

1-1-2013

Mass Spectrometry and Tandem Mass Spectrometry Characterization of Protein Patterns, Protein Markers and Whole Proteomes For Pathogenic Bacteria

Jennifer Intelicato-Young
University of South Carolina

Follow this and additional works at: <http://scholarcommons.sc.edu/etd>

Recommended Citation

Intelicato-Young, J.(2013). *Mass Spectrometry and Tandem Mass Spectrometry Characterization of Protein Patterns, Protein Markers and Whole Proteomes For Pathogenic Bacteria*. (Master's thesis). Retrieved from <http://scholarcommons.sc.edu/etd/2102>

This Open Access Thesis is brought to you for free and open access by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact SCHOLARC@mailbox.sc.edu.

MASS SPECTROMETRY AND TANDEM MASS SPECTROMETRY
CHARACTERIZATION OF PROTEIN PATTERNS, PROTEIN MARKERS AND WHOLE
PROTEOMES FOR PATHOGENIC BACTERIA

by

Jennifer Intelicato-Young

Bachelor of Arts
Bloomsburg University of Pennsylvania, 2006

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in

Biomedical Science

School of Medicine

University of South Carolina

2013

Accepted by:

Alvin Fox, Director of Thesis

Kim Creek, Reader

Cory Robinson, Reader

Lacy Ford, Vice Provost and Dean of Graduate Studies

ACKNOWLEDGEMENTS

Support for this work was provided by the National Science Foundation (#0959427, J. Rose, P.I.).

ABSTRACT

There have been many recent reviews published on MALDI-TOF MS (matrix assisted laser desorption/ionization time-of-flight) MS (mass spectrometry) for identification of bacteria particularly with relevance to clinical microbiology. MALDI-TOF MS is now a mature technique for bacterial identification with great promise. The purpose of this review is to put into perspective MALDI-TOF MS and other widely used mass spectrometry methods for characterization of proteins. MALDI-TOF MS is used for rapid determination of a mass pattern of proteins for bacterial characterization; these proteins are generally not identified. Alternatively, after gel separation, MALDI-TOF-TOF MS-MS (tandem mass spectrometry) or on-line LC-ESI MS-MS (liquid chromatography-electrospray tandem mass spectrometry) specific protein markers can be identified and peptide sequence variation among species assessed. Unlike direct MALDI-TOF MS, sample preparation for gel separation/MALDI-TOF-TOF MS and MS-MS remains quite demanding. Specific marker proteins are readily identified. Sample preparation is quite straightforward for LC-MS-MS. Massive amounts of information (whole proteomes) are provided but bioinformatics is complex. Chromatography and electrospray mass spectrometry instrumentation is also not widely used among microbiologists. Thus, there is a need for further development in sample preparation and instrumental development for rapid and simplified analysis.

As MS-MS for microbial characterization reaches maturity, it is to be anticipated that further developments in bioinformatics will also become essential. The genome codes for all proteins that might be synthesized under certain growth conditions but only direct protein identification can prove that specific proteins or networks of proteins are actually expressed which might be of relevance in improving our understanding of bacterial pathogenesis.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
ABSTRACT	iii
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: HISTORY OF MALDI-TOF MS	4
CHAPTER 3: MODERN “DIRECT” MALDI-TOF MS ANALYSIS OF WHOLE CELL PROTEINS OR FRACTIONS.....	6
CHAPTER 4: GEL SEPARATION FOLLOWED BY MALDI-TOF MS AND MALDI-TOF-TOF MS-MS.....	14
CHAPTER 5: LIQUID CHROMATOGRAPHY ELECTROSPRAY TANDEM MASS SPECTROMETRY (LC-MS-MS) ANALYSIS OF WHOLE CELL PROTEINS OR FRACTIONS.....	17
CHAPTER 6: FUTURE DIRECTIONS AND CONCLUSIONS.....	20
REFERENCES	22

CHAPTER 1

INTRODUCTION

There are few branches of science that have not gained benefit from the vast amount of information encoded in the genes of all living organisms. Looking into the DNA of an organism gives us a snapshot of everything that “could be”. It provides us with data that can be used for taxonomic discrimination, clinical microbiology, epidemiology, infectious disease initiation and perpetuation, forensics and biodetection. Techniques such as cloning, PCR (polymerase chain amplification) DNA-DNA hybridization and microarrays are widely used to isolate, purify, amplify, identify and manipulate genes. Correct identification and classification of bacteria is crucial to clinical and environmental microbiologists. However beneficial, genomic techniques and the information they provide can have limitations in classifying microbes. As an example, the polymerase chain reaction (PCR), although universally used, has some drawbacks in bacterial characterization. The procedure itself can be complicated for a routine clinical microbiology laboratory even in real-time format. It is also necessary to have some prior knowledge of the gene of interest in order to create operational primers. Additionally, effective primer selection requires there be two conserved regions for genetic recognition and amplification, thereby limiting the number of useful genes (Jabbour et al., 2010; Fox et al., 2011).

Analysis of 16S rRNA gene sequences is also greatly exploited in discriminating and classifying bacteria. The 16S rRNA gene is highly conserved between different bacterial species. The variable sequence regions found within the gene have been long considered the gold standard for bacterial identification and taxonomical classification. However, in certain cases, 16S rRNA sequences are nearly identical for closely related bacteria; therefore sequencing is often useless in discriminating bacteria past the genus level. Species level discrimination is even further complicated when introducing environmental isolates since most have not had their 16S rRNA sequenced or cataloged as they are not of clinical importance. This is particularly difficult when the environmental isolates are genetically closely related to industrially important or clinically relevant species that may also be present in the same milieu (Tani et al., 2012). For example, the 16S rRNA gene is so conserved among several species for the *Bacillus cereus* group that they cannot be differentiated to the species level. Better discrimination can be obtained using variations in the SASP protein sequence (Castanha et al. 2006a, 2006b, Callahan et al., 2009) but more work remains to be done differentiating some of the species. The same can be said for APEO_n (alkylphenol polyethoxylate) -degrading bacteria e.g. strain BSN20 and *Sphingopyxis terrae* strain NBRC 15098, which share 99.9% sequence identity (Dworzanski et al., 2010, Hotta et al., 2012). It has been suggested that using less conserved genes (e.g. gyrase [gyrB], RNA polymerase [rpoB], aconitate hydratase and superoxide dismutase [sodA]) may yield a higher discriminatory power when trying to speciate certain bacterial strains (Dworzanski et al., 2010). Protein patterns, protein markers and proteomic techniques have become an increasingly popular solution to such issues. Regardless of whether a protein is conserved or variable in

sequence it can be analyzed directly without amplification. Furthermore, if one potential protein marker does not provide adequate information, another can be readily selected as an alternative. Additionally, the use of mass spectrometry for full-scale proteomics offers an alternative to genomic techniques.

CHAPTER 2

HISTORY OF MALDI-TOF MS

The use of mass spectrometers for the identification and classification of microorganisms is not a new phenomenon. In one of the earliest applications, John Anhalt and Catherine Fenselau in 1975 described a technique coupling pyrolysis (heating to high temperatures in the absence of oxygen) with mass spectrometry by which lyophilized bacteria were directly inserted into the ion source of a mass spectrometer. Unfortunately, the complex and elegant monomeric, oligomeric and polymeric structures present in bacteria are converted into low mass volatiles, destroying most of the information present in the original cellular biochemistry. These early attempts were able to differentiate certain bacteria to some extent. While the spectra of gram-negative bacteria were less distinct, the mass spectrum of *S. aureus* was found to be markedly distinguishable from that of *S. epidermidis*. In addition to Catherine Fenselau, Henk Meuzelaar, was another pioneer who persisted in attempting to popularize pyrolysis MS (Dworzanski et al. 1990). Although pyrolysis has since fallen to the wayside, the model of obtaining the mass spectra of microbes and comparing them as a means of characterization has become widely accepted.

The 1980's brought advancements in ionizing molecules of higher molecular weights. In 1988, Franx Hillenkamp and co-workers used an organic matrix along with ultraviolet laser desorption to detect compounds with a mass range greater than 10,000

Da (Karas et al., 1985, Karas et al.1985). Earlier in 1988, Koichi Tanaka employed an “ultra fine metal plus liquid matrix method” that allowed obtainment of mass spectra up to 100,000 m/z (Tanaka et al., 1988). Tanaka developed a laser desorption ionization TOF MS instrument that, when coupled with the use of the ultra fine metal plus liquid matrix method, was able to detect lysozyme (mw 14,306Da), chymotrypsinogen (mw 25,717Da), Poly(propylene glycol) (4kDa), and PEG20K (20kDa). These early advancements in detection unknowingly set the stage for improving the expediency of mass spectrometry in the field of microbiology.

Several groups in the mid-1990s: (including Edward-Jones et al., 2000; Holland et al., 1996, 1999; Krishnamurphy et al., 1996, 1999; Wall et al., 1999) should be credited with the introduction of MALDI-TOF MS for bacterial analysis. At this time the resolution of the mass spectrometers was poor (both delayed extraction and reflectron mode were introduced much later) and databases had not been developed. Each of these developments helped in the success of modern MALDI-TOF MS. The development of commercial MALDI-TOF-TOF MS-MS (largely attributed to Marvin Vestal) has also helped immensely.

CHAPTER 3

MODERN “DIRECT” MALDI-TOF MS ANALYSIS OF WHOLE CELL PROTEINS OR FRACTIONS

The usefulness of rapid, accurate identification and classification of bacteria is far-reaching. Expedient and precise identification of pathogens in a clinical microbiology laboratory is essential for proper diagnosis and treatment of patients. It is also crucial for homeland security threats, as was witnessed by the 2001 anthrax attacks on the US postal system. Additionally, such techniques greatly benefit strain identification for food safety and taxonomical classification in environmental microbiology. Currently, the gold standard in the clinical diagnosis of microbial pathogens still relies heavily on phenotypic differences and growth patterns for identification of bacteria. Although simple to perform, for less studied species the results are often less than optimal.

Recent studies have shown that whole cell (WC) matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) can reduce the average identification time after growth to 10 min or less for identifying clinical microbiological isolates (Croxatto et al., 2011; Dubois et al., 2012; Giebel et al., 2010; Kliem and Sauer, 2012; Martiney et al., 2012; Risch et al., 2010; Safferet et al., 2012; Seibold et al., 2012; Welkner and Moore, 2011). However it must be emphasized that prior growth is still required which takes, from 5–48 hours for rapidly growing organisms and or in extreme

cases up to 12 weeks (e.g. mycobacteria). Direct analysis in clinical samples without culture is currently beyond the capability of MALDI-TOF MS.

Simply put, bacterial colonies are directly streaked onto a MALDI target plate, coated with a matrix and put into a mass spectrometer where mass spectra of the samples are generated and then analyzed by compatible software (Kliem, 2012). The mass spectrum of that sample is then compared to other spectra stored in a database. At least twenty isolates may be prepared, analyzed and subjected to automated software handling in 2.5 hours or less (Karger 2010, Seng 2009, Shitikov 2012, Wolters 2011). It has also been estimated that use of direct MALDI-TOF MS could be 22–32% of the cost of the current methods routinely employed (Seng et al., 2009). The premise is that each genera, species, or even strain of bacteria will produce a distinct protein profile that can be used as a means of identification.

As an example, the purpose of one study was to evaluate MALDI-TOF MS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) as a means to discriminate 12 *Escherichia* strains retaining 96–100% 16S rRNA gene sequence homology. MALDI-TOF MS was not only able to distinguish *Escherichia fergusonii* from other *E. coli* strains possessing 98.9–99.7% 16S rRNA sequence homology, but was also able to discriminate two strains of *E. coli* sharing identical 16S rRNA sequence identity (Masashi et al., 2011).

More than 95% of clinical isolates were identified in another study to the genus level, with greater than 98% of *Enterobacteriaceae* and the HACEK group being identified to the species level by comparing the mass spectra generated from MALDI-

TOF MS (van Veen et al., 2010). The HACEK group of bacteria is comprised of *Haemophilus*, *Actinobacillus*, *Cardiobacterium*, *Eikenella*, and *Kingella* species, and is categorized based on their potential to cause endocardial infections. Cluster analysis of the mass spectra of clinical isolates of mycobacteria allocated 67 samples into two distinct groups, the *Mycobacterium tuberculosis* complex (MTBC) and non-tuberculosis *Mycobacterium* (NTM). Although it was necessary to include a protein extraction step using 96% ethanol, 70% formic acid, and 100% acetonitrile with subsequent centrifugation to mitigate the extra stability offered by the lipids found in mycobacterium cell walls. With the exception of *M. bovis*, this technique was able to identify all species in the NTM group (Shitikov et al., 2012).

Oftentimes, as in the case of disease outbreaks, genus and species level identification is not sufficient. Therefore, strain level discrimination is sometimes desired, as is the case with methicillin-resistant *Staphylococcus aureus* (MRSA) (Du et al., 2002). The protein profile of MRSA has been shown to differ from that of methicillin-susceptible *Staphylococcus aureus* (MSSA) within the m/z range of 500–3500 Da in MALDI-TOF MS analysis. Moreover, MRSA was shown to have distinguishing peaks at m/z 891, 1140, 1165, 1229 and 2127, while MSSA had unique peaks at m/z 2548 and 2647 (Du et al., 2002, Edward-Jones et al., 2000). Even more discriminatory power was shown in a 2011 study where strains of hospital-acquired MRSA (methicillin resistant *Staphylococcus aureus*) were differentiated based on 18 m/z values found within the m/z range of 2500 to 7734 (Wolters et al., 2011).

Current studies are proving the effectiveness of using protein profiles to characterize and discriminate a variety of clinically relevant microbes, but there are some

genera and species of bacteria that are too closely related to be accurately discriminated. Such an issue arises when trying to differentiate pathogenic *Streptococcus pneumoniae* from the commensal, and sometimes pathogenic, viridians group streptococci (VGS). This may be because *S. pneumoniae* shares over 99% 16S rRNA sequence homology with *S. mitis*, *S. pseudopneumoniae*, and *S. oralis*. It has also been hypothesized that the streptococci are exceptionally competent bacteria and can readily take up DNA from their environment, or even transfer virulence factors via interspecies recombination, thus changing their expression of significant proteins. Under these circumstances, identification can be difficult. Phenotypic changes have been noted as well, as demonstrated in a 2011 study. Fifty streptococcal isolates were divided based on their susceptibility to optochin. Of the 50 isolates, 21 proved to be optochin-susceptible, and were subsequently frozen. Eighteen months later, the samples were thawed and re-tested. Only one isolate retained its optochin-susceptibility. MALDI-TOF MS analysis in this same study yielded 100% congruency with the typical genotypic and phenotypic identification tests at the species level, but only 4 out of 17 VGS were correctly identified and each of the 8 *S. mitis* isolates were incorrectly identified as being *S. pneumoniae* (Ikryannikova et al., 2011). Seven specific m/z peaks may prove beneficial in differentiating streptococcal species. These seven particular peaks were found to comprise 6 different MALDI-TOF MS profiles: MP1: m/z 2936, 5878; MP2: m/z 2623, 2936, 5253 5878; MP3: m/z 2623, 5253, 5878; MP4: m/z 2910, 5824, 6955; MP5: m/z 6955, and MP6: m/z 2623. The peak combinations found in MP1 appear to be biomarkers for *S. pneumoniae*, while those found in MP2 and MP3 were associated with *S. pseudopneumoniae*. MP4, MP5 and MP6 were characteristic of *S. mitis* (Werno et al.,

2012). Although promising, this study would need to be reproduced using more samples; n=48, with 14 isolates being *S. pneumoniae* (6 reference strains, 8 clinical isolates), 14 *S. pseudopneumoniae* (1 reference strain, 13 clinical isolates), 19 *S. mitis* (1 reference strain, 18 clinical isolates) and 1 reference strain of *S. oralis*.

The identification and discrimination of environmental species is even more difficult to discern. For one, the sheer volume and diversity of environmental species is overwhelming. Many of these species are difficult, if not impossible to culture. Thus the vast majority of environmental bacteria remain uncharacterized and grossly underrepresented in the databases. Although not without obstacle, mass spectrometry is proving to be as useful to the environmental microbiologist as is to the clinician (Rezzonico et al., 2010). However, the clinical perspective would greatly benefit from a method that could quickly and inexpensively differentiate known and novel pathogens isolated from the environment.

MALDI-TOF MS analysis has been performed on cultured bacteria isolated from airborne dust collected from occupied and unoccupied schoolrooms. It was initially shown that there is an increase in bacterial markers, such as cell wall peptidoglycan, and lipopolysaccharide in the occupied rooms as compared to the unoccupied rooms (Fox et al., 2005). In addition the skin epithelial keratin, k10, is the most abundant protein found in indoor air, pointing to humans as being a major contributor to the bacterial load in occupied buildings (Fox et al., 2008). This can have clinical implications. One mode of transmission of *Staphylococcus aureus* infections is via the air-borne route. A swift and economical technique is needed to exclude environmental species from further characterization. Out of 36 strains analyzed by MALDI-TOF MS, 8 were found to be

staphylococci (*S. aureus*: three; *S. hominis* two; *S. warneri*; two and one was only discerned to the genus level), 5 strains were not identified, and 23 were found to be micrococci. The micrococcus is also a resident of human skin, and is almost indistinguishable from staphylococci phenotypically. As such, it is crucial to be able to quickly discriminate between the two. Biochemical tests were also administered in this study, and the results were compared to those obtained from the MALDI-TOF MS. The isolates were first tested for glucose and glycerol fermentation. Staphylococci are able to ferment both, while micrococci are unable to ferment either. The results were largely in agreement with the MALDI-TOF MS identification. The fermentation tests showed more discriminatory power over the MS results in regard to the 5 unidentified strains. Of these strains, 3 were further characterized as being staphylococci. Among the Micrococcus strains, 2 strains showed aberrant fermentation but were correctly identified by MALDI-TOF MS analysis of tryptic peptides agreeing with direct MALDI-TOF MS analysis (Kooken et al, 2012). Additionally, the API Staph Ident Strip (Biomérieux, l'Etoile, France), a set of biochemical tests, was also used as a comparison. Only certainty above 85% is considered significant. Using this criterion, the API strip was only able to identify 3/8 of the staphylococcal isolates to the species level (Fox et al., 2010; Fox et al., 2011).

In regard to strain characterization, other methods such as multi-locus sequence analysis (MLSA) and pulsed-field gel electrophoresis (PFGE) are capable of sub-species identification, although typically 12-24 hours is required following bacterial isolation. In contrast, MALDI-TOF MS as mentioned above, takes a few minutes (Martens et al., 2008; Dworzanski et al., 2010; Karlsson et al., 2012).

A limiting factor in MALDI-TOF MS analysis is insufficient and/or incorrect database entries. There are well-established databases for MALDI-TOF MS (e.g. Bruker Biotyper protein profiling software and strain library (Bruker Leipzig, DE)). It has been shown that adding certain species to the database significantly improves MALDI precision in identification. Almost half of the *Streptococcus pneumoniae* strains were misidentified as *Streptococcus parasanguinis* in a study of 1660 isolates, because of an insufficient number of such species included in the Bruker database. Seven strains of *Stenotrophomonas maltophilia* were misidentified as *Pseudomonas hibiscicola* due to a mislabeling in the database. Additionally, all strains of *Shigella sonnei* were misidentified as *Escherichia coli*, as *S. sonnei* was absent from the database. Adding the correct spectra to the database corrected these issues and further identification was accurate (Seng et al., 2009).

A separate study used MALDI-TOF MS to identify 327 clinical isolates. Ten isolates were misidentified. Seven of the isolates were assigned the correct genus, but wrong species identification (minor error), while the remaining 3 had the wrong genus identification (major error). Of the major errors, *Acinetobacter* species was wrongly identified as *Campylobacter jejuni* and 2 *Sphingomonas* species were wrongly identified as *Paracoccus yeei*. *Paracoccus* was not included in the BioTyper database. Thirteen isolates had no identification, six of which were unable to be identified, while 7 did not show consistent identification between duplicate trials. Mislabeling in the BioTyper database was thought to be the cause of 12% of the *S. maltophilia* species being wrongly identified as *Pseudomonas geniculata*, *Pseudomonas hibiscicola*, or *Pseudomonas beteli* (Van Veen et al., 2010).

Many groups are creating in house databases in order to more accurately identify microbes. A technique, aptly named S10-GERMS (s10-spc-alpha operon gene-encoded ribosomal mass spectrum) was able to discriminate *Sphingopyxis terrae* strain NBRC 15098 from APEO_n-degrading bacteria (alkylphenol polyethoxylate) strain BSN20 by using s14, a ribosomal subunit protein, as a biomarker. The mass difference between both strains differed by an m/z of 14 (APEO_n-degrading bacteria strain BSN20: s14 m/z = 11513.6; *S. terrae* strain NBRC 15098: s14 m/z = 11527.6). These particular strains share a 16S rRNA sequence identity of 99.9%, having only a single base difference between the two, alluding to the assumption that there may be other factors contributing to the difficulty in discriminating species of streptococci. Perhaps it concerns the database in use. In this study, a ribosomal subunit protein database was created by correcting the virtual protein masses of the s10 and spc operon sequences, with the actual protein masses observed from whole cell MALDI-TOF MS. The database included the sequenced s10-spc-alpha operon of strains of *Sphingomonadaceae* and related APEO_n-degrading bacteria. This operon was able to produce 9 ribosomal subunit protein biomarkers (L18, L22, L24, L29, L30, S08, S14, S17, and S19) to differentiate *S. terrae* from other APEO_n-degrading bacteria. By creating a database that contains only the selected ribosomal subunit protein masses, one can avoid unwanted, or unnecessary biomarker peaks. The s10-GERMS method was also able to distinguish 3 strains of *S. terra* (NBRC 15593, NBRC 15598 and NBRC 15599) as well (Hotta et al., 2012).

CHAPTER 4

GEL SEPARATION FOLLOWED BY MALDI-TOF MS AND MALDI-TOF-TOF MS-MS

Gel separation followed by MALDI-TOF MS and MALDI-TOF-TOF MS-MS is well established for studying protein expression (e.g. bacterial pathogenesis). However, there has been limited use in bacterial taxonomy. Proteins are isolated (SDS-PAGE or 2D gel electrophoresis), enzymatically digested (in gel trypsin digest), and further analyzed for identification and sequence information. Several days are required for sample preparation. MS allows identification of distinct protein patterns among bacterial species but additionally MS-MS allows identification of proteins by their peptide sequences.

Experimental MS-MS patterns are also compared to virtual MS-MS patterns. More than one peptide sequence (at least two and preferably more) is required to be confident of protein identification. Once a protein has been identified, MS alone is all that is needed to identify a group of species (Fox et al., 2011, Kooken et al., 2012). Provided the genomic databases are available, sequence variation can be predicted using a program such as Protein Prospector.

Sometimes whole cell mass fingerprinting also has its limitations (Carbonelle et al., 2012). As previously mentioned, environmental bacteria are largely uncharacterized. This is reflected in their underrepresentation in MS m/z databases as compared to clinical pathogens (Alexander, 2012). Oftentimes, species-specific information can be found in

larger proteins as well. MALDI-TOF MS preferentially ionizes lower molecular weight proteins (<10kDa).

With MALDI-TOF-TOF MS-MS, however, one can separate desired intact proteins (bottom-down approach) using techniques such as gel electrophoresis or liquid chromatography, then analyze higher molecular weight proteins (50–100kDa) by enzymatically digesting the proteins into their respective peptides (Fox et al., 2011). The results of MS-MS can enhance the protein profile information obtained in MS.

Biomarkers can be selected by performing MS-MS on the most abundant and reproducible peaks found in a MALDI-TOF MS profile. A follow up to a previous study described above is one case in point. Improved bacterial speciation was possible for coagulase negative staphylococci isolated from the environment after homing in on a 100kDa band excised from 1D SDS PAGE gels. This band was identified as aconitate hydratase based on the sequence information obtained from 13 different peptides (derived from 6 staphylococcal species) from MS-MS analysis.

This study analyzed the 8 environmental strains from the aforementioned investigation and 5 reference strains for the following staphylococcal species: *S. cohnii*, *S. hominis*, *S. saprophyticus*, *S. warneri* and *S. chromogenes*. *S. capitis* was used as a negative control. With this approach, ion peaks of actual spectra obtained from MS-MS can be compared to virtual spectra found in a database (for example, NCBI). A standard search engine, i.e. Mascot, will generate a list of most probable peptides and their source, based on the sequence homology between actual and virtual spectra found within the database. However, use of a peak-picking program, such as Distiller, in conjunction with Mascot (both Matrix Science) allows utilization of the isotopic envelope in improving

peak selection. This probability is conferred by Mascot as a Mowse score, with $p < 0.05$ indicating extensive sequence homology. Out of the 4 staphylococcal species studied by this method, 3 were considered to have significant Mowse scores. Mascot revealed that for only half of the environmental isolates in this study, the top hits for protein identification were shown to be aconitate hydratase from staphylococcal species, with 3 of the 4 having significant Mowse scores. In contrast 7/8 environmental isolates were identified using visual discrimination and confirmed using custom-written cluster analysis software. This study demonstrates the complementary value of sequence information found in tryptic peptides previously separated by 1D gels and subsequent MS-MS analysis.

CHAPTER 5

LIQUID CHROMATOGRAPHY ELECTROSPRAY TANDEM MASS SPECTROMETRY (LC-MS-MS) ANALYSIS OF WHOLE CELL PROTEINS OR FRACTIONS

Whole cell extracts (or fractions) are subjected to trypsin digestion and then analyzed on-line using ESI-MS-MS. The flow from the liquid chromatograph (usually at nanoliter per ml flow rates [nanospray] for optimal sensitivity) is nebulized (sprayed) into the electrospray source. The development of electrospray MS analysis by John Fenn (who was awarded the Nobel Prize along with Koichi Tanaka) made these developments possible. In the mass spectrometer source the charged droplets evaporate placing multiple charges on each peptide. Thus electrospray spectra are quite complex since each peptide can have multiple charge states, each producing a distinct m/z value. These spectra can be simplified by deconvolution, where the multiple m/z peaks for each peptide are converted to a mass; the result resembles a MALDI spectrum. Mathematically, this is quite straightforward since the mass is the same but only the charge varies in adjacent m/z values for an individual spectrum. In the case of ESI MS-MS it is more difficult since, on fragmentation of a parent ion, both the mass and charge for daughter ions can vary. Using high resolution MS-MS (e.g. the LTQ-Orbitrap or Fourier transform ion resonance tandem mass spectrometry [FTICR]) multiple peaks (often 3 or 4 for peptides in the 1000-3000 mass range) differing by one or more units are resolved for each charge state due to differences in isotopic abundance (primarily hydrogen, deuterium and tritium). The charge is readily calculated; e.g. if there is one charge, then the difference in

mass in peaks in the isotopic window will be 1 (as in MALDI-TOF MS); for a charge of 2–0.5 mass units, 3–0.33, etc. There is often minimal sample preparation for LC-MS-MS but the mixture of peptides is derived from a complex matrix of proteins derived from the bacterial cell. It can be difficult to be confident that individual peptides (found in unconnected places in the chromatogram) are derived from the same protein. Furthermore the chromatograms, in whole proteomes, may not allow the same peptide to be identified in every run.

One group created a bacterial proteome database consisting of computer simulated tryptic peptides of translated, known protein-coding genes. Peptides from the experimental product ion mass spectra were obtained using LC tandem MS and identified using the SEQUEST algorithm. These peptides were then compared to those generated by the in house database. Any peptide sequence having a probability score of 95% or higher were used to produce a binary matrix that would match the peptides to genus, species or strain specific proteins. Peptides matched to a protein were assigned a score of 1, while those unmatched were assigned a score of 0. Twenty-one double blind microbial samples were matched with one of the 881 bacteria having sequenced genomes in the in house database by matching the number of unique peptides found between the two. Any degenerate peptides were eliminated using in-house created software (BACid) that retains only unique experimental peptides that correctly match those found in the NCBI protein database with a 99% confidence level or greater. This method effectively identified bacteria either singularly or in a mixture, with and without having a sequenced genome. The BACid algorithm correctly identified a strain of *Bacillus subtilis* based on its 212 unique peptides found in the sample. A mixture of 5 different bacteria was also

accurately identified in this manner; two to the species level (*S. aureus* and *B. thuringiensis*) and three to the strain level (*E. faecalis* V583, *B. thailandensis* E264, and *P. aeruginosa* PA01). Strain level identification was obtained for *S. aureus* ATCC 3359, a strain whose genome has yet to be sequenced. All of the unsequenced bacteria in the blind sample were correctly speciated, further demonstrating that comparing the experimentally obtained peptides to the virtual digests in the proteome could yield statistically significant identification (Jabbour et al., 2010). It is apparent that the utility and precision of MADLI-TOF MS, MALDI-TOF-TOF MS/MS and LC-ESI-MS-MS are all reliant on the development and maintenance of accurate and representative databases.

CHAPTER 6

FUTURE DIRECTIONS AND CONCLUSIONS

MALDI-TOF MS has reached maturity for bacterial identification. It is extremely simple to perform and database searching is straightforward. However, there are cases where discrimination is not optimal (Carbonelle et al., 2012). This is probably related to actual proteins that ionize from the sample and may not be improved by software changes alone. At this time gel separation followed by MALDI-MS-MS for protein identification and MALDI-TOF MS for subsequent protein sequence discrimination provides improved identification by greatly simplifying the sample to be analyzed and improving the knowledge gained from the analysis. However there is a real need for simplified protein separation. One system commercialized as the Gelfree 8100 Fractionation System (Expedion) allows in-liquid separation eliminating the steps for extracting proteins from gels (Tran and Doucette. 2009). Another commercial system (the Protea GPR System), more recently introduced, can elute intact proteins from gels in approximately 20 minutes. In both cases there is still the need for enzymatic release of peptides and limited sample cleanup. Others have described analysis of intact proteins using LC-MS (Everley and Croley, 2008; Krishnamurthy et al., 1999). However, ESI MS-MS is not optimal for analysis of intact protein of higher mass (approximately 30 kDa or above) (Tang et al., 2011). At this point there are no databases available for advanced MALDI or ESI MS and MS-MS methods for tryptic peptides or intact proteins since the technology is developmental in nature. Clearly there is a need for further improvements in sample

preparation, instrumental analysis and data handling. However, the future remains bright and we look forward to innovations in this field.

It should be stressed at this point both protein profiling and sequencing (due primarily to sensitivity issues) have been primarily limited to identification after culture. Indeed only abundant proteins are generally detected directly in complex matrices (whether clinical or environmental) without prior growth (Ram et al., 2005). However, similar to gene amplification approaches, there are proteomic techniques involving protein amplification prior to MALDI-TOF MS analysis. In one technique, a specific bacteriophage is amplified by multiplication within an infected bacterial host and the phage proteins are targeted for detection. (Madonna et al., 2003). Another technique consists of incubating exotoxin-producing bacteria with a synthetic peptide substrate, which upon enzymatic cleavage generates peptide products. In both instances, the peptide products are readily detected by MALDI-TOF MS with exquisite sensitivity. In a clinical microbiology laboratory, such methods can be an excellent choice (Boyer et al., 2009). However, amplification of DNA or proteins targets adds biological complexity to a readily automatable analytical chemical system necessary for biodetection. Furthermore, such focused methods are not optimal for biomarker discovery. The development of direct mass spectrometry techniques for biodetection is suggested to be a fruitful area for future research.

REFERENCES

- Boyer, A, et al., 2009. Kinetics of lethal factor and poly-D-glutamic acid antigenemia during inhalation anthrax in rhesus macaques. *Infect. Immun.* 77, 3432-3441.
- Callahan C., Fox K. Fox A., 2009. The small acid soluble proteins of *Bacillus weihenstephanensis* and *Bacillus mycoides* group 2 are the most distinct among the *Bacillus cereus* group. *Mol. Cell Probes.* 23, 291-297.
- Carbonnelle, E., Grohs, P., Jacquier, H., Day, N., Tenza, S., Dewailly, A., Vissouarn, O., Rottman, M., Herrmann, J-L., Podglajen, I., Raskine, L., 2012. Robustness of two MALDI-TOF mass spectrometry systems for bacterial identification. *J. Microbiol. Meth.* 89, 133-136.
- Castanha E., Fox K.F., Fox A., 2006. Rapid discrimination of *Bacillus anthracis* from other members of the *B. cereus* group by mass and sequence of intact small acid soluble proteins (SASPs) using mass spectrometry. *J. Microbiol. Meth.* 67, 230-240.
- Castanha E.R., Vestal V., Fox A., Fox K.F., Dickinson D. 2007. *Bacillus cereus* strains fall into two clusters (one closely and one more distantly related) to *Bacillus anthracis* according to amino acid substitutions in small acid-soluble proteins as determined by tandem mass spectrometry. *Mol. Cell. Probes.* 21, 190-201
- Croxatto, A., Prod'hom, G., Greub, G., 2011. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *Fed. Euro. Microbiol. Soc.*, 36, 380-407.
- Du, Z., Yang, R., Guo, Z., Song, Y., Wang, J., 2002. Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.* 74, 5487-5491.
- Dubois, D., Grare, M., Prere, M-F., Segonds, C., Marty, N., Oswald, E., 2012. Performances of the MALDI-TOF mass spectrometry system VITEK MS for the rapid identification of bacteria in routine clinical microbiology. *J. Clin. Microbiol.*, 50, 2568-2576.
- Dworzanski, J. P., Berwald, L., Meuzelaar, H. L. C., 1990. Pyrolytic methylation-gas chromatography of whole bacterial cells for rapid profiling of cellular fatty acids. *Appl. Environm. Microbiol.*, 56, 1717-1724.

- Dworzanski, J., Dickinson, D.N., Deshpande, S.V., Snyder, A.P., Eckenrode, B.A., 2010. Discrimination and phylogenomic classification of *Bacillus anthracis-cereus-thuringiensis* Strains, based on LC-MS/MS analysis of whole cell protein digests. *Anal. Chem.* 82, 145-155.
- Edward-Jones, V., Claydon, M.A., Evason, D.J., Walker, J., Fox, A.J., Gordon, D.B., 2000. Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. *J. Med. Micro.* 49, 295-300.
- Everley R. A, Croley T. R. 2008. Ultra-performance liquid chromatography/mass spectrometry of intact proteins. *J. Chromatograph. A.* 1192, 239–247.
- Fox A., Harley W., Feigley C., Salzberg D., Toole C., Sebastian A., Larsson L., 2005. Large particles are responsible for elevated bacterial marker levels in school air upon occupation. *J. Environ. Monit.* 7, 450 – 456.
- Fox K., Castanha E., Fox A., Feigley C., Salzberg, C., 2008. Human K10 epithelial keratin is the most abundant protein in airborne dust of both occupied and unoccupied school rooms. *J. Environ. Monit.* 10, 55–59.
- Fox, K., Fox, A., Elßner, T., Feigley, C., Salzberg, D., 2010. MALDI-TOF mass spectrometry speciation of staphylococci and their discrimination from micrococci isolated from indoor air of schoolrooms. *J. Environ. Monit.*, 12, 917-923.
- Fox, K., Fox, A., Rose, J., Walla, M., 2011. Speciation of coagulase negative staphylococci, isolated from indoor air, using SDS page gel bands of expressed proteins followed by MALDI-TOF MS and MALDI-TOF-TOF MS-MS analysis of tryptic peptides. *J. Microbiol. Meth.* 84, 243-250.
- Giebel, R., Worden, C., Rust, S.M., Kleinheinz, G.T., Robbins, M., Sandrin, T.R., 2010. Microbial Fingerprinting using matrix-assisted laser desorption Ionization time-of-flight mass spectrometry (MALDI-TOF MS): applications and challenges. *Advances Appl. Microbiol.* 71, 150-178.
- Holland, R.D., Wilkes, G., Rafii, F., Sutherland, J.B., Persons, C.C., Voorhees, K.J., Lay, J.O. Jr., 1996. Rapid Identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 10, 1227-1232.
- Holland, R.D., Duffy, C.R., Rafii, F., Sutherland, J.B., Heinze, T.M., Holder, C.L., Voorhees, K.J., Lay, J.O. Jr., 1999. Identification of Bacterial Proteins Observed In MALDI-TOF Mass Spectra from Whole Cells. *Anal. Chem.*, 71, 3226-3230.

- Hotta, Y., Sato, H., Hosoda, A., Tamura, H., 2012. MALDI-TOF analysis of ribosomal proteins coded in S10 and spc operons rapidly classified the *Sphingomonadaceae* as alkylphenol polyethoxylate-degrading bacteria from the environment. Federation of European Microbiological Societies. 330, 23-29.
- Ikryannikova, L.N., Lapin, K.N., Malakhova, M.V., Filimonova, A.V., Ilina, E.N., Dubovickaya, V.A., Sidorenko, S.V., Govorun, V.M., 2011. Misidentification of alpha-hemolytic streptococci by routine tests in clinical practice. Infect. Genet. Evol. 11, 1709-1715.
- Jabbour, R.E., Deshpande, S.V., Wade, M.M., Stanford, M.F., Wick, C.H., Zulich, A.W., Skowronski, E.W., Snyder, A.P., 2010. Double-Blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics. Appl. Environ. Microbiol. 76, 3637-3644.
- Karas, M., Bachmann, D., Hillenkamp, F., 1985. Influence of the wavelength in High-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. Anal. Chem. 57, 2935-2939.
- Karas, M., Hillenkamp, F., 1988. Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Anal. Chem. 60, 2299-2301.
- Karger, A., Ziller, M., Bettin, B., Mintel, B., Schares, S., Geue, L., 2010. Determination of Serotypes of Shiga Toxin-Producing *Escherichia coli* isolates by intact cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. 77, 896-905.
- Karlsson, R., Davidson, M., Svensson-Stadler, L., Karlsson, A., Olesen, K., Carlsohn, E., Moore, E.R.B., 2012. Strain-level typing and identification of bacteria using mass spectrometry-based proteomics. J. Proteome Res. 11, 2710-2720.
- Kliem, M., Sauer, S., 2012. The essence on mass spectrometry based microbial diagnostics. Curr. Opin. Microbiol. 15, 397-402.
- Kooker J.F., Fox K.F., Fox, A., 2012. Characterization of *Micrococcus* strains isolated from indoor air. Mol. Cell. Probes. 26, 1-5.
- Krishnamurthy, T., Ross, P.L., Rajamani, U., 1996. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time of flight mass spectrometry. Rapid Commun. Mass Spectrom. 10, 883-888.
- Krishnamurthy, T., Ross, P.L., 1996. Rapid Identification of bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. Rap. Commun. Mass Spectrom. 10, 1992-1996.

- Krishnamurthy, T., Davis, M.T., Stahl, D.C., Lee, T.D., 1999. Liquid chromatography/microspray mass spectrometry for bacterial investigations. *Rapid Commun. Mass Spectrom.* 13, 39-49.
- Madonna, A., Cuyk, S., Voorhees, K., 2003. Detection of *Escherichia coli* using immunomagnetic separation and bacteriophage amplification coupled with matrix assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 17, 257–263.
- Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P., Willems, A., 2008. Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *Int. J. System. Evol. Microbiol.* 58, 200-214.
- Martiny, D., Busson, L., Wybo, I., El Haj, R.A., Dediste, A., Vandenberg, O., 2012. Comparison of the Microflex LT and Vitek MS Systems for routine identification of bacteria by matrix-assisted laser desorption Ionization- time of flight mass Spectrom. *J. Clin. Microbiol.* 50, 1313-1325.
- Muroi, M., Shima, K., Nakagawa, Y., Tanamoto, K., 2011. Application of matrix-assisted laser desorption ionization-time of flight mass spectrometry for discrimination of *Echerichia* strains possessing highly conserved ribosomal RNA gene sequences. *Biol. Pharm.* 34, 430-432.
- Ram, R., VerBerkmoes, N., Thelen, M., Tyson, G., Baker, B., Blake, C., Shah, M., Hettich, R., Banfield, J., 2005. Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1919.
- Rezzonico, F., Vogel, G., Duffy, B., Tonolla, M., 2010. Application of whole-Cell matrix- assisted laser desorption ionization-time of flight mass spectrometry for rapid identification and clustering analysis of *Pantoea* species. *Appl. Environ. Microbiol.* 4497–4509.
- Risch, M., Radjenovic, D., Nam Han, J., Wydler, M., Nydegger, U., Risch, L., 2010. Comparison of MALDI-TOF with conventional identification of clinically relevant bacteria. *Swiss Med Wkly.* 140: w13095
- Saffert, R.T., Cunningham, S.A., Mandrekar, J., Patel, R., 2012. Comparison of three preparatory methods for detection of bacteremia by MALDI-TOF mass spectrometry. *Diag. Microbiol. Infectious Dis.* 73, 21-26.
- Seibold, E., Maier, T., Kostrzewa, M., Zeman, E., Splettstoesser, W., 2010. Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. *J. Clin. Microbiol.*, 48, 1061-1069.

- Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P-E, Rolain, J.M., Raoult, D., 2009. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Infect. Dis.* 49, 543-551.
- Shitikov, E., Ilina, E., Chernousova, L., Borovskaya, A., Rukin, I., Afanas'ev, M., Smirnova, T., Vorobyeva, A., Larionova, E., Andreevskaya, S., Kostrzewa, M., Govorun, V., 2012. Mass spectrometry based methods for the discrimination and typing of mycobacteria. *Infection, Gen. Evol.* 12, 838-845.
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., 1988. Protein polymer analysis up to m/z 100000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2, 151-153.
- Tang, K., Page, J.S., Marginean, I., Kelly, R.T., Smith, R.D., 2011. Improving liquid chromatography-mass spectrometry sensitivity using a subambient pressure Ionization with nanoelectrospray (SPIN) interface. *J. Am. Soc. Mass Spectrom.* 22, 1318-1325.
- Tani, A., Sahin, N., Matsuyama, Y., Enomoto, T., Nishimura, N., Yokota, A., Kimbara, K., 2012. High-throughput identification and screening of novel *Methylobacterium* species using whole-cell MALDI-TOF MS analysis. *PLoS One* 7, 1-13.
- Tran J., A. A. Doucette. 2009. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal. Chem.* 2009, 81, 6201-6209.
- Van Veen, S.Q., Claas, E.C.J., Kuijper, E.J., 2010. High-Throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories. *J. Clin. Microbiol.* 48, 900-907.
- Wall, D.B., Lubman, D.M., Flynn, S.J., 1999. Rapid profiling of induced proteins in bacteria using MALDI-TOF mass spectrometric detection of nonporous RP HPLC-separated whole cell lysates. *Anal. Chem.* 71, 3894-3900.
- Welkner, M., Moore, E.R.B., 2011. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *System. Appl. Microbiol.*, 34, 2-11.
- Werno, A.M., Christner, M., Anderson, T.P., Murdoch, D.R., 2012. Differentiation of *Streptococcus pneumoniae* from non-pneumococcal streptococci of the *Streptococcus mitis* group by MALDI-TOF MS. *J. Clin. Microbiol.*, 50, 2863-2867.

- Wolters, M., Rohde, H., Maier, T., Belmar-Campos, C., Franke, G., Scherpe, S., Aepfelbacher M., Christner, M., 2011. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. *Internat. J. Med. Microbiol.* 301, 64-68.
- Wynne, C., Fenselau, C., Demirev, P.A., Edwards, N., 2009. Top-Down identification of protein biomarkers in bacteria with unsequenced genomes. *Anal. Chem.* 81, 9633-9642.