue to expand the *Bacillus* profile library with thoroughly characterized reference strains that represent the vast diversity of this genus. This will certainly mean using multiple strains of a given species for reference as opposed to type strains which do not always encompass all the phenotypic

properties within a species as was evident in the BACT group analysis. The addition of strains to the database would be best managed by a multidisciplinary team which includes scientists from microbiology, chemistry, and statistics, where each member brings a different view of the analysis to the table.

Changing the statistical treatment of the spectra for identification is also inevitable. While the linear correlation analysis has worked throughout the course of this research, the manual interpretation of the spectra is the limiting factor in the speed and analysis of these species. The liberal criteria used for the correlation analysis allowed for effective species discrimination but often ignored the finer points of strain differentiation which required manual interpretation of peaks. The solution is to move towards using peak picking algorithms for profile identification. The use of peak picking algorithms should allow us to weight species-specific biomarker peaks in the spectra, which, should in turn allow us to deal with strain variation and atypical strains more efficiently. The use of peak picking algorithms has also been shown to be successful for the analysis of spore mixtures. Preliminary investigation of the Algoworks statistical software developed at Pacific Northwest National Laboratories is currently underway for this purpose.

To identify the species-specific biomarker peaks in the spectra, we have to patiently wait for more *Bacillus* genome projects to be completed and translated into protein information in the databases or perform *de novo* sequencing of the extracted proteins. Newly developed mass spectrometry techniques such as electron capture dissociation allow for the sequencing of whole proteins and could be utilized for this purpose. For this type of analysis however, it will be necessary to isolate the protein of interest prior to

analysis. Because of the limited resolution of 1D gel electrophoresis, either 2D gel electrophoresis or chromatography will have to be developed for this purpose.

The research here provides a platform for the design of diagnostics for a wide range of applications that can include source tracking, epidemiological studies, forensic investigations, determination of resistances, niche, and natural selection and direct environmental monitoring. While the ultimate design of the instrument and the sample requirements may vary in these different applications, the fundamental and polyphasic approach used here allows for the rapid identification of new characters that have diagnostic value for the differentiation of the *Bacillus* genus.

LIST OF REFERENCES

1. Sonenshein, A. L.; Hoch, J. A.; Losick, R. *Bacillus Subtilis and Other Gram-positive Bacteria Biochemistry, Physiology, and Molecular Genetics,* American Society for Microbiology: Washington, D.C, 1993.

2. Logan, N. A. *Applications and Systematics of Bacillus and Relatives,* Blackwell Scientific Publications: Oxford, 2002.

3. Madigan, M. T. *Brock Biology of Microorganisms,* 8th ed.; Prentice-Hall: Upper Saddle River, New Jersey, 1997.

4. Nicholson, W. L.; Fajardo-Cavazos, P.; Rebeil, R.; Slieman, T. A.; Riesenman, P. J.; Law, J. F.; Xue, Y. M. *Int. J. Gen. Mol. Microbiology* **2002,** *81*, 27-32.

5. Driks, A. *Microbiol. Mol. Biol. Rev.* **1999,** *63*, 1-20.

6. Driks, A. *Cell Mol. Life Sci.* **2002,** *59*, 389-91.

7. Nicholson, W. L.; Setlow P. in *Molecular Biological Methods for Bacillus*, Harwood, C. R.; Cutting, S. M., Eds.; Wiley: Chichester, MA, 1990.

8. Towner, K. J.; Cockayne, A. *Molecular Methods for Microbial Identification and Typing,* 1st ed; Chapman & Hall: London, 1993.

9. Logan, N. A. *Bacterial Systematics,* Blackwell Scientific Publications: Oxford, 1994.

10. Busse, H. J.; Denner, E. B.; Lubitz, W. *J. Biotechnol.* **1996,** *47*, 3-38.

11. Dass, C. *Principles and Practice of Biological Mass Spectrometry,* John Wiley: New York, 2001.

12. Fung, E. T.; Wright, G. L.; Dalmasso, E. A. *Cur. Opin. Mol. Ther.* **2000,** *2*, 643-50.

13. Fung, E. T.; Enderwick, C. *Biotechniques* **2002,** *Suppl*, 34-1.

14. Weinberger, S. R.; Dalmasso, E. A.; Fung, E. T. *Cur. Opin. Chem. Biol.* **2002,** *6*, 86-91.

15. Cain, T. C.; Lubman, D. M.; Weber, W. J. *Rapid Commun. Mass Spectrom.* **1994,** *8*, 1026-30.

16. van Baar, B. L. M. *FEMS Microbiol. Rev.* **2000,** *24*, 193-219.

17. Goldman, R. C.; Tipper, D. J. *J. Bacteriol.* **1981,** *147*, 1040-48.

18. Fox, A.; Stewart, G. C.; Waller, L. N.; Fox, K. F.; Harley, W. M.; Price, R. L. *J. Microbiol. Methods.* **2003,** *54*, 143-52.

19. Birmingham, J.; Demirev, P.; Ho, Y. P.; Thomas, J.; Bryden, W.; Fenselau, C. *Rapid Commun. Mass Spectrom.* **1999,** *13*, 604-06.

20. Arnold, R. J.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1998,** *12*, 630-36.

21. Holland, R. D.; Rafii, F.; Heinze, T. M.; Sutherland, J. B.; Voorhees, K. J.; Lay, J. O. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 911-17.

22. Vaidyanathan, S.; Rowland, J. J.; Kell, D. B.; Goodacre, R. *Anal. Chem.* **2001,** *73*, 4134-44.

23. Welham, K. J.; Domin, M. A.; Scannell, D. E.; Cohen, E.; Ashton, D. S. *Rapid Commun. Mass Spectrom.* **1998,** *12*, 176-80.

24. Easterling, M. L.; Colangelo, C. M.; Scott, R. A.; Amster, I. J. *Anal. Chem.* **1998,** *70*, 2704-09.

25. Elhanany, E.; Barak, R.; Fisher, M.; Kobiler, D.; Altboum, Z. *Rapid Commun. Mass Spectrom.* **2001,** *15*, 2110-16.

26. Hathout, Y.; Demirev, P. A.; Ho, Y. P.; Bundy, J. L.; Ryzhov, V.; Sapp, L.; Stutler, J.; Jackman, J.; Fenselau, C. *Appl. Environ. Microbiol.* **1999,** *65*, 4313-19.

27. Holland, R. D.; Wilkes, J. G.; Rafii, F.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O. *Rapid Commun. Mass Spectrom.* **1996,** *10*, 1227-32.

28. Krishnamurthy, T.; Ross, P. L. *Rapid Commun. Mass Spectrom.* **1996,** *10*, 1992-96.

29. Krishnamurthy, T.; Ross, P. L.; Rajamani, U. *Rapid Commun. Mass Spectrom.* **1996,** *10*, 883-88.

30. Lay, J. O. *Trac-Tr. Anal. Chem.* **2000,** *19*, 507-16.

31. Lay, J. O. *Mass Spectrom. Rev.* **2001,** *20*, 172-94.

32. Ryzhov, V.; Hathout, Y.; Fenselau, C. *Appl. Environ. Microbiol.* **2000,** *66*, 3828-34.

33. Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001,** *20*, 157-71.

34. Ryzhov, V.; Bundy, J. L.; Fenselau, C.; Taranenko, N.; Doroshenko, V.; Prasad, C. R. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 1701-06.

35. Ullom, J. N.; Frank, M.; Gard, E. E.; Horn, J. M.; Labov, S. E.; Langry, K.; Magnotta, F.; Stanion, K. A.; Hack, C. A.; Benner, W. H. *Anal. Chem.* **2001,** *73*, 2331-37.

36. Melly, E.; Genest, P. C.; Gilmore, M. E.; Little, S.; Popham, D. L.; Driks, A.; Setlow, P. *J. Appl. Microbiol.* **2002,** *92*, 1105-15.

37. Lai, E. M.; Phadke, N. D.; Kachman, M. T.; Giorno, R.; Vazquez, S.; Vazquez, J. A.; Maddock, J. R.; Driks, A. *J. Bacteriol.* **2003,** *185*, 1443-54.

38. Hathout, Y.; Setlow, B.; Cabrera-Martinez, R. M.; Fenselau, C.; Setlow, P. *Appl. Environ. Microbiol.* **2003,** *69*, 1100-07.

39. Paidhungat, M.; Setlow, B.; Driks, A.; Setlow, P. *J. Bacteriol.* **2000,** *182*, 5505-12.

40. Demirev, P. A.; Ramirez, J.; Fenselau, C. *Anal. Chem.* **2001,** *73*, 5725-31.

41. Liu, H. B.; Bergman, N. H.; Thomason, B.; Shallom, S.; Hazen, A.; Crossno, J.; Rasko, D. A.; Ravel, J.; Read, T. D.; Peterson, S. N.; Yates, J.; Hanna, P. C. *J. Bacteriol.* **2004,** *186*, 164-78.

42. Kuwana, R.; Kasahara, Y.; Fujibayashi, M.; Takamatsu, H.; Ogasawara, N.; Watabe, K. *Microbiol.* **2002,** *148*, 3971-82.

43. Dalluge, J. J. *Fres. J. Anal. Chem.* **2000,** *366*, 701-11.

44. Williams, T. L.; Andrzejewski, D.; Lay, J. O.; Musser, S. M. *J. Am. Soc. Mass Spectrom.* **2003,** *14*, 342-51.

45. Vaidyanathan, S.; Winder, C. L.; Wade, S. C.; Kell, D. B.; Goodacre, R. *Rapid Commun. Mass Spectrom.* **2002,** *16*, 1276-86.

46. Madonna, A. J.; Basile, F.; Ferrer, I.; Meetani, M. A.; Rees, J. C.; Voorhees, K. J. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 2220-29.

47. Cohen, S. L.; Chait, B. T. *Anal. Chem.* **1996,** *68*, 31-37.

48. Bornsen, K. O.; Gass, M. A. S.; Bruin, G. J. M.; vonAdrichem, J. H. M.; Biro, M. C.; Kresbach, G. M.; Ehrat, M. *Rapid Commun. Mass Spectrom.* **1997,** *11*, 603-09.

49. Breaux, G. A.; Green-Church, K. B.; France, A.; Limbach, P. A. *Anal. Chem.* **2000,** *72*, 1169-74.

50. Green-Church, K. B.; Limbach, P. A. *Anal. Chem.* **1998,** *70*, 5322-25.

51. Chapman, J. R. *MassSspectrometry of Protein and Peptides*; Humana Press: Totowa, NJ, 2000.

52. Zhang, N.; Doucette, A.; Li, L. *Anal. Chem.* **2001,** *73*, 2968-75.

53. Pace, N. R.; Marsh, T. L. *Orig. Life Evol. Biosph.* **1985,** *16*, 97-116.

54. Venkateswaran, K.; Moser, D. P.; Dollhopf, M. E.; Lies, D. P.; Saffarini, D. A.; MacGregor, B. J.; Ringelberg, D. B.; White, D. C.; Nishijima, M.; Sano, H.; Burghardt, J.; Stackebrandt, E.; Nealson, K. H. *Int. J. Syst. Bacteriol.* **1999,** *49 Pt 2*, 705-24.

55. Yamada, S.; Ohashi, E.; Agata, N.; Venkateswaran, K. *Appl. Environ. Microbiol.* **1999,** *65*, 1483-90.

56. Yamamoto, S.; Harayama, S. *Int. J. Syst. Bacteriol.* **1996,** *46*, 506-11.

57. Nakamura, L. K. *Int. J. Syst. Bacteriol.* **1989,** *39*, 295-300.

58. Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A. M.; Alloni, G.; Azevedo, V.; Bertero, M. G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S. C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.; Choi, S. K.; Codani, J. J.; Connerton, I. F.; Cummings, N. J.; Daniel, R. A.; Denizot, F.; Devine, K. M.; Dusterhoft, A.; Ehrlich, S. D.; Emmerson, P. T.; Entian, K. D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S. Y.; Glaser, P.; Goffeau, A.; Golightly, E. J.; Grandi, G.; Guiseppi, G.; Guy, B. J.; Haga, K.; Haiech, J.; Harwood, C. R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M. F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; KlaerrBlanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S. M.; Levine, A.; Liu, H.; Masuda, S.; Mauel, C.; Medigue, C.; Medina, N.; Mellado, R. P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; OReilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S. H.; Parro, V.; Pohl, T. M.; Portetelle, D.; Porwollik, S.; Prescott, A. M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlan, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S. J.; Serror, P.; Shin, B. S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H. F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. *Nature* **1997,** *390*, 249-56.

59. Venkateswaran, K.; Satomi, M.; Chung, S.; Kern, R.; Koukol, R.; Basic, C.; White, D. *Syst. Appl. Microbiol.* **2001,** *24*, 311-20.

60. La Duc, M. T; Kern, R.; Venkateswaran, K. *Microb. Ecol.***2004,** in press.

61. La Duc, M. T; Nicholson, W. L.; Kern, R.; Venkateswaran, K. *Environ. Microbiol.* **2003,** *5*, 977-85.

62. Schaeffer, P.; Millet, J.; Aubert, J. P. *Proc. Natl. Acad. Sci.* **1965,** *54*, 704-08.

63. Dickinson, D. N.; Powell, D. H.; Winefordner, J. D.; Kempf, M. J.; Venkateswaran, K. Proc. Am. Soc. Microbiol. **2002**, Salt Lake City, UT.

64. Dickinson, D. N.; Haskins, W. E.; Powell, D. H.; Winefordner, J. D.; La Duc, M. T.; Venkateswaran, K. Proc. Am. Soc. Mass Spectrom. **2003**, Montreal, Canada.

65. Gornushkin, I. B.; Ruiz-Medina, A.; Anzano, J. M.; Smith, B. W.; Winefordner, J. D. *J. Anal. At. Spectrom.* **2000,** *15*, 581-86.

66. Gornushkin, I. B.; Smith, B. W.; Nasajpour, H.; Winefordner, J. D. *Anal. Chem.* **1999,** *71*, 5157-64.

67. Gornushkin, S. I.; Gornushkin, I. B.; Anzano, J. M.; Smith, B. W.; Winefordner, J. D. *Appl. Spectrosc.* **2002,** *56*, 433-36.

68. Priest, F. G. in *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, Sonenshein, A. L.; Losick, R.Hock, J. A., Eds.; American Society for Microbiology: Washington, D.C., 1993.

69. Ash, C.; Farrow, J. A.; Dorsch, M.; Stackebrandt, E.; Collins, M. D. *Int. J. Syst. Bacteriol.* **1991,** *41*, 343-46.

70. La Duc, M. T; Satomi, M.; Venkateswaran, K. *Int. J. Syst. Bacteriol.* **2004,** in press.

71. Bright, J. J.; Claydon, M. A.; Soufian, M.; Gordon, D. B. *J. Microbiol. Methods.* **2002,** *48*, 127-38.

72. Jarman, K. H.; Cebula, S. T.; Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Kingsley, M. T.; Wahl, K. L. *Anal. Chem.* **2000,** *72*, 1217-23.

73. Dickinson, D. N.; La Duc, M. T.; Haskins, W. E.; Gornushkin, I.; Winefordner, J. D.; Powell, D. H.; Venkateswaran, K. *Appl. Environ. Microbiol.* **2004,** *70*, 475-82.

74. Link, L.; Sawyer, J.; Venkateswaran, K.; Nicholson, W. *Microb. Ecol.* **2003**.

75. Nicholson, W. L.; Munakata, N.; Horneck, G.; Melosh, H. J.; Setlow, P. *Microbiol. Mol. Biol. Rev.* **2000,** *64*, 548-+.

76. Riesenman, P. J.; Nicholson, W. L. *Appl. Environ. Microbiol.* **2000,** *66*, 620-26.

77. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning a laboratory manual,* 2nd ed; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y, 1989.

78. Swofford, D. PAUP: phylogenetic analysis using parsimony. [3.0]. 1990. Champaign, IL, USA, Illinois Natural History Survey.

79. Ezaki, T.; Hashimoto, Y.; Yabuuchi, E. *Int. J. Syst. Bacteriol.* **1989,** *39*, 224-29.

80. Satomi, M.; Kimura, B.; Mizoi, M.; Sato, T.; Fujii, T. *Int. J. Syst. Bacteriol.* **1998,** *48*, 332.

81. Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987,** *4*, 406-25.

82. Felsenstein, J. *Syst. Biol.* **1997,** *46*, 101-11.

83. Wayne, L. G.; Brenner, D. J.; Colwell, R. R.; Grimont, P. A. D.; Kandler, O.; Krichevsky, M. I.; Moore, L. H.; Moore, W. E. C.; Murray, R. G. E.; Stackebrandt, E.; Starr, M. P.; Truper, H. G. *Int. J. Syst. Bacteriol.* **1987,** *37*, 463-64.

84. Jarman, K. H.; Daly, D. S.; Petersen, C. E.; Saenz, A. J.; Valentine, N. B.; Wahl, K. L. *Rapid Commun. Mass Spectrom.* **1999,** *13*, 1586-94.

85. Jarman, Kristin H.; Daly, Don S.; Anderson, Kevin K.; Wahl, Karen L. *Chemometrics and Intelligent Laboratory Systems* **2003,** *69*, 61-76.

86. Wahl, K. L.; Wunschel, S. C.; Jarman, K. H.; Valentine, N. B.; Petersen, C. E.; Kingsley, M. T.; Zartolas, K. A.; Saenz, A. J. *Anal. Chem.* **2002,** *74*, 6191-99.

87. Turnbull, P. C. .*J Appl. Microbiol.* **1999,** *87*, 237-40.

88. La Duc, M. T; Satomi, M.; Agata, N.; Venkateswaran, K. *J. Microbiol. Methods.* **2004,** *56*, 383-94.

89. Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Gantt, S. L.; Jarman, K. H.; Kingsley, M. T.; Wahl, K. L. *Rapid Commun. Mass Spectrom.* **1999,** *13*, 1580-85.

90. Henderson, I.; Duggleby, C. J.; Turnbull, P. C. *Int. J. Syst. Bacteriol.* **1994,** *44*, 99-105.

91. Ramisse, V.; Patra, G.; Garrigue, H.; Guesdon, J. L.; Mock, M. *FEMS Microbiol. Lett.* **1996,** *145*, 9-16.

92. Keim, P.; Kalif, A.; Schupp, J.; Hill, K.; Travis, S. E.; Richmond, K.; Adair, D. M.; Hugh-Jones, M.; Kuske, C. R.; Jackson, P. *J. Bacteriol.* **1997,** *179*, 818-24.

93. Jackson, P. J.; Hill, K. K.; Laker, M. T.; Ticknor, L. O.; Keim, P. *J. Appl. Microbiol.* **1999,** *87*, 263-69.

94. Harrell, L. J.; Andersen, G. L.; Wilson, K. H. *J. Clin. Microbiol.* **1995,** *33*, 1847-50.

95. Helgason, E.; Okstad, O. A.; Caugant, D. A.; Johansen, H. A.; Fouet, A.; Mock, M.; Hegna, I.; Kolsto, A. B. *Appl. Environ. Microbiol.* **2000,** *66*, 2627-30.

96. Dickinson, D. N.; La Duc, M. T.; Winefordner, J. D.; Powell, D. H.; Venkateswaran, K. *J. Microbiol. Methods.* **2004,** in press.

97. Hathout, Y.; Ho, Y. P.; Ryzhov, V.; Demirev, P.; Fenselau, C. *J. Nat. Prod.* **2000,** *63*, 1492-96.

98. Bacher, G.; Korner, R.; Atrih, A.; Foster, S. J.; Roepstorff, P.; Allmaier, G. *J. Mass Spectrom.* **2001,** *36*, 124-39.

99. Warscheid, B.; Jackson, K.; Sutton, C.; Fenselau, C. *Anal. Chem.* **2003,** *75*, 5608-17.

100. Warscheid, B.; Fenselau, C. *Anal. Chem.* **2003,** *75*, 5618-27.

101. Dickinson, D. N.; La Duc, M. T.; Haskins, W. E.; Gornushkin, I.; Winefordner, J. D.; Powell, D. H.; Venkateswaran, K. *Appl. Environ. Microbiol.* **2004,** *70*, 475-82.

102. Cutting, S.; Oke, V.; Driks, A.; Losick, R.; Lu, S.; Kroos, L. *Cell* **1990,** *62*, 239-50.

103. Conway, G. C.; Smole, S. C.; Sarracino, D. A.; Arbeit, R. D.; Leopold, P. E. *J. Mol. Microbiol. Biotech.* **2001,** *3*, 103-12.

104. Demirev, P. A.; Lin, J. S.; Pineda, F. J.; Fenselau, C. *Anal. Chem.* **2001,** *73*, 4566-73.

105. Pineda, F. J.; Lin, J. S.; Fenselau, C.; Demirev, P. A. *Anal. Chem.* **2000,** *72*, 3739-44.

106. Demirev, P. A.; Ho, Y. P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999,** *71*, 2732-38.

107. Holland, R. D.; Duffy, C. R.; Rafii, F.; Sutherland, J. B.; Heinze, T. M.; Holder, C. L.; Voorhees, K. J.; Lay, J. O. *Anal. Chem.* **1999,** *71*, 3226-30.

108. Krishnamurthy, T.; Rajamani, U.; Ross, P. L.; Eng, J.; Davis, M.; Lee, T. D.; Stahl, D. S.; Yates, J. *Natural and Selected Synthetic Toxins* **2000,** *745*, 67-97.

109. Krishnamurthy, T.; Rajamani, U.; Ross, P. L.; Jabhour, R.; Nair, H.; Eng, J.; Yates, J.; Davis, M. T.; Stahl, D. C.; Lee, T. D. *J. of Tox.-Toxin Rev.* **2000,** *19*, 95-117.

110. Harris, W. A.; Reilly, J. P. *Anal. Chem.* **2002,** *74*, 4410-16.

111. Vaidyanathan, S.; Rowland, J. J.; Kell, D. B.; Goodacre, R. *Anal. Chem.* **2001,** *73*, 4134-44.

112. Haskins, W. E.; Wang, Z. Q.; Watson, C. J.; Rostand, R. R.; Witowski, S. R.; Powell, D. H.; Kennedy, R. T. *Anal. Chem.* **2001,** *73*, 5005-14.

113. Yates, J. R., III; Morgan, S. F.; Gatlin, C. L.; Griffin, P. R.; Eng, J. K. *Anal. Chem.* **1998,** *70*, 3557-65.

114. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999,** *20*, 3551-67.

115. Aronson, A. I.; Song, H. Y.; Bourne, N. *Mol.Microbiol.* **1989,** *3*, 437-44.

116. Sonenshein, A. L.; Hoch, J. A.; Losick, R. *Bacillus subtilis and its closest relatives,* ASM Press: Washington, D.C, 2002.

117. Little, S.; Driks, A. *Mol. Microbiol.* **2001,** *42*, 1107-20.

118. Bauer, T.; Little, S.; Stover, A. G.; Driks, A. *J. Bacteriol.* **1999,** *181*, 7043-51.

119. Beall, B.; Driks, A.; Losick, R.; Moran, C. P., Jr. *J. Bacteriol.* **1993,** *175*, 1705-16.

BIOGRAPHICAL SKETCH

Danielle N. Dickinson was born in St. Petersburg, Florida on December 23, 1974. She is the daughter of Connie Putman and Sidney P. Rowe who, at 2-year intervals, provided her with two siblings, a sister, Deborah Shannon Rowe; and a brother, Michael Aaron Rowe. Her father passed away in 1986, and her mother later married James W. Hancock.

At Brookwood High School her fascination with science began; thanks largely to the individual guidance and encouragement she received from her AP Chemistry instructor, Dr. Pat Saulson. When she graduated with honors in 1992 it was no surprise that she chose to attend Embry Riddle Aeronautical University (ERAU) on a full Air Force ROTC scholarship; ostensibly providing the opportunity to pursue her childhood dream of becoming an astronaut. While attending ERAU, she was active in Alpha Xi Delta Sorority and the Future Professional Women in Aviation, played several intramural sports including volleyball and softball, and served as a summer-camp counselor.

Unable to retain her scholarship due to a medical misdiagnosis, and unable to remain at ERAU without the scholarship, in the fall of 1994 she departed ERAU and enrolled at North Georgia College to pursue a degree in Chemistry. There, she was a member of the state championship flag football team, the Fellowship of Christian Athletes, the Baptist Student Union, Gamma Sigma Sigma Service Sorority, Omicron Delta Kappa leadership honor society, and Phi Kappa Phi Honor Sorority. She readily became an integral part of the chemistry department. As such, she was a teaching

assistant for most of the chemistry classes, and rebuilt a donated capillary electrophoresis instrument as an undergraduate research project. She spent a summer at Clemson University in the National Science Foundation Summer Undergraduate Research Program doing inorganic chemistry research under Dr. Hwu. She received her baccalaureate degree in chemistry, Magna Cum Laude, December 12th, 1997, and was recognized as the Chemistry Department's Outstanding Graduate for academic year 1996-1997. During her senior year, she was an intern working in research and development for Ciba Vision in Duluth, Georgia, and post-graduation continued with them full time as a scientist in the analytical, formulations, and sterilization technology groups.

On July 4th of 1997 she met then Marine Corporal Owen Dickinson, who was to become her husband; they wed on February 27th, 1999. She joined the analytical division of the chemistry department at the University of Florida in the fall of 1999, and began working with personnel from both the Kennedy Space Center and Jet Propulsion Laboratory (JPL). While there she received guidance from Dr. James Winefordner, Dr. David Powell and JPL's Dr. Kasthuri Venkateswaran. She was supported in her graduate education by a NASA Graduate Student Research Program Fellowship, and was also a Grinter and a Rue-Gammer Fellow. In her spare time, she enjoyed participating in intramural sports, found time to begin an annual Women in Science Conference, and served as a chapter adviser to her sorority. She received her Doctor of Philosophy degree in chemistry in August 2004. Danielle has been awarded an ORISE postdoctoral fellowship to work in the Counterterrorism and Forensics Unit at the FBI Academy in Quantico, Virginia.