ABSTRACT

Title of Document: INTEGRATION OF ASP-SPECIFIC MICROWAVE-ACCELERATED ACID HYDROLYSIS INTO PROTEOMIC ANALYSES

Stephen J. Swatkoski, Ph.D., 2007

Directed By: Dr. Catherine Fenselau, Department of Chemistry and Biochemistry

Presented in this work is a novel “bottom up” proteomics approach to protein identification and characterization that utilizes microwave-accelerated acid hydrolysis combined with mass spectrometric methods and bioinformatics. Results of this study demonstrate that this strategy is a robust alternative to residue specific enzymatic cleavage methods for generating peptides. Cleavage of proteins was shown to be site-specific at aspartate, as evaluated by matrix-assisted laser desorption ionization time of flight mass spectrometry and liquid chromatography tandem mass spectrometry. A bioinformatic analysis indicates that proteins will be cleaved at Asp to provide peptides that are longer than tryptic peptides, and which contain more basic residues, on average.

The feasibility of this digestion method was first demonstrated on a pure protein standard, ovalbumin, and further developed for applications in rapid microorganism identification, and whole organelle processing. Digestion of ovalbumin and analysis by mass spectrometry provided ~80% sequence

coverage. Bacillus spores, the RNA virus, bacteriophage MS2, and the DNA virus, human adenovirus type 5, were identifiable, based on the analysis digestion products generated by rapid digestion of protein biomarkers released by acid. Whole ribosomes isolated from *Saccharomyces cervasiae* were processed directly to peptides, which enabled identification of 58 of 79 ribosomal proteins by LC tandem mass spectrometry. Finally, this digestion method was shown to be compatible with a proteolytic ^{18}O labeling strategy, which enables rapid relative quantitation of proteins.

INTEGRATION OF ASP-SPECIFIC MICROWAVE-ACCELERATED ACID HYDROLYSIS INTO PROTEOMIC ANALYSES

By

Stephen James Swatkoski

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Catherine Fenselau, Chair
Professor Sang Bok Lee
Professor Neil Blough
Professor Dorothy Beckett
Professor Jonathan Dinman

Dedication

I dedicate this dissertation to my parents, John and Cindy, and my sister, Lisa. Their unwavering support enabled me to overcome all of my struggles and allowed me pursue all of my aspirations with confidence. Thank You.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Catherine Fenselau, for making my graduate school experience better than I had ever imagined. Her guidance has allowed me to endure much success both inside and outside of the laboratory, and for that, I will be forever grateful.

I would like to thank my colleagues, especially Dr. Scott Russell, Dr. Partick Pribil, Dr. Amir Rahbar, and Dr. Faith Hays. I appreciate all of their support, valuable advice, and most of all, their friendships.

I also need to thank Dr. Nathan Edwards, Dr. Peter Gutierrez, and Dr. Jonathan Dinman for the time and effort they put forward in several collaborations.

Lastly, I would like to acknowledge my Grandfather, both of my late Grandmothers, and Uncle Vic and Aunt Irene for providing me with their incredible wisdom. I would also like to thank my best friend, Mr. Brian Van Scoy, for being someone I can always count on.

Table of Contents

Dedication……...ii
Acknowledgements……...iii
Table of Contents.....iv
List of Abbreviations.....v
List of Tables……vi
List of Figures.....vii
Chapter 1: Introduction.....1
Proteomics……...1
Mass Spectrometry……3
Bioinformatics and Protein Identification……..13
Microwave-Accelerated Acid Hydrolysis…….....16
Hypothesis and Objectives……20
Chapter 2: Analysis of Ovalbumin…….....22
Introduction.....22
Materials and Methods…….....24
Results and Discussion……...25
Chapter 3: Rapid Analysis of Bacillus Spores……31
Introduction…….....31
Materials and Methods……...32
Results and Discussion……...25
Chapter 4: Analysis of Bacteriophage MS2……58
Introduction……....58
Materials and Methods……...59
Results and Discussion……...61
Chapter 5: Analysis of Human Adenovirus Type 5……...72
Introduction…….....72
Materials and Methods……73
Results and Discussion……...77
Chapter 6: Analysis of *Saccharomyces cerevisiae* Ribosomal Proteins…....88
Introduction…….....88
Materials and Methods…….....92
Results and Discussion……...94
Chapter 7: Relative Quantitation of Proteins……..……..108
Introduction…….....108
Materials and Methods……...109
Results and Discussion……..111
Chapter 8: Conclusions and Prospectus…….....131
Bibliography…….....134

List of Abbreviations

MALDI: Matrix Assisted Laser Desorption Ionization

ACN: Acetonitrile

Bt: *Bacillus turingiensis*

Ba: *Bacillus anthracis*

Bs: *Bacillus subtilis*

CFR: Curved Field Reflectron

CHCA: alpha-Cyano-4-hydroxycinnamic acid

ESI: Electrospray Ionization

Gst: *Geobacillus steareothermophilus*

LC MS/MS: Liquid chromatography tandem mass spectrometry

LID: Laser Induced Dissociation

NCBInr: National Center for Biotechnology Information (non redundant)

PMF: Peptide Mass Fingerprinting

Q-TOF: Quadrupole Time of Flight

RMIDB: Rapid Microorganism Identification Database

SASP: Small Acid Soluble Protein

TFA: Trifluoroacetic acid

TIC: Total Ion Chromatogram

TOF: Time of Flight

TIGR: The Institute for Genomic Research

TrEMBL: Translated European Molecular Biology Laboratory

XIC: Extracted Ion Chromatogram

List of Tables

Table 2.1: List of experimentally observed of Asp-specific peptides from ovalbumin……. 28

Table 3.1: Theoretical formic acid digest products from the SASP proteins from *B. subtilis* str. 168 and *B. anthracis* str. Sterne…….. 37

Table 3.2: MASCOT Search results for formic acid digest products from *B. anthracis* str. Sterne and *B. subtilis* str.168 against gram positive bacteria…….....41

Table 3.3: Search of digest peptide sequences against theoretical acid digest products from the SASP family of proteins 48

Table 3.4: MASCOT Search results for formic acid digest products from *B. anthracis* str. Sterne and *B. subtilis* str.168 against all entries, for the next nearest match…… 50

Table 4.1: List of experimentally observed peptide masses (Figure 1C) from the bacteriophage MS 2 capsid protein and the corresponding amino acid sequences assigned…….. 67

Table 5.1: MASCOT search results from LC MS/MS analysis of Human Adenovirus Type 5 digestion products 87

Table 6.1: Table of identified peptides from the *Saccharomyces cerevisiae* ribosome 99

Table 6.2: MASCOT search results from LC MALDI TOF-TOF analysis of high mass peptides (>3 kDa) 107

List of Figures

Figure 1.1: Schematic representation of the MALDI process 5

Figure 1.2: Schematic representation of the principle of TOF mass analysis…… 7

Figure 1.3: Schematic representation of a Reflectron TOF mass spectrometer…… 10

Figure 1.4: Schematic representation of the ESI process…… 11

Figure 1.5: Schematic representation of a Quadrupole Time of Flight mass spectrometer 12

Figure 1.6: Diagram of peptide cleavage at the amide bonds of the peptide backbone to generate b- and y- ions 15

Figure 1.7: Mechanisms for the acid hydrolysis of proteins at aspartic acid residues 18

Figure 1.8: Four potential products of internal Asp-specific cleavage of the protein, VEGLIDSSYMPFRKRGDVAPSS…. 19

Figure 2.1: Amino acid sequence of ovalbumin……. 23

Figure 2.2: Mass spectrum (m/z 1000-5000) of ovalbumin digestion products (top). Mass spectrum (m/z 5000-8400) of ovalbumin digestion products (bottom) 27

Figure 2.3: Mass spectrum of phosphorylated peptides from ovalbumin 29

Figure 3.1: Mass spectrum of products of a microwave assisted formic acid digestion of *B. anthracis* str. Sterne spores 39

Figure 3.2: Fragmentation spectra (LID) of peptides from the microwave assisted formic acid digestion of *B. anthracis* str. Sterne spores (a) fragments from m/z 3087.1 (b) m/z 3073.0 40

Figure 3.3: Mass spectrum of undigested SASPs from *B. anthracis* str. Sterne selectively solubilized with 12 % formic acid 43

Figure 3.4: Mass spectrum of products of a microwave assisted formic acid digestion *B. subtilis* str. 168 spores.....45

Figure 3.5: Fragmentation spectra (LID) of peptides from the microwave assisted formic acid digestion of *B. subtilis* str. 168 spores fragments from m/z 2213.8...46

Figure 3.6: Mass spectrum of of a 90 second microwave assisted formic acid digestion from *B. anthracis* str. Sterne ..... …52

Figure 3.7: Fragmentation spectra (LID) of peptide m/z = 2164.7 from the microwave assisted formic acid digestion of *B. anthracis* str. Sterne spores.. 54

Figure 3.8: Microwave assisted formic acid digestion carried out for 90 seconds on 1:1 mixture. (a) A representative spectrum showing the acid digest peptide products of *B. anthracis* str. Sterne and *B. subtilis* str. 168 and (b) expanded mass window showing observed intact proteins and peptide products..... …56

Figure 4.1: Mass spectrum of the bacteriophage MS2 suspension following 130 W (A) 160 W (B) 190 W (C) 220 W (D) of microwave irradiation for 15 sec. above 108° C.....63

Figure 4.2: Mass spectrum of bacteriophage MS2 suspended in 50% acetic acid/1% triton X-100.....65

Figure 4.3: Multiple sequence alignment of all coat protein sequences from Enterobacteria phage MS2 strains.....70

Figure 5.1: Superimposed mass spectra of “crude” Human adenovirus Type 5 (blue) and “pure” Human adenovirus Type 5 (black) both treated with a 1:1 mixture of 1% Triton X-100 and 50% Acetic acid from m/z 11500-20000 (A) 20000-50000 (B) and 50000-110000. 78

Figure 5.2: Mass spectrum of “crude” Human adenovirus Type 5 following a 90 sec. digestion with acetic acid (A) undigested (B).....80

Figure 5.3: Fragmentaion spectrum (LID) of m/z = 1182.8. 81

Figure 5.4: Fragmentation spectrum (LID) of m/z = 1196.8. 83

Figure 5.5: TIC (A) TOF Mass Spectrum (B) XIC (C) and MS/MS mass spectrum of m/z 903.84 (D) from HPLC Q-TOF analysis……..86

Figure 6.1: Distribution of peptide lengths from tryptic cleavage (A) Asp-specific cleavage (B)……90

Figure 6.2: Distribution of basic residues from tryptic cleavage (A) Asp-specific cleavage (B)……91

Figure 6.3: MALDI TOF mass spectrum (m/z 800-8000) of ribosomal protein digestion products (A). MALDI TOF mass spectrum (m/z 1800-2900) of ribosomal protein digestion products 96

Figure 6.4: LTQ-Orbitrap TIC of ribosomal protein digestion products..... 97

Figure 6.5: LTQ-Orbitrap tandem mass spectrum of a doubly charged peptide (m/z 600.85)..... 98

Figure 6.6: MALDI TOF-TOF fragmentation spectrum of m/z = 3172.3 106

Figure 7.1: Mass spectrum of a the peptide, STRTQINKVVRFD, $[M+H]^+_{calc.}$ = 1563.9 after digestion with 12.5% acetic acid in $H_2{}^{16}O$ (A) and $H_2{}^{18}O$ (B) for 90 sec 112

Figure 7.2: Theoretical isotope cluster for the peptide, STRTQINKVVRFD, $[M+H]^+_{calc.}$ = 1563.9 from ovalbumin 113

Figure 7.3: LID Mass spectrum of a the peptide, AQGAMTKALELFRND, $[M+H]^+_{calc.}$ = 1664.8 following 5 min digestion with 25% acetic acid. Inset is a list of all predicted fragment ion masses..... 115

Figure 7.4: Mass spectrum of the C-terminal myoglobin peptide, IAAKYKELGFQG, $[M+H]^+_{calc.}$ = 1324.8…..... 116

Figure 7.5: Theoretical isotope cluster for the peptide, IAAKYKELGFQG, $[M+H]^+_{calc.}$ = 1324.8 from myoglobin..... 117

Figure 7.6: Mass spectrum of a peptide pair from ovalbumin generated from a 90 sec (A), 3min (B), and 5 min (C) digest with 12.5% acetic acid 119

Figure 7.7: Mass spectrum of a the peptide, STRTQINKVVRFD, $[M+H]^+_{calc.}$ = 1563.9 from a 1:1 (A) and 2:1 (B) mixtures.....121

Figure 7.8: Schematic representation of experimental protocol for the chemical digestion and ^{18}O labeling of Bacillus spores.....123

Figure 7.9: Mass spectrum of the peptide, ARSTNKLAVPGAESALD, $[M+H]^+_{calc.}$ = 1699.9, from SASP B from *B.cereus* following digestion in $H_2{}^{16}O$ (A). Mass spectrum of the same peptide from *B.anthracis* Sterne (B) following digestion in $H_2{}^{18}O$….....124

Figure 7.10: Mass spectrum of the peptide, ARSTNKLAVPGAESALD, $[M+H]^+_{calc.}$ = 1699.9, from SASP B from *B.cereus* and *B.anthracis* Sterne mixed 1:1(A) and 2:1 (B).....126

Figure 7.11: Mass spectrum of the peptide, PEYWNFRNGD, $[M+H]^+_{calc.}$ = 1297.6, from the chemical digestion of the fiber protein from a 24 hour culture of adenovirus (A) from a 48 hour culture of adenovirus (B)……...128

Figure 7.12: Mass spectrum of the peptide, PEYWNFRNGD, $[M+H]^+_{calc.}$ = 1297.6, from the adenovirus fiber protein after combining the 24 hour (unlabeled) and 48 hour (labeled) cultures 1:1..……..……129

Figure 7.13 Theoretical isotope clusters of a 1:2 (unlabled:labeled) mixture of the peptide, PEYWNFRNGD, $[M+H]^+_{calc.}$ = 1297.6 from the adenovirus fiber protein……130

Chapter 1: Introduction

Proteomics

Proteomics can be defined as the study of the total protein complement of the genome.[1] Over the past decade, proteomics has been utilized to help provide a better understanding of a variety of biological phenomena including changes in protein expression levels, protein-protein interactions, and post-translational modifications. As a result, proteomics has had a profound affect, helping researchers understand how biological environments can affect the proteins within a given system. Recent technological advancements in protein separation techniques, mass spectrometric methods and bioinformatics has made large-scale analysis of proteins from plasma, biopsies, whole cell lysates, fractionated cellular organelles, plants, and intact microorganism samples possible.

Perhaps the most important component of proteomics research is the identification and characterization of those proteins that undergo change. In proteomics research, proteins are identified and characterized using two different strategies, termed “bottom-up” or “top-down” proteomics. While either of these approaches can be employed independently, the coupling of both strategies can be advantageous because they can provide complementary information about the protein(s) of interest.

In a “bottom-up” proteomic analysis, proteins are characterized based on information provided by their constituent peptides generated by enzymatic or chemical cleavage. A variety of enzymes and chemical agents is available for the

cleavage of proteins. However, the use of a cleavage agent that exhibits high digestion specificity greatly facilitates peptide and protein identification. Two "bottom-up" approaches are most commonly used for the large-scale analysis of protein mixtures. One approach involves first separating of the protein mixture by gel electrophoresis and excising the protein spot from the gel. The protein is then subjected to a site-specific proteolytic digestion, usually with trypsin, and the peptide products are analyzed by mass spectrometry to identify the proteins. Alternatively, the entire protein mixture is digested prior to separation and the resulting peptide mixture is then analyzed by multi-step separations and liquid chromatography tandem mass spectrometry (LC-MS-MS) to obtain peptide and protein identifications.

When using a "top-down" approach, the intact proteins are studied directly. The proteins are identified based on their intact molecular weights in combination with partial protein sequence information provided by MS/MS techniques.[2] "Top-down" proteomics can be a powerful technique for protein characterization and identification, but requires complex instrumentation and laborious interpretation of complicated MS/MS mass spectra.

Proteomics has also been shown to provide valuable quantitative information about the proteins within a given sample. In comparative proteomic studies, relative quantitation of protein abundance levels can provide important information about differences in concentration of the proteins in two separate protein pools (e.g. healthy and diseased). The ability to measure differences in protein levels has led to a better understanding of the relationships between

protein functions and the biological conditions within a given system. Moreover, detection of proteins with altered abundances has facilitated the discovery of new therapeutic targets. [3]

Various mass-spectrometry based labeling strategies that involve isotope tagging for the relative quantitation of proteins have been proposed. Using these approachs, isotope ratio measurements of two identical peptides from two separate peptide pools, unlabeled and labeled, are obtained to generate quantitative information about the difference in protein abundances between the two samples. [4] Alternatively, gel-based techniques can be used to obtain quantitative information in a comparative proteomic study. Using this approach, the intensities of same gel spot on two different gels from different biological samples are compared to measure differences in protein abundance. [5]

Mass Spectrometry

Mass spectrometry has become the foundation technique in proteomic research. Since the advent of two soft ionization techniques, matrix assisted laser desorption ionization[6, 7] (MALDI) and electrospray ionization [8] (ESI), mass spectrometry has proved to be well suited for the analysis of large biological molecules such as proteins and peptides. Once a biological molecule is ionized, mass spectrometry provides an unparalleled means of accurately measuring its molecular weight and in the case of peptides, its primary structure.

MALDI requires that the sample analyte first be mixed and co-crystallized with a matrix, which is a chromophoric organic acid present in a large molar excess. The function of the matrix during the ionization process is to (a) absorb

laser energy (b) isolate the sample molecules, and (c) provide photoexcited acidic or basic sites for ionization of the sample molecules.[9] Upon absorption of the laser energy by the matrix, the sample and matrix molecules expand into a gaseous plume. The sample molecules are ionized by proton transfer from the photoexcited matrix molecules in both the solid and gaseous states.[8] Perhaps the most appealing aspect of MALDI is the ability to produce high molecular weight ions in the gas phase without significant thermal degredation.[10] MALDI is most compatible with rapid analyses that require minimal sample preparation. MALDI has been shown to produce high quality mass spectra in the presence of high salt concentrations.[11] Furthermore, MALDI can tolerate the addition of strong acid and/or low concentrations of detergents that are commonly used to solubilize proteins prior to analysis. **Figure 1.1** is a schematic representation of the MALDI process.

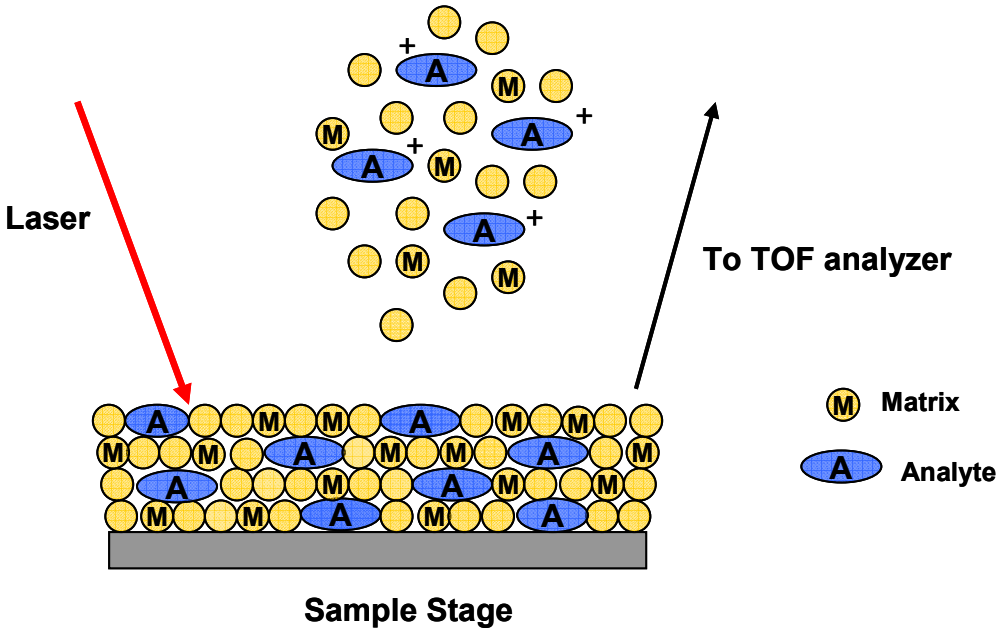


Figure 1.1: Schematic representation of the MALDI process.

MALDI is most commonly coupled with a time-of-flight (TOF) mass analyzer. A TOF mass analyzer is perhaps the most simplistic of all mass analyzers used in mass spectrometry. The fundamental principle behind TOF mass spectrometry is that ions are separated based on the amount of time required for them to reach the detector. After ionization, the ions are accelerated to a constant kinetic energy by the application of an accelerating potential, V. When the ions reach the field-free drift region, they traveling at a velocity, *v,* which is related to their m/z values by the following equation:

$$v = \left(\frac{2zV}{m} \right)^{1/2}$$

where, v, is ion's velocity; z is the charge on the ion, m is the mass of the ion, and V is the accelerating potential.[12] As a result, lighter ions will arrive at the detector first, followed by heavier ions.[10] **Figure 1.2** is a schematic representation of the principle of TOF mass spectrometry.

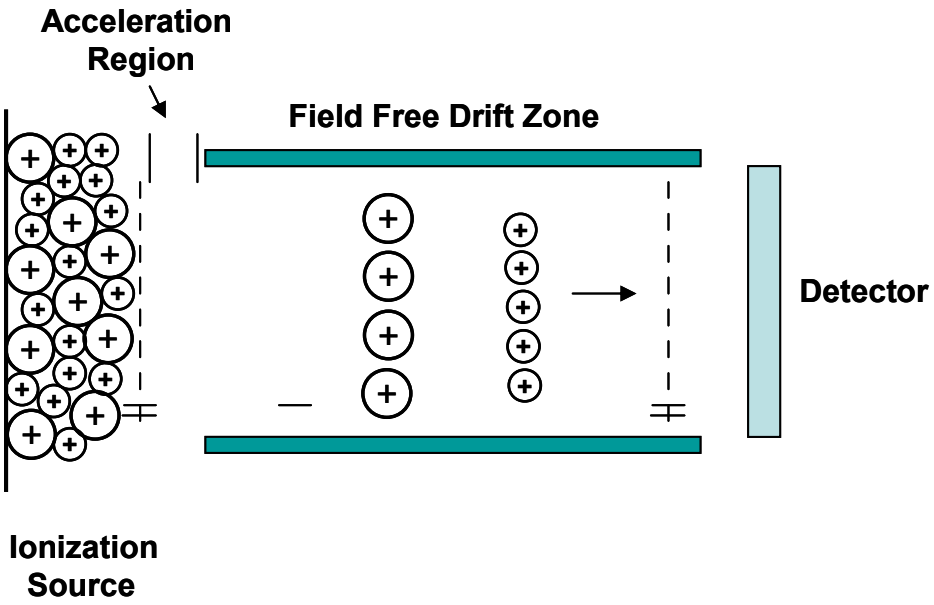


Figure 1.2: Schematic representation of the principle of TOF mass analysis. (Adapted from reference 10)

Some TOF mass analyzers are also equipped with a curved field reflectron, which helps to improve the resolution of TOF measurements. A reflectron provides improved resolution by compensating for the slight differences in the kinetic energies of ions with the same mass, after they have entered the flight tube. The reflectron serves to ensure the ions of the same mass reach the detector at the same time.[12] **Figure 1.3** is a schematic of a reflectron time-of-flight instrument. TOF instruments equipped with a reflectron can be used to acquire important structural information of precursor ions through the analysis of fragment ions generated by a process known as laser induced dissociation (LID). LID of precursor ions occurs because of excess energy introduced by the ionization process and/or because of opportunistic collisions with residual gases in the vacuum chamber. [12]

In contrast to the MALDI process, ESI produces gaseous ions from analytes in the liquid phase. The sample is typically mixed with an electrospray solvent and sprayed in the presence of an electric field at atmospheric pressure. As a result, charged, solvated droplets are produced, which entrain the sample molecules. During the electrospray process, solvent is evaporated from the charged droplets when a drying gas of either N_2 or He is applied, resulting in the production of molecular ions in the gaseous state. These are steered into a high vacuum region and the mass spectrometer. ESI is the most common ionization method employed when liquid chromatography (LC) is used for separation of analytes before introduction into the mass spectrometer. **Figure 1.4** is a schematic representation of the ESI process.

A quadrupole time of flight (Q-TOF) mass analyzer is frequently combined with ESI source because of its excellent MS/MS capabilities. This hybridization of mass analyzers enables accurate TOF mass measurements of precursor ions as well as precursor ion selection followed by collisionally induced dissociation (CID). **Figure 1.5** is a schematic representation of a Q-TOF mass spectrometer. As ions enter a Q-TOF instrument, a precursor ion is selected by the first quadrupole, Q1, and subsequently fragmented by collisionally induced dissociation in the second quadrupole, Q2. Accurate molecular weight measurements of the product ions are then made by the TOF mass analyzer.

A more recently developed hybrid mass analyzer; the linear iontrap/orbitrap, can alternatively be combined with LC ESI interfaces to obtain high resolution, high accuracy MS/MS measurements of peptides. In an LTQ-orbitrap instrument, precursor ions are measured with high accuracy by the orbitrap and fragmentation is carried out in the linear ion trap.

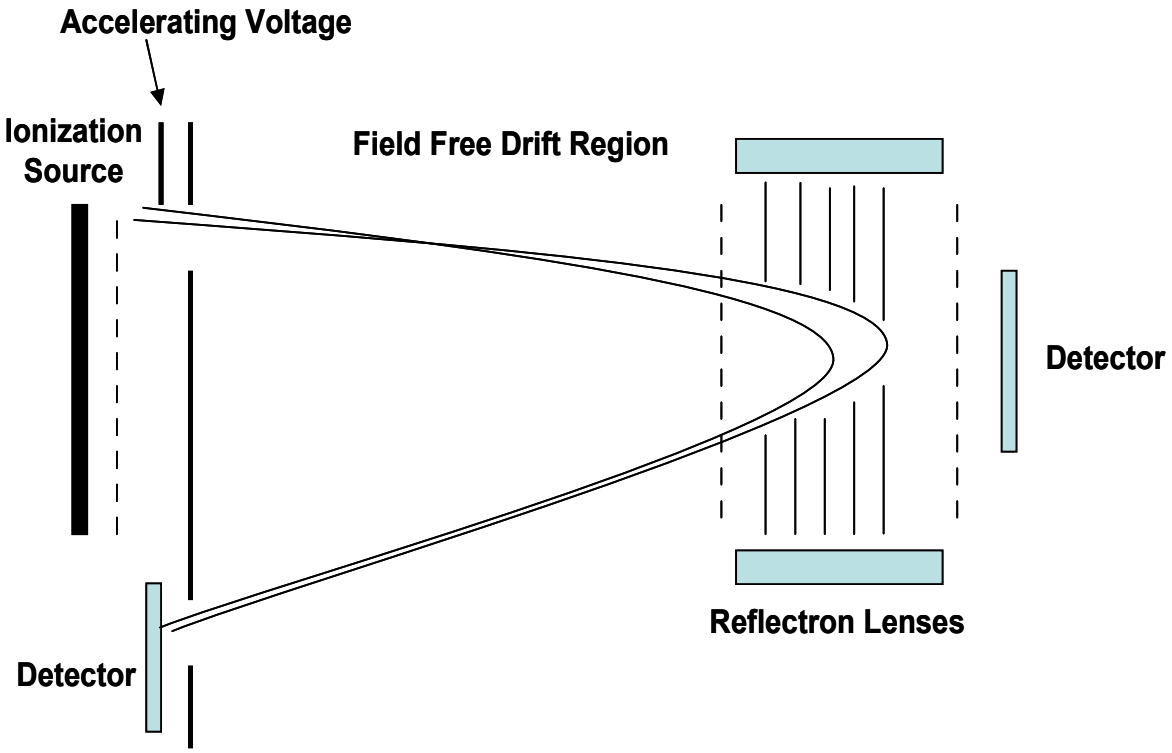


Figure 1.3: Schematic representation of a Reflectron TOF mass spectrometer. (Adapted from reference 10)

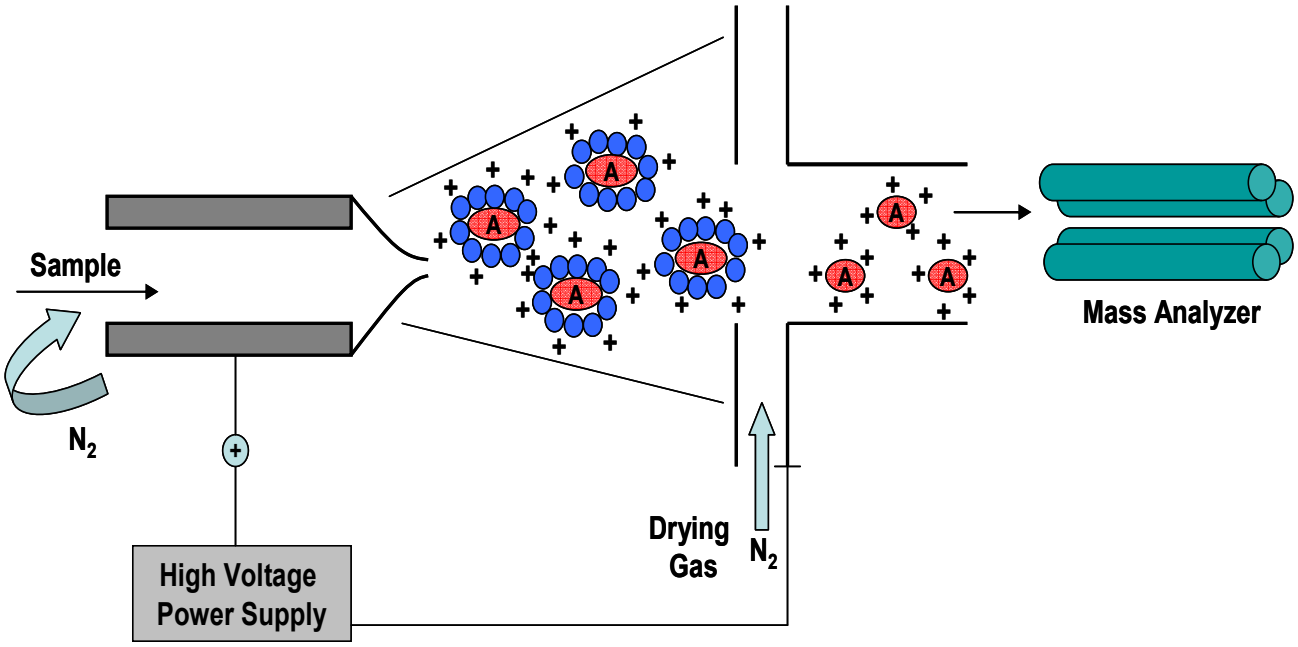


Figure 1.4: Schematic representation of the ESI process.

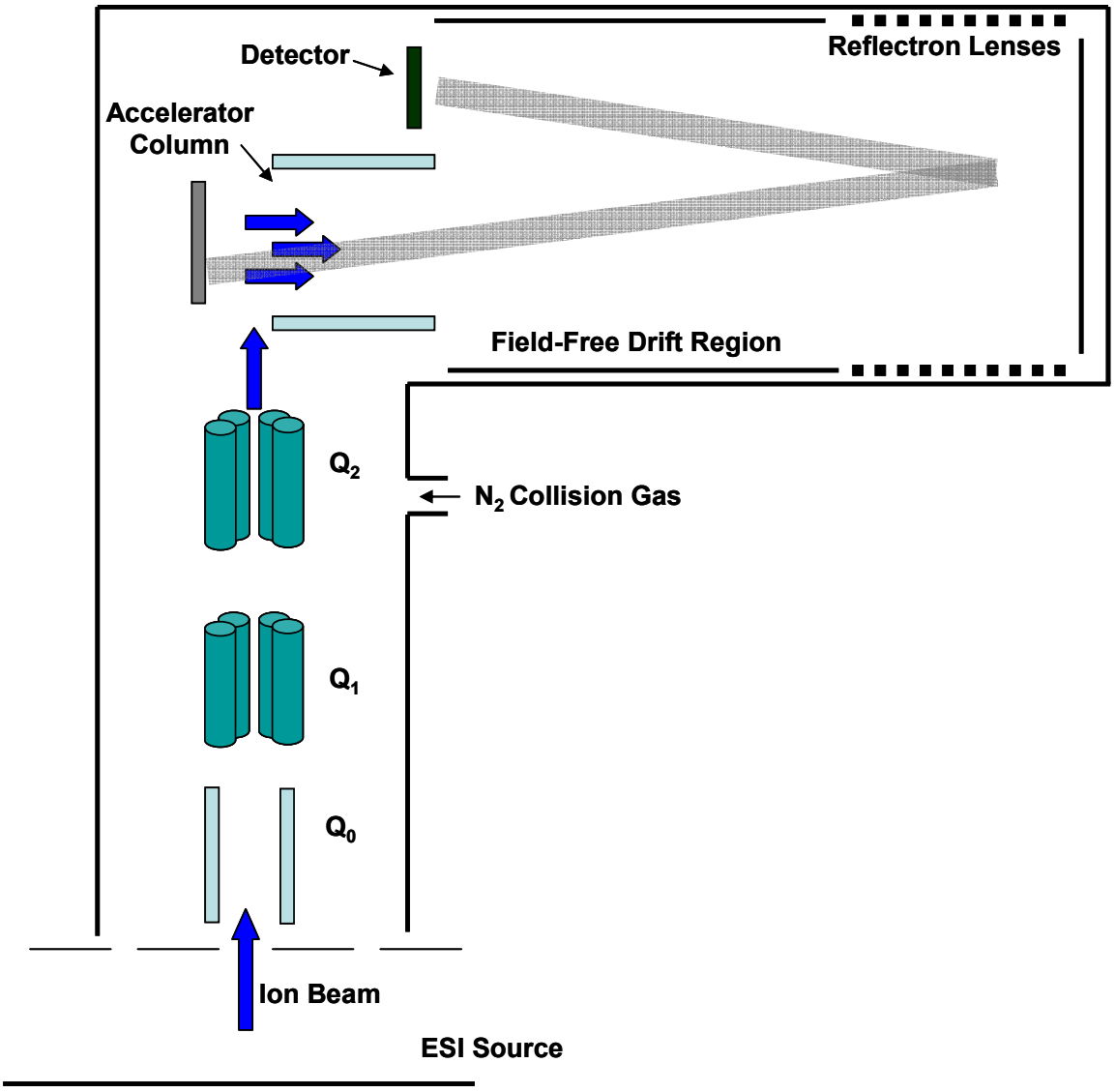


Figure 1.5: Schematic representation of a Quadrupole Time of Flight mass spectrometer. (Adapted from reference 10)

Bioinformatics and Protein Identification

Protein identification using mass spectrometric based techniques can be accomplished using two different strategies, peptide mass fingerprinting (PMF) or microsequencing. PMF is ordinarily used to identify only a single protein or a very simple protein mixture. This approach is commonly used for identification of proteins isolated from gel spots. The protein is identified by correctly matching molecular masses from a peptide mixture to theoretical masses predicted for each entry of a protein database.

Protein identification using microsequencing can be performed using two methodologies, “MS/MS spectral searching” or “sequence tagging”. Unlike peptide mass fingerprinting, this approach is well suited for the analysis of peptides resulting from the digestion of protein mixtures. An attractive feature of microsequencing and of MS/MS spectral searching is that a protein can be confidently identified based on the reliable identification of only a single peptide.

MS/MS data can provide valuable information about a precursor peptide because fragmentation patterns induced by low-energy gas phase collisions are predictable.[13] Low energy collisions have been shown to cause fragmentation along the peptide backbone at amide bonds, which creates a ladder sequence of b- and y-ions. B-ions refer to those fragments that retain a charge on the n-terminus, whereas y-ions are those that retain charge on the c-terminus. [13] **Figure 1.6** depicts the fragmentation pattern of a theoretical peptide. Protein identification by MS/MS ion searching requires that the entire MS/MS mass

spectrum combined with mass of the precursor and any cleavage specificity be submitted to the protein database. The database then calculates all possible peptides with the given precursor mass that can be produced with a specified enzyme. The fragment ion masses of those peptides are then compared to the fragment ion mass spectra obtained experimentally to identify the best match.[14]

Alternatively, the protein may be identified using sequence tagging. When using sequence tagging, a partial, manually interpreted amino acid sequence is submitted to the database in addition to the precursor peptide mass. When combined with the enzyme specificity, enough information is provided to locate potential peptide candidates from a sequence database. Once the candidates are chosen, their predicted fragment ions can compared to those observed experimentally to verify a match.[15]

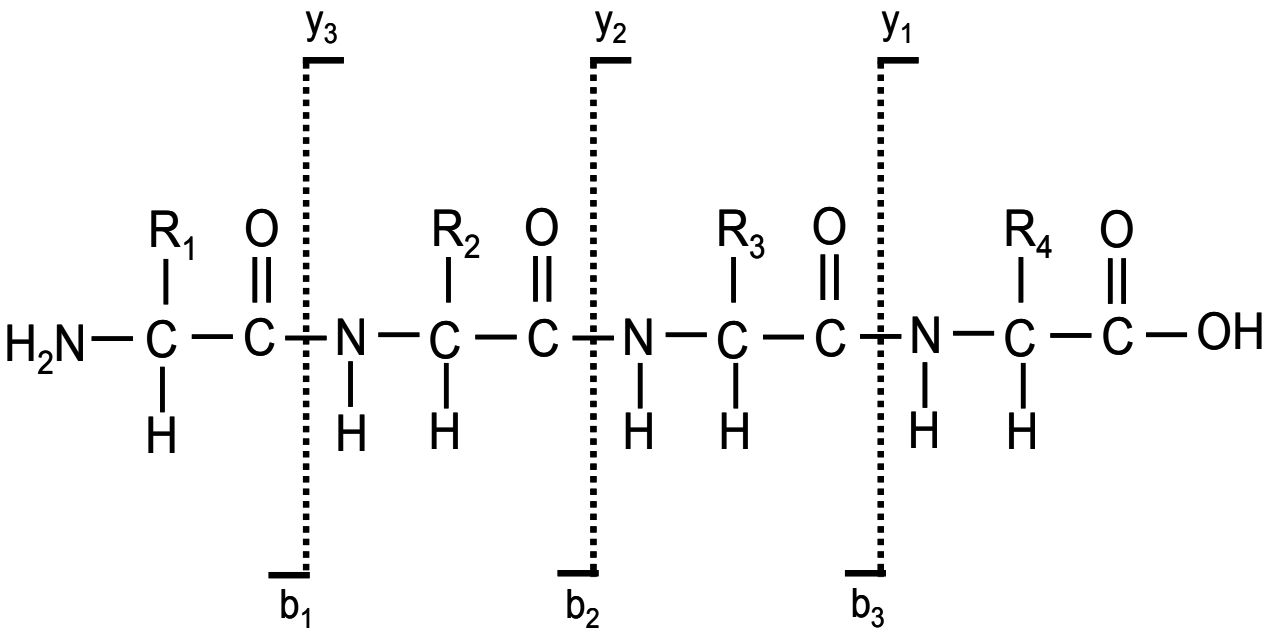


Figure 1.6: Diagram of peptide cleavage at the amide bonds of the peptide backbone to generate b- and y- ions.

Microwave-Accelerated Acid Hydrolysis

The most commonly used protein cleavage agent is trypsin because of its high specificity, the distribution of lysine and arginine residues, and the advantages of localizing charge on the C-terminus of peptide products.[16] Other useful enzymes include Glu-C protease, Arg-C protease and Lys-C protease. Several alternative methods to enzymatic digestion have been proposed for use in "bottom-up" proteomics studies. Of significant interest is the use of acids to generate peptide products for analysis. Various reports in the literature have shown that proteins, in the presence of dilute acid, undergo preferential cleavage at aspartic acid (Asp) residues. Shultz first reported that the peptide bonds of aspartic acid residues are cleaved at a rate at least 100 times greater than other bonds.[17] In conjunction with the results of this work, Inglis later reported that the primary requirement to achieve cleavage is to maintain a pH below the pKa = 2.1 of the side chain of the carboxyl group of aspartic acid residues.[18] In addition, he found that temperatures exceeding 108º C significantly accelerate the rate of Asp cleavage. Inglis proposed that the reaction for Asp cleavage can proceed via two possible pathways, at different rates, which can result in the cleavage on either the N- or C-terminal sides of Asp residues.[18, 19] The results of these studies also indicated that this cleavage mechanism is independent of type of acid used. **Figure 1.7** displays the reaction pathways for Asp cleavage. In addition, cleavage may also occur at both the N- and C-termini, resulting in the removal of the Asp residue. The possible cleavage products of a model protein, VEGLIDSSYMPFRKRGDVAPSS are shown in **Figure 1.8**.

Until recently, conventional thermal sources were used to accelerate the reaction for Asp cleavage of proteins. However, several recent reports have shown that microwave irradiation can reduce digestion times from 2 hr to only a few minutes.[20, 21] These studies used high concentrations of TFA and HCl, which resulted in the production of many nonspecific cleavage products. This resulted in an increase in mass spectral complexity which caused ambiguous interpretations of digestion product mass spectra.[20, 21] Recently, dilute formic acid (3%) combined microwave heating was reported to produce a greater amount of Asp-specific cleavages.[22] However, it is speculated that excessive digestion times also caused a significant number of cleavages at other amino acid positions. Prior to the analyses detailed in this dissertation, the suitability of the peptide products generated by Asp specific acid cleavage for integration into high throughput proteomic strategies has not been evaluated theoretically or experimentally.

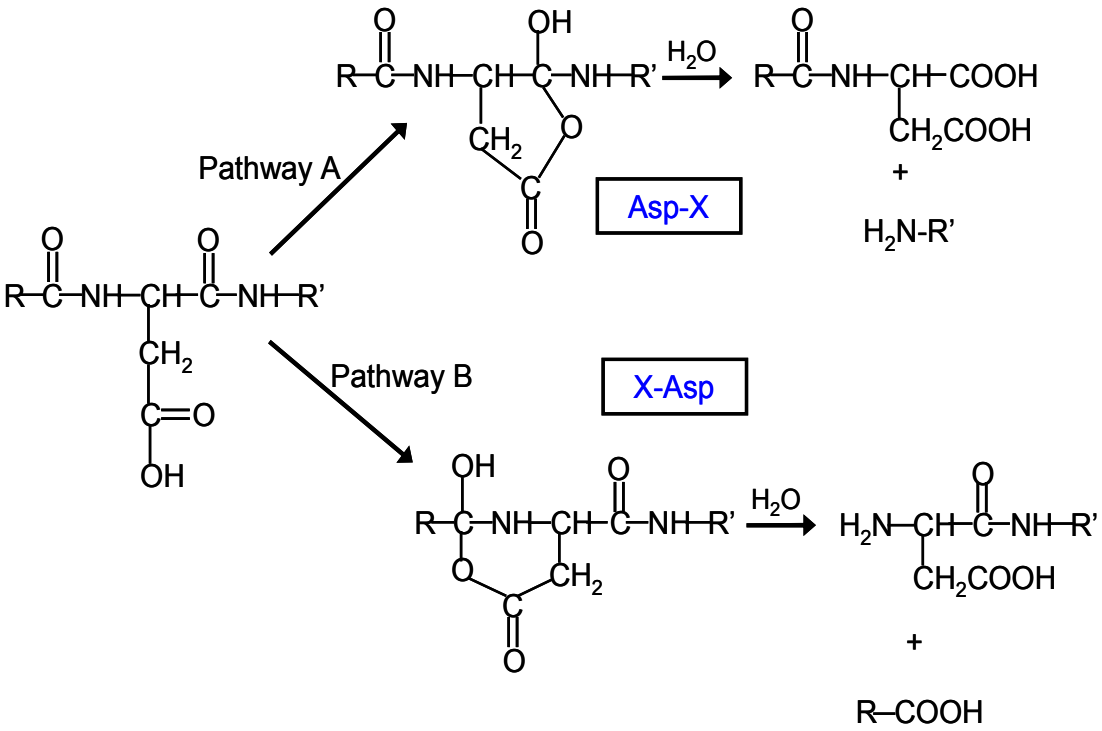


Figure 1.7: Mechanisms for the acid hydrolysis of proteins at aspartic acid residues. (Adapted from Reference 19)

1. DSSYMPFRKRGD
2. SSYMPFRKRGD
3. DSSYMPFRKRG
4. SSYMPFRKRG

Figure 1.8: Four potential products of internal Asp-specific cleavage of the protein, VEGLIDSSYMPFRKRGDVAPSS. Asp residues appear in red and are underlined.

Hypothesis and Objectives

Given that an organism's genome is considered to be rather static, it has become apparent that in order to better understand the complexity of biological activity within a given system, the dynamics of its proteome must be studied extensively. Thus, as the field of proteomics continues to grow, so does the need for new developments in proteomics technologies and sample preparation techniques to facilitate protein identification. One major focus in the proteomics community is developing new amino acid specific cleavage methods to generate peptides. While trypsin has become the most widely used proteolytic cleavage agent, enzymatic digestion has several limitations. Examples of these difficulties include a) inefficient digestion of hydrophobic and very basic proteins; b) production of protease autolysis products which can complicate mass spectra and c) chemical noise resulting from buffer contamination.[19] Furthermore, while enzymes perform well in a control laboratory setting, their application to highly automated and fieldable analyses can be limited due to uncontrollable environmental conditions.

Development of a reliable and efficient chemical digestion method for use in proteomic analyses may offer a more forgiving and flexible set of experimental constraints. Therefore, the objective of this study is to develop and evaluate a novel, non-enzymatic, digestion strategy for the identification of proteins from a variety of biological samples, which exhibits high specificity, and is suitable for use in automated proteomic workflows. This method relies on the rapid digestion of proteins at aspartic acid residues using microwave-accelerated acid

hydrolysis. In order to test Asp-specific microwave-accelerated acid hydrolysis as part of proteomic strategies, it will be utilized for a) the characterization of pure proteins, b) selective solubilization and subsequent digestion of protein biomarkers for microorganism identification c) processing of intact subcellular organelles into proteotypic peptides and d) relative quantitation of proteins using stable isotope labeling.

Chapter 2: Analysis of Ovalbumin

Introduction

Microwave-accelerated acid hydrolysis was first applied to the pure protein standard, ovalbumin, to evaluate its use as an efficient and specific cleavage method. Ovalbumin is a 385 amino acid glycoprotein with a calculated molecular weight of 42750 Da., based on its amino acid sequence. Ovalbumin has two potential glycosylation sites at ^{292}Asn and ^{311}Asn. However, only ^{292}Asn is known to be glycosylated.[23-25] In addition to glycosylation, ovalbumin is known to contain several other post-translational modifications, which include: a) a disulfide bond between ^{74}Cys and ^{121}Cys, b) two potential phosphorylation sites at ^{69}Ser and ^{345}Ser and C) an acetylated glycine residue at ^{1}Gly.[26] **Figure 2.1** shows the amino acid sequence of ovalbumin. The post-translational modifications present in ovalbumin make it an ideal model protein to evaluate the efficiency of microwave-accelerated acid hydrolysis for characterization of proteins by mass spectrometry-based proteomic techniques.

Figure 2.1: Amino acid sequence of ovalbumin. Modifications are highlighted.

Materials and Methods

Materials.

Glacial acetic acid was obtained from Fischer Scientific (Fair Lawn, NJ). α-cyanohydroxycinnamic acid (CHCA) and albumin, isolated from chicken egg, were obtained directly from Sigma (St. Louis, MO). Samples were reported to be a minimum of 98% pure by agarose gel electrophoresis. Protein solutions were prepared at 0.1 mg/ml in Mill-Q grade water.

Microwave-Accelerated Acid Hydrolysis.

Microwave-accelerated acid hydrolysis of ovalbumin was carried out in a Discover Benchmate (CEM Corp., Matthews, NC) microwave system equipped with 45 ml digestion vessel and a fiber optic temperature probe. A 43.8 µL aliquot of the protein solution was acidified to 12.5% with 6.3 µL glacial acetic acid in a 200 µl glass sample holder and placed in the digestion vessel. All digestions were carried out at a constant temperature of 140 ± 5 ˚C for 5 min.

MALDI-TOF Mass Spectrometry.

For the analysis of ovalbumin digestion products, 0.5 µL of the peptide mixture was applied to the sample plate and allowed to air dry. 0.5 µL of saturated α-cyano-4-hydroxycinnamic acid at 10 mg/mL in 70 % ACN and 0.1 % TFA was then applied to the sample plate and allowed to air dry. Peptide mass spectra were collected in Linear mode on an Axima CFR Plus MALDI-TOF mass

spectrometer (Shimadzu, Columbia,MD). All spectra were acquired with a 337 nm N_2 laser as averages of 200 profiles for TOF scans.

Database Searching

An “in house” version of MASCOT was used to carry out all database searches following data collection. The “formic acid” cleavage option was edited from its original form, which allows for cleavage on only the C-terminal side of Asp residues, to allow for cleavage on both the N- and C-termini of Asp residues.[27]This change allowed for the database to accommodate peptides that were generated by any of the proposed cleavage pathways described in the **Introduction**. All masses labeled in figure 13 were submitted as a peptide mass fingerprint with a peptide tolerance of 2.0 D., allowing for up to 7 missed cleavages. The masses were searched against all entries of the Swissprot database. “Acetylation” and “pyro-glutamine” were selected as “possible modifications” in the MASCOT search window.

Results and Discussion

Ovalbumin contains 14 potential aspartic acid cleavage sites. Thus, Asp-specific digestion of ovalbumin *in silico* produces fifteen potential peptide products without any missed cleavages. **Figure 2.2** is a representative mass spectrum of ovalbumin, obtained following a 5 min microwave-accelerated acid digestion with acetic acid. As evaluated by mass spectrometry, the digestion provided high specificity for Asp residues. Twenty-nine peptides were observed in the mass spectra that correlated well with predicted masses of Asp-specific

peptides generated *in silico.* The observed peptides in **Figure 2.2** confirm that cleavage occurred at all 14 aspartic acid cleavage sites and provided 84% sequence coverage. Interestingly, twelve peptide pairs, separated by 115 Da., were observed, which indicate that cleavage occurred on both the N- and C-termini of the aspartic acid residues. The observation of peptide pairs proved advantageous for assigning digestion products and helped confirm Asp-specificity.

Table 2.1 provides a list of the experimentally observed peptide masses, their corresponding calculated masses and amino acid sequences, and the length of the amino acid sequence. Parentheses indicate that aspartic acid residues retained on internal peptides can reside on either the N- or C-termini.

As indicated in **Table 2.1**, several masses were assigned to Asp-specific peptides by accounting for known post-translation modifications. The peptide at 1399.5 corresponds to a sodiated (Na^+) form of the peptide, 1-13, which is acetylated at ^{1}Gly. The peptide pairs at m/z = 5198 and m/z = 5082, m/z = 6743 and m/z = 6627, and m/z = 7458 and 7344 correspond to the acetylated forms of the peptides, 1-47(46), 1-60(59), and 1-67(66), respectively, with and without an Asp retained on the C-terminus. The peptide observed at m/z = 4636, corresponds to the phosphorylated form (+80 Da) of the peptide, 309-345, which was observed as a peptide pair at m/z = 4556 and m/z = 4441. **Figure 2.3** shows the region of the spectrum where the phosphorylation site was detected.

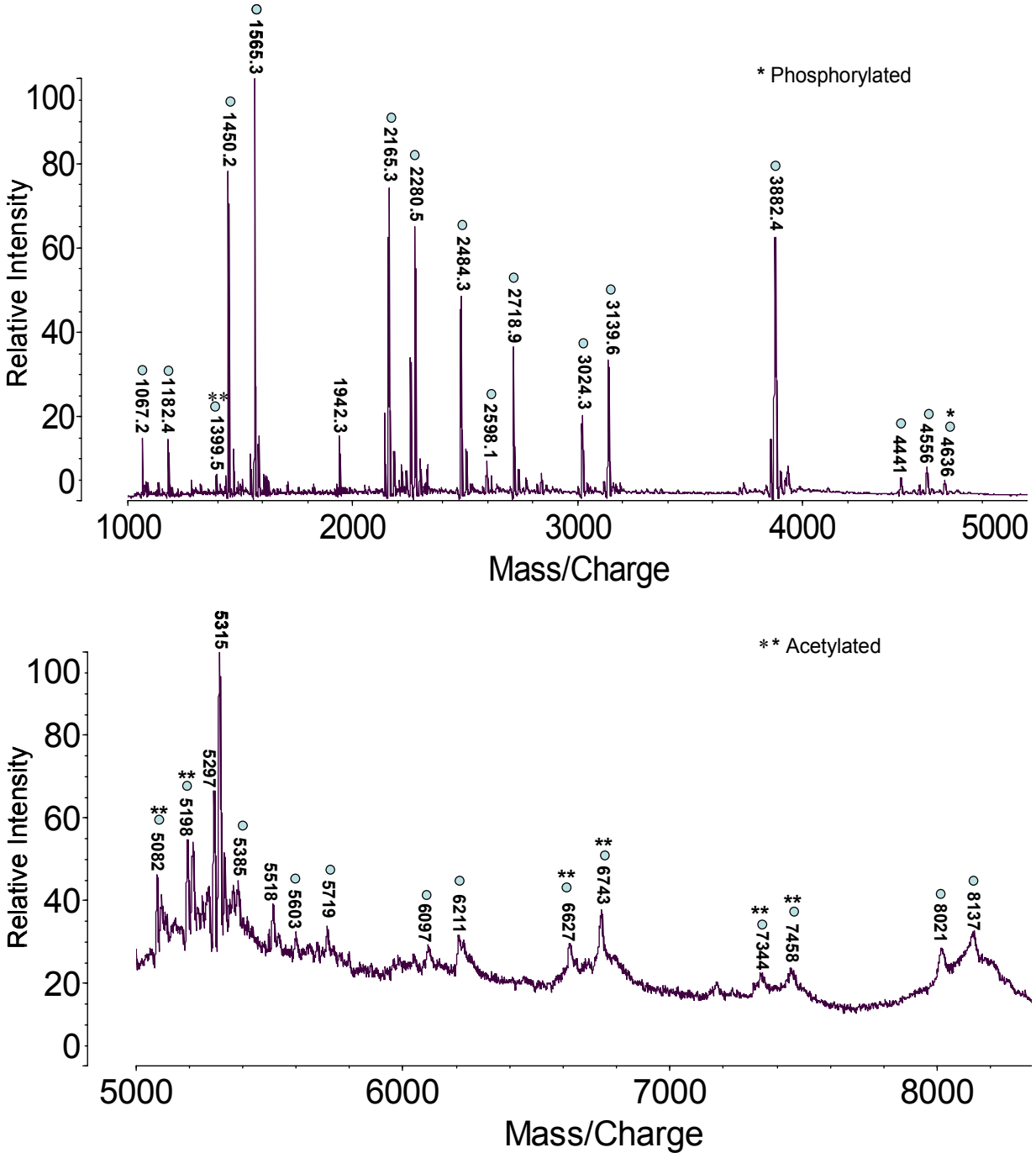


Figure 2.2: Mass spectrum (m/z 1000-5000) of ovalbumin digestion products (top). Mass spectrum (m/z 5000-8400) of ovalbumin digestion products (bottom). Green circles indicate mass matches to predicted Asp-specific peptides.

[M+H]+ calc.	[M+H]+ obs.	AA Position	Amino Acid Sequence
1182.2	1182.4	350-360, 351-361	(D)AASVSEEFRA(D)
1067.2	1067.2	351-360	AASVSEEFRA
1335.4	1399.5** +Na	1-13	GSIGAASMEFCFD
1564.8	1565.3	47-59, 48-60	(D)STRTQINKVVRF(D)
1449.8	1450.2	48-59	STRTQINKVVRF
2279.6	2280.5	47-66,48-67	(D)STRTQINKVVRFDKLPGFG(D)
2164.9	2165.3	48-66	STRTQINKVVRFDKLPGFG
2597.1	2598.1	167-189, 168-190	(D)SQTAMVLVNAIVFKGLWEK AFK(D)
2482.1	2484.3	168-189	SQTAMVLVNAIVFKGLWEKAFK
2719.4	2718.9	362-385	HPFLFCIKHIATNAVLFFG RCVSP
3155.5	3139.6	139-166	pyro-EARELINSWVESQTNGIIR NVLQPSSVD
3040.5	3024.3	139-165	pyro-EARELINSWVESQTNGIIR NVLQPSSV
3882.6	3882.4	351-385	AASVSEEFRADHPFLFCIK HIATNAVLFFGRCVSP
4555	4556, 4636*	304-349, 305-350	(D)VFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGV(D)
4440	4441	305-349	VFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGV
5155	5197**	1-47	GSIGAASMEFCFDVFKELKV HHANENIFYCPIAIMSALAM VYLGAKD
5040	5082**	1-46	GSIGAASMEFCFDVFKELKV HHANENIFYCPIAIMSALAM VYLGAK
5384.3	5385	13-59, 14-60	(D)VFKELKVHHANENIFYCPI AIMSALAMVYLGAKDSTRTQINKVVRF(D)
5718.1	5719	304-360, 305-361	(D)VFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGVDAASVSEEFRA(D)
5603.1	5603	305-360	VFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGVDAASVSEEFRA
6210	6211	192-246, 193-247	(D)TQAMPFRVTEQESKPVQMMYQIGLFRVASMASEKMKILELPFASGTMSMLVLLP(D)
6095	6097	193-246	TQAMPFRVTEQESKPVQMMYQIGLFRVASMASEKMKILELPFASGTMSMLVLLP
6701	6743**	1-60	GSIGAASMEFCFDVFKELKV HHANENIFYCPIAIMSALAM VYLGAKDSTRTQINKVVRFD
6586	6627**	1-59	GSIGAASMEFCFDVFKELKV HHANENIFYCPIAIMSALAM VYLGAKDSTRTQINKVVRF
7416	7458**	1-67	GSIGAASMEFCFDVFKELKV HHANENIFYCPIAIMSALAM VYLGAKDSTRTQINKVVRFD KLPGFGD
7301	7344**	1-66	GSIGAASMEFCFDVFKELKV HHANENIFYCPIAIMSALAM VYLGAKDSTRTQINKVVRFD KLPGFG
8137	8137	96-166, 97-167	(D)VYSFSLASRLYAEERYPILPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRNVLQPSSV(D)
8022	8021	97-166	VYSFSLASRLYAEERYPILPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRNVLQPSSV

Table 2.1: List of experimentally observed of Asp-specific peptides from ovalbumin. A single asterisk is used to indicate phosphorylation and double asterisk indicates acetylation.

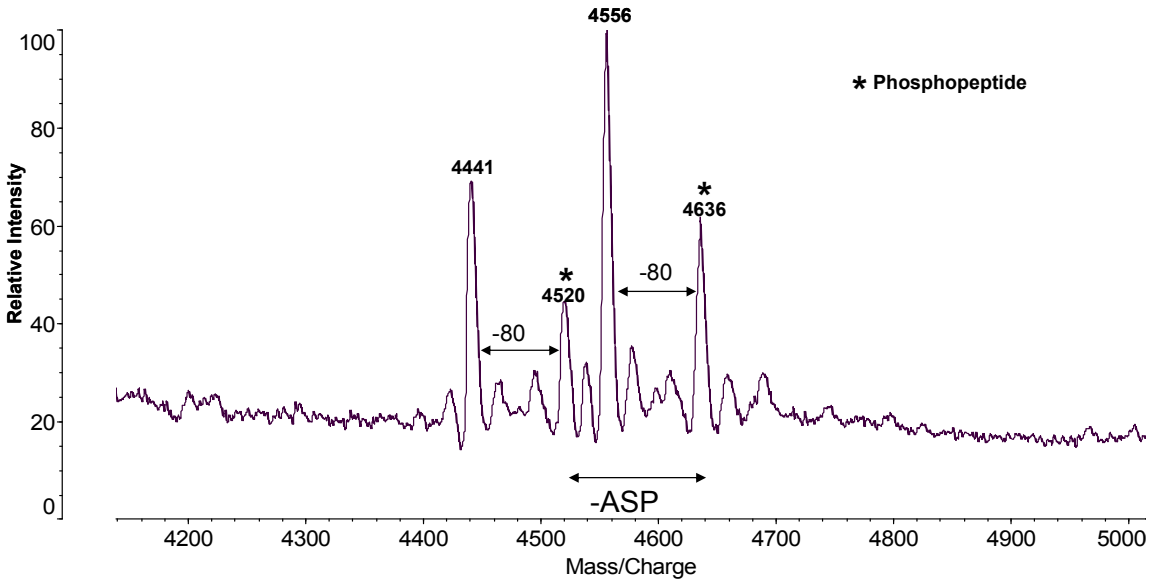


Figure 2.3: Mass Spectrum of phosphorylated peptides from ovalbumin.

The peptide pair at m/z = 3139.6 and m/z = 3124.3 did not correlate well with any predicted peptide masses upon initial investigation. However, this peptide pair was interpreted to be the peptide, 139-166 and 139-165, with its N-terminal Gln converted to pyro-glutamic acid. Conversion of the N-terminal glutamine to pyro-glutamic acid occurs via a cyclization reaction. This conversion results in a decrease in the molecular weight of a peptide by 17 Da. through the loss of NH_3. [28] Several reports have shown that a variety of factors such as high temperature, low pH, and salt concentration can drive this reaction. [29-31] Cyclization is thought to have occurred during microwave heating, although other studies have shown that artifactual modification is possible during the ionization process in mass spectrometric analyses. [29, 32]

Although the identity of this model protein was known prior to analysis, a database search was performed to identify the protein in order to demonstrate that MASCOT can identify proteins cleaved by this method. Ovalbumin was confidently identified with a score of 117 and an E-value of 3.7 e^{-7} based on 26 peptide matches, when the peptide masses from **Figure 2.1** were submitted. MASCOT results were considered significant if an e-value < 0.05 was recorded. MASCOT e-values represent the number of matches with equal or better scores that can occur by chance alone.

Chapter 3: Rapid Analysis of *Bacillus* Spores

This chapter has been reproduced in part with permission from: Swatkoski, S.; Russell, S. C.; Edwards, N.; Fenselau, C. *Rapid chemical digestion of small acid-soluble spore proteins for analysis of Bacillus spores.* Analytical Chemistry **2006**, *78*, 181-188.

Introduction

In this chapter is the application of microwave-accelerated acid hydrolysis to Bacillus spores is discussed. Bacillus spores have been well characterized by mass spectrometric based methods.[33-38] Traditionally, Bacillus spores were identified by detecting a selectively released family of intact, basic proteins referred to as small acid soluble proteins (SASPs), which constitutes 8-15% of the mass of Bacillus species. [37, 38,39] The SASPs were released with the addition of acid directly on a MALDI probe and their molecular weights were determined using MALDI-TOF MS.[37, 38] The masses of the SASPs were then matched to the molecular weights predicted by their genome. More recently, Bacillus spores were characterized using a “bottom-up” proteomics approach which combined on-slide release and enzymatic digestion of the SASPs with MALDI-TOF post source decay sequencing (LID) and database searching.[33-36] In this study, we have evaluated the use of microwave-accelerated acid hydrolysis for the rapid identification of Bacillus spores, which relies on the selective release of the SASPs by treatment with acid and subsequent digestion in a single step.

Materials and Methods

Materials.

Formic acid (88%), and sinapinic acid were obtained from Sigma (St. Louis, MO).

Source of *Bacillus* Spores.

Bacillus anthracis str. Sterne, and *Bacillus subtilis* str.168 vegetative cells were grown in nutrient media at 37^0C and subsequently sporulated on new sporulation media at 30^0C. Spores were harvested in-house using methods described previously.[40] Briefly, vegetative cells were grown and sporulated in new sporulation media. Sporulation occurred following a 48 hr incubation period at 37° C followed by 72 hr incubation at room temperature. Spores were collected, washed with Milli-Q water and lyophilized. Lyophilized spores were stored at -20 °C until use.

Microwave-assisted acid digestions of *Bacillus* Spores

Bacillus spore samples were digested using a domestic microwave (Goldstar 745 W). Spore samples were suspended in deionized water at a concentration of 2 mg/mL. Two hundred µL of each spore suspension was mixed with an equal volume of 12 % formic acid and placed in a loosely capped 5 mL glass vial. The glass vial was placed in the center of a domestic microwave (Goldstar 745 W) and irradiated at the highest power setting for 90 seconds. As a safety precaution a beaker containing 20 mL of water was also placed in the microwave to absorb excess energy and prevent the vials from exploding.

Analytical Methods. ***MALDI TOF Mass spectrometry of Bacillus Spores***

Digestion products from *Bacillus* spore proteins were analyzed directly by MALDI-TOF-MS. No sample clean-up, or fractionation of digestion products was performed prior to collection of MALDI TOF mass spectra. The crude mixture samples were prepared using a matrix sandwich method. The analyte was deposited between bottom and top matrix layers. Each layer was allowed to air dry prior to deposition of subsequent layers. The bottom matrix layer was prepared as a saturated solution of alpha-cyano hydroxycinamic acid (CHCA) in acetone, while the top matrix layer was CHCA at 10 mg/mL in 70 % ACN and 0.1 % TFA. For analysis of the HPLC fractions, 0.5 µL of analyte was spotted and allowed to air dry, followed by application of 0.5 µL of CHCA in 70 % ACN and 0.1 % TFA.

All samples were analyzed with a Kratos Axima-CFR Plus MALDI-TOF equipped with a curved field reflectron, which provided laser induced dissociation (LID) capabilities. All spectra were acquired in reflectron mode with a 337 nm N_2 laser as averages of 100 profiles for TOF scans, and 250 profiles for LID experiments.

Separation of Bacillus Spore Digestion Products.

To confirm the identities of the peptide products observed in the crude mixture, an offline LC-MALDI-TOF method was employed, and coupled with LID for peptide fragmentation. After digestion, the crude mixtures were centrifuged at 10,000 g for 10 minutes to remove any insoluble cell debris. The supernatant was subsequently dried down and reconstituted in 500 µL of mobile phase A (18MΩ water 0.1% TFA). The mixtures were separated by reverse phase HPLC on a C_{18} column (Phenomenex, Torrance, CA) with a 30 minute linear gradient from 2% to 60 % ACN, with a 1 mL per minute flow rate. Fractions were collected at a rate of 1 per minute and subsequently analyzed with MALDI-TOF-MS.

Database Searching

LID spectra were submitted to MASCOT with mass tolerance's of +/- 1.0 Da and +/- 1.5 Da for parent and fragment ions, respectively, with one missed cleavage allowed. The spectra were first searched to establish peptide identity by restricting the search to the Gram-positive bacteria of NCBInr's protein sequence database. The spectra were then searched against the entire NCBInr database to establish the specificity of the identified peptides – not only with respect to exact sequence matches but also with respect to near isobaric peptides that may represent false-positive identifications to a LID spectrum.

Theoretical Acid digestion of SASP Protein Sequences

The Pfam [*] and TIGRFAM [*] databases were accessed to compile a database of proteins belonging to the SASP family based on sequence homology.[41, 42] The Pfam database defines 7 SASP protein families: Small, acid-soluble spore proteins, alpha/beta type (PF00269); Small, acid-soluble spore protein, gamma-type (PF04259); and the Small, acid-soluble spore protein H, O, K, N, and P families (PF08141, PF08176, PF08177, PF08175, and PF08179, respectively). The TIGRFAM database defines 1 SASP protein family: small, acid-soluble spore protein, gamma-type (TIGR01442). Hidden Markov Models (HMMs) built from multiple sequence alignments of known family members provided a statistical certificate for family membership. All protein sequences from SwissProt, TrEMBL, GenBank, and TIGR that are annotated as belonging to one of these protein families were considered to be a SASP protein. A total of 776 sequences (268 distinct sequences) and 27 species were represented.

An *in silico* acid digestion allowing for cleavage on both sides of each aspartyl residue, was performed on all SASP proteins as defined above. The possiblility of an N-terminal methionine loss was also accounted for in the theoretical digestion. The m/z values of the peptide products from the *in silico* digestion were matched against prominent peaks observed in the digest spectrum for preliminary peak identification.

Results and Discussion

When performing the *in silico* digestion, all cleavage possibilities described in **Chapter 1** were accounted for. In addition, the possibility of N-

terminal methionine loss, which is common in prokaryotic post-translational modification, was accounted for. Using the SASP-B protein found in *B. anthracis* str. Sterne as an example; the N-terminal peptide may be cleaved on either side of the first aspartic acid residue, producing two potential masses (**Table 3.1**). In addition, each peptide can either retain or lose its N-terminal methionine, bringing the total number of potential peptides to four at m/z 1584.7, 1699.87, 1715.91, and 1830.91. An internal peptide digest product can be associated with two, one, or no aspartic acid residues at m/z 2041.93, 1926.93, and 1811.93, respectively. Finally, two masses are possible for the C-terminal peptide depending if the aspartic acid residue is retained (m/z 3201.7), or not (m/z 3086.7). This theoretical digest algorithm was applied to the SASP family of proteins.

Table 3.1 lists the selected set of peptides that were observed from both *B. anthracis* str. Sterne and *B. subtilis* str. 168, and indicates their matches to peptides from the theoretical digestion. This comprehensive theoretical digestion not only aided in peptide assignments, but also allowed for a greater understanding of the trends in formic acid cleavage applied to microorganisms.

Example: *B. anthracis* SASP- B (Avg_{MW}: 6679) Accession numbers: ti/BA4898 & ti/GBAA4898 & tr/Q81KU1_BACAN

MARSTNKLAVPGAESALDQMKYEIAQEFGVQLGADATARANGSVGGEITKRLVSLAEQQLGGFQK

Potential Met loss

Potential acid digest sites

Organism	Precursor SASP Protein Accession No.	Formic Acid Digest Product Peptide Sequences	-Met [M+H]+	-Met + Aspartic Acid	-Met - Aspartic Acid	+Met [M+H]+	+Met + Aspartic Acid	+Met - Aspartic Acid
B. anthracis str. Sterne	ti/BA4898 & ti/GBAA4898 & tr/Q81KU1_BACAN	**ARSTNKLAVPGAESALD**	**1699.9**	N/A	1584.9	1830.9	N/A	1715.9
		QMKYEIAQEFGVQLGAD	1926.9	2041.9	1811.9			
		ATARANGSVGGEITKRLVSLAEQQLGGFQK	**3086.7**	3201.7	N/A			
	ti/BA0858 & ti/GBAA0858 & tr/Q81UL0_BACAN	ANQNSSNQLVVPGATAAID	1869.9	N/A	1754.9	2001.0	N/A	1886.0
		QMKYEIAQEFGVQLGAD	1926.9	2041.9	1811.9			
		STARANGSVGGEITKRLVAMAEQSLGGFHK	**3072.6**	3187.6	N/A			
B. subtilis str. 168	sp/SAS1_BACSU & ti/NT01BS3753 & ti/NTL01BS2951	ANNNSGNSNNLLVPGAAQAID	2054.0	N/A	1939.0	2185.0	N/A	2070.0
		QMKLEIASEFGVNLGAD	1821.9	1936.9	1706.9			
		TTSRANGSVGGEITKRLVSFAQQNMGGGQF	**3098.5**	3213.6	N/A			
	tr/Q9AH72_BACSU & tr/Q9AH73_BACSU	ANSNNFSKTNAQQVRKQNQQSAAGQGQFGTEFASETN AQQVRKQNQQSAAGQQGQFGTEFASETD	7008.3	N/A	6893.2	7139.3	N/A	7024.3
		AQQVRQQNQSAEQNKQQNS	**2214.1**	2329.1	N/A			
	sp/SASG_BACSU & ti/NT01BS1114 & ti/NTL01BS0865	ANSNNFSKTNAQQVRKQNQQSAAGQGQFGTEFASETN AQQVRKQNQQSAGQQGQFGTEFASETD	6937.2	7052.2	6822.2	7068.3	N/A	6953.2
		AQQVRQQNQSAEQNKQQNS	**2214.1**	2329.1	N/A			

Table 3.1: Theoretical formic acid digest products from the SASP proteins from *B. subtilis* str. 168 and *B. anthracis* str. Sterne. Sequences and masses highlighted in red were observed experimentally. All masses listed are monoisotopic.

Identification of Digestion Products

A representative spectrum obtained from *B. anthracis* str. Sterne after incubation in the microwave with 12 % formic acid for 90 sec is shown in **Figure 3.1**. Five peaks, at m/z 1700.0, 2356.3, 3073.0, 3087.1 and 3097.6, were found to closely match m/z values predicted for acid digestion products. However, m/z 2356.3 was also observed in spectra of undigested *B. anthracis* spores, leaving four peaks tentatively assigned as digestion products. In order to confirm the sequences and identity of these peaks, peptide products were purified by HPLC. Three of the four digestion products were readily found in the HPLC fractions and were subjected to LID analysis. LID mass spectra for m/z 3087.1, and 3073.0 are presented in **Figure 3.2a & b.** These and other LID spectra were submitted to MASCOT to establish the identity of the peptides, by searching against the Gram-positive bacteria of NCBI's nr protein sequence database. The restricted database search was used in order to apply the prior knowledge of the source of the sample. The MASCOT search results are shown in **Table 3.2**.

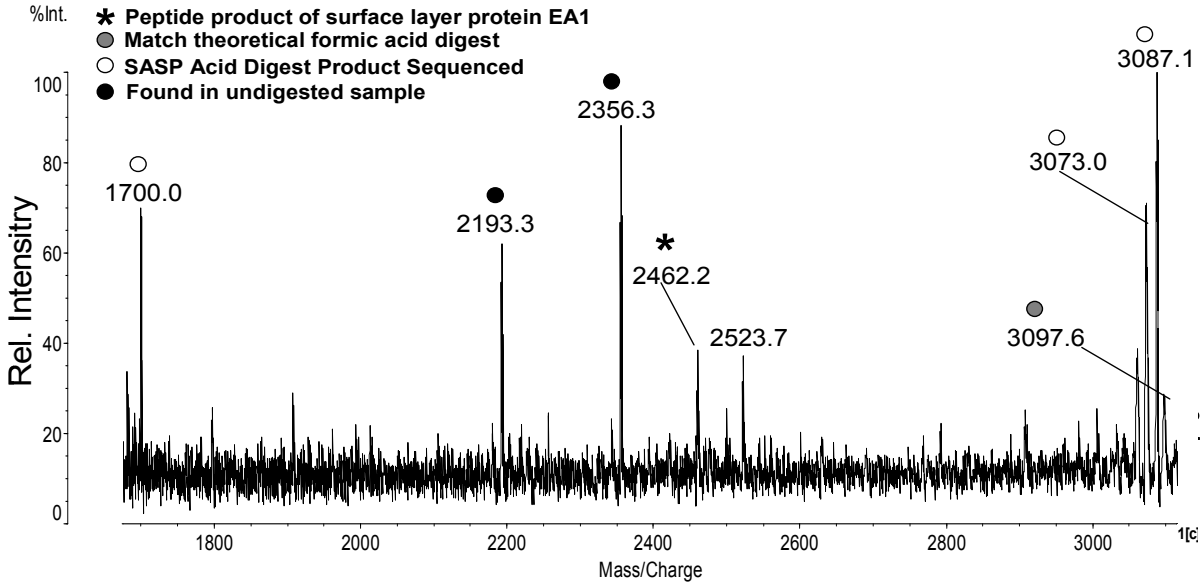


Figure 3.1: Mass spectrum of products of a microwave assisted formic acid digestion of (a) *B. anthracis* str. Sterne spores. The spore suspensions were mixed with an equal volume of 12 % formic acid and subjected to microwave irradiation for 90 seconds

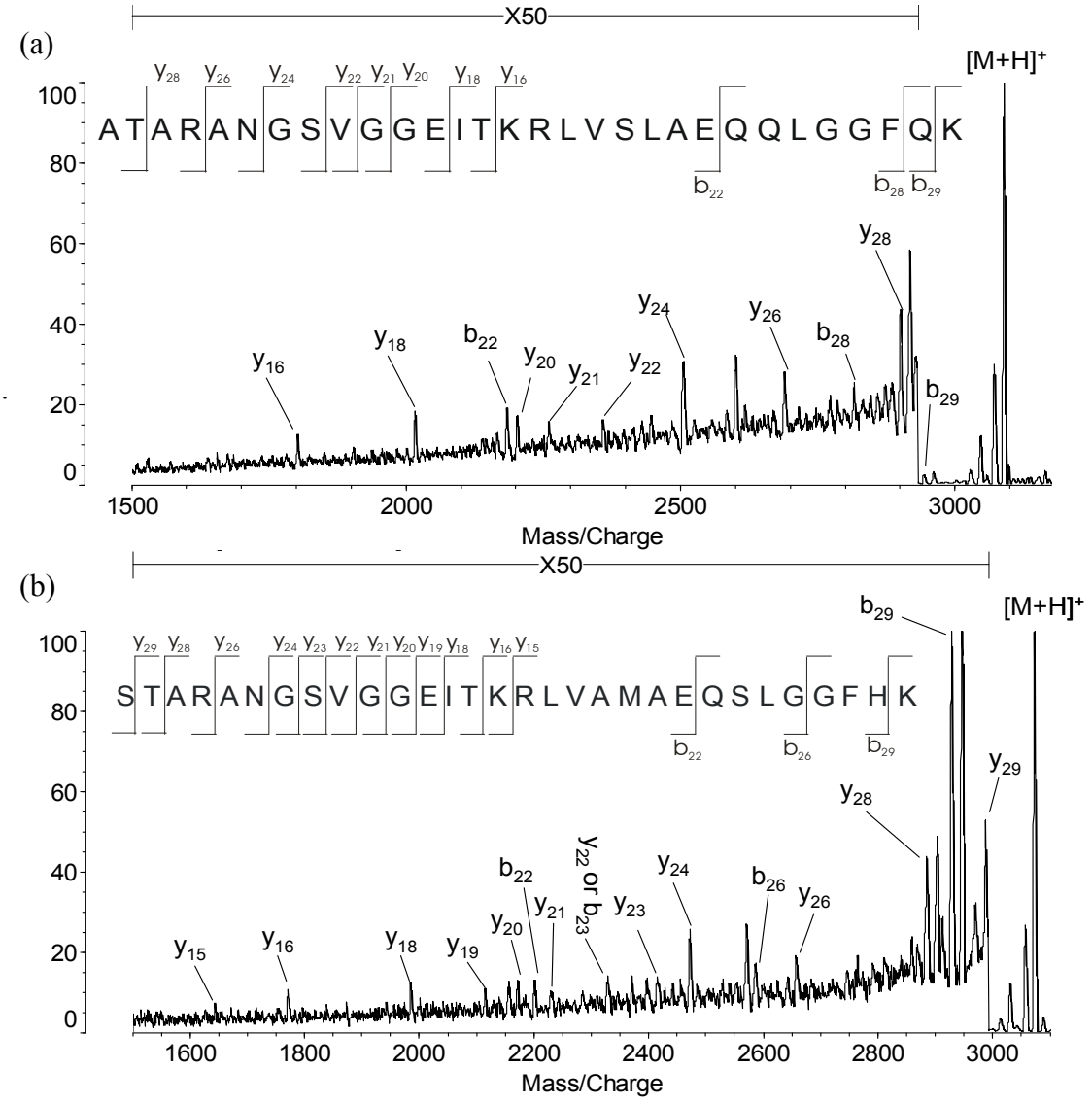


Figure 3.2: Fragmentation spectra (LID) of peptides from the microwave assisted formic acid digestion of *B. anthracis* str. Sterne spores (a) fragments from m/z 3087.1 (b) m/z 3073.0

Organism Digested	$[M+H]^+_{Obs.}$	$[M+H]^+_{Calc}$ [a]	Sequence	Score [b]	Expect value [b]	Protein Match	Accession number	Organism Match
B. anthracis Sterne	1700.0	1699.9	ARSTNKLAVPGAESALD	37	1.60E-01	SASP-B	gi|49187538	*B. anthracis* Sterne
	2462.1	2461.3	KDVITSEIGSQAVHVNVLNNPNL	66	2.00E-04	s-lp ea1	gi|49183866	*B. anthracis* Sterne
	3073.0	3072.6	STARANGSVGGEITKRLVAMAEQSLGGFHK	59	6.80E-04	SASP	gi|49183839	*B. anthracis* Sterne
	3087.1	3086.7	ATARANGSVGGEITKRLVSLAEQQLGGFQK	78	8.50E-06	SASP-B	gi|49187538	*B. anthracis* Sterne
B. subtilis 168	2214.0	2214.1	AQQVRQQNQSAEQNKQQNS	71	6.00E-05	SASP-γ	gi|16077932	*B. subtilis* 168
	3099.0	3098.5	TTSRANGSVGGEITKRLVSFAQQNMGGGQF	39	6.90E-02	SASP-α	gi|16080009	*B. subtilis* 168

[a] Masses were calculated as monoisotopic masses

[b] MASCOT Searches were performed against the NCBInr database against gram positive bacteria
Peptide and MS/MS fragment mass tolerences were set to +/- 1.0 and +/- 1.5, respectively.

Table 3.2: MASCOT Search results for formic acid digest products from *B. anthracis* str. Sterne and *B. subtilis* str. 168 against gram positive bacteria.

The MASCOT searches provided assignments for many of the peptides observed in the whole cell digest. For *B. anthracis* str. Sterne, peptides with m/z 1700.0 and 3087.1 were found in HPLC fractions 15 and 22, respectively. While the MASCOT score for 1700.1 was low (**Table 3.2**), the confident observation of another digest product from the same protein at m/z 3087.1 argued for the identity of this peptide. These two peptides represent 72% of SASP-B (gi|49187538). The peak at m/z 3073.0 was observed in HPLC fraction 23 and its LID spectrum matched the C-terminal digest product from SASP gi|49183839 with a score of 59, representing 45% sequence coverage. The three observed peptides originated from the most abundant intact proteins isolated by treatment of *B. anthracis* str. Sterne spores with 1M HCl.[40] Similar relative intensities were observed for SASP protein ions isolated from *B. anthracis* str. Sterne by treatment with 12 % formic acid without microwaving (**Figure 3.3).** These observations support a strong correlation between proteins extracted and peptide products detected.

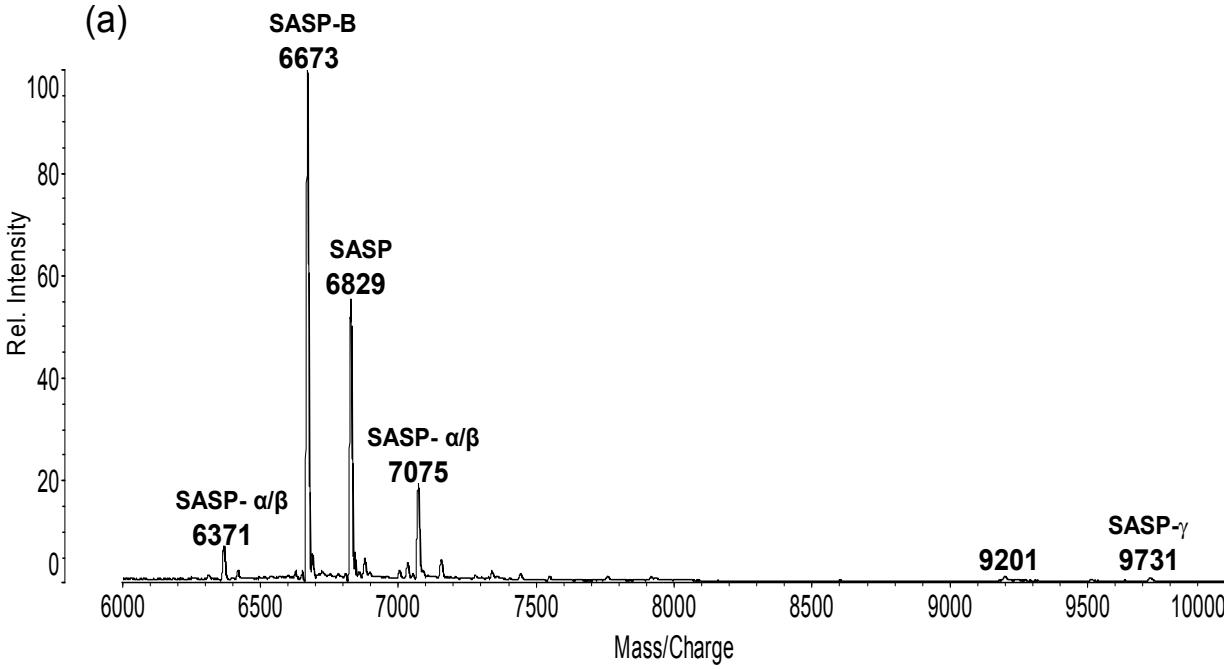


Figure 3.3: Mass spectrum of undigested SASPs from *B. anthracis* str. Sterne selectively solubilized with 12 % formic acid.

For *B. subtilis* str. 168, two peaks were observed, at m/z 2213.8 and 3099.5 were matched to SASP acid digestion products (**Figure 3.4**). A LID spectrum for m/z 2213.8 was obtained from HPLC fraction 8 for *B. subtilis* str. 168 (**Figure 3.5**). The fragment ions matched C-terminal peptides from two SASP-γ proteins (gi|71657050 & gi|16077932), found in *B. subtilis* str. 168, with a MASCOT score of 71 (**Table 3.2**). It is speculated that the primary source of ions with m/z 2213.8 is the latter protein, as it was observed at relatively high intensity in the spectrum of the undigested sample (data not shown). The second digest product, with ions at m/z 3099.5, was found in fraction 21, and its LID spectrum matched a peptide from *B. subtilis* str. 168 SASP-γ (gi|16080009) with a score of 39 (**Table 3.2**). It is worthwhile to note that for both species, five of the six SASP peptides originated from the C-terminus of their proteins. It is not yet known whether this trend is a consequence of the digestion chemistry, a bias in ionization efficiency, or a coincidence.

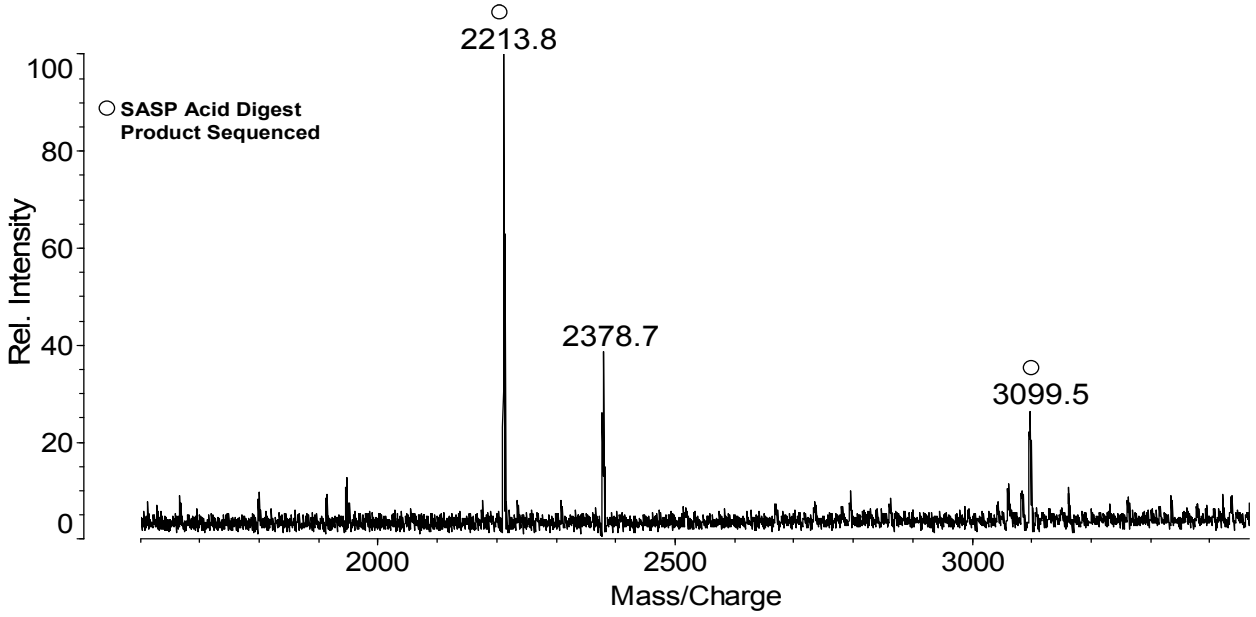


Figure 3.4: Mass spectrum of products of a microwave assisted formic acid digestion *B. subtilis* str. 168 spores. The spore suspensions were mixed with an equal volume of 12 % formic acid and subjected to microwave irradiation for 90 seconds.

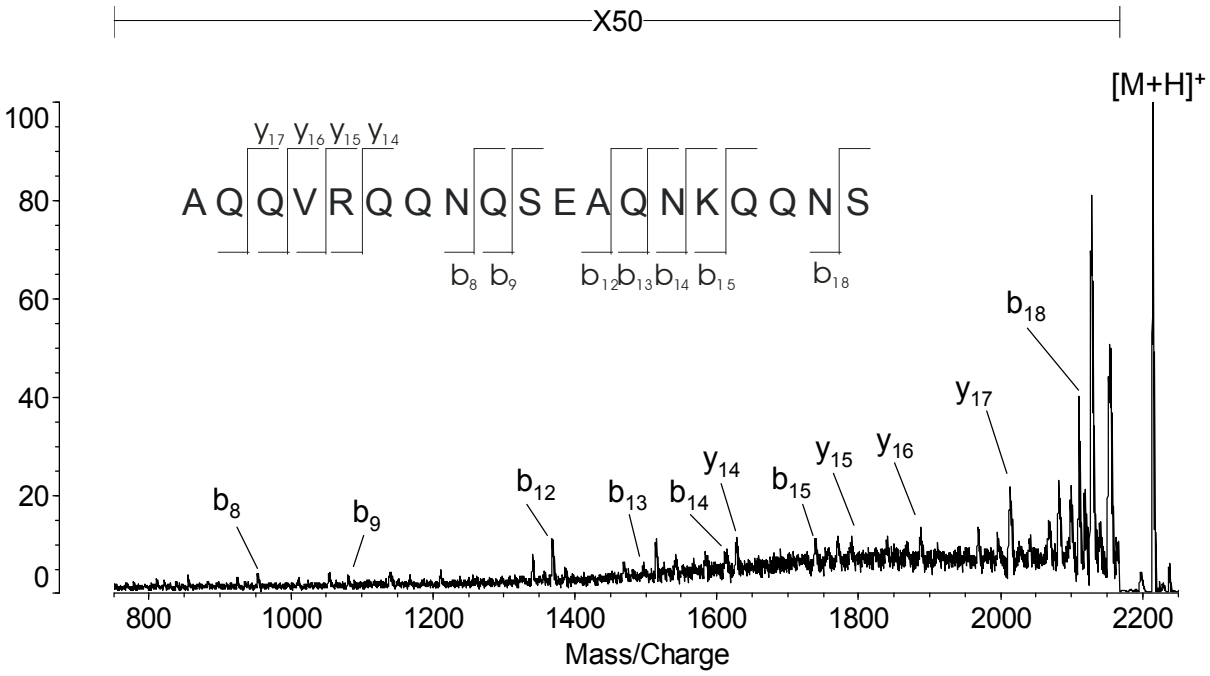


Figure 3.5: Fragmentation spectra (LID) of peptides from the microwave assisted formic acid digestion of *B. subtilis* str. 168 spores fragments from m/z 2213.8.

Specificity of SASP Peptides in Acid Digests

Once the peptides in the microwave digests were identified as belonging to the SASP family, it was of interest to evaluate their specificity for bacterial identification. The peptide sequences determined experimentally from *B. anthracis* str. Sterne and *B. subtilis* str. 168 were checked against all SASP proteins, to find those species whose SASP proteins would yield the same peptides when subjected to an acid digestion. Matches are summarized in **Table 3.3**. The four ions identified as SASP peptides from *B. anthracis* str. Sterne were found to be *cereus* group specific.[43, 44] The *B. anthracis* str. Sterne peptide sequence of the ion at m/z 1700.0 was also found in *B. cereus* and *B. thuringinesis*. The *B. anthracis* str. Sterne peptide sequence of the ion at m/z 3087.1 was also found in *B. cereus*. The *B. anthracis* str. Sterne peptide sequence of the ion at m/z 3073.1 was also found in *B. thuringiensis*, and *B. cereus*. The putative *B. anthracis* str. Sterne peptide sequence of the ion at m/z 3097.6, assigned on the basis of an *in silico* SASP digestion, was also found in *B. thuringinesis*. The *B. subtilis* str. 168 peptide sequence of the ion at m/z 2213.8 was also found in *Geobacillus steareothermophilus*, but no other Bacillus species. It should be noted that *Geobacillus steareothermophilus* was previously classified in the *Bacillus* genus.[45, 46] The *B. subtilis* str. 168 peptide sequence for the ion at m/z 3099.1 was found only within the species *B. subtilis*.

$[M+H]^+_{calc}$ [a]	Sequence	*Species* [b]
1699.9	ARSTNKLAVPGAESALD	Bt, Ba, Bc
3072.6	STARANGSVGGEITKRLVAMAEQSLGGFHK	Bt, Ba, Bc
3086.7	ATARANGSVGGEITKRLVSLAEQQLGGFQK	Ba, Bc
$[M+H]^+_{calc}$ [a]	Sequence	
2214.1	AQQVRQQNQSAEQNKQQNS	Gst, Bs
3098.5	TTSRANGSVGGEITKRLVSFAQQNMGGGQF	Bs

[a]Monoisotopic masses calculated for peptide ions (N-terminal peptides calculated with methionine losses[17, 41])

[b]Ba, *B. anthracis*; Bc, *B. cereus*; Bs, *B. subtilis;* Bt, *B. thuringiensis*; Gst, *G. steareothermophilus*;

Table 3.3: Search of digest peptide sequences against theoretical acid digest products from the SASP family of proteins.

The specificity of these peptide sequences, and the LID spectra from which they were derived, were also evaluated by searching the spectra against the entire NCBI nr protein sequence database. The objective of this search was not to identify the peptides, but to determine if any known non-SASP peptide might contain the same or a near isobaric sequence that would constitute a false positive species identification in the field. **Table 3.4** shows the scores of the best non-SASP peptides identified by MASCOT for each of the proposed SASP peptide sequence markers. Only one of the peptides at m/z 1700.0 had a non-SASP match of significant homology with a score of 31. However, this score match was to a hypothetical protein from the Eukaryotic parasite *Theileria annulata*.[47] While this may pose a risk as a false positive, the observation of the complementary peptide ion at m/z 3087.1, which also originates from SASP-B, would rule out such a case.

$[M+H]^+_{Calc}$ [a]	Sequence	Highest Score	Next nearest Score	Accession number	Protein Name [Source Organism]
1699.9	ARSTNKLAVPGAESALD	37	31	gi|65302884	hyp. protein [Theileria annulata]
2461.3	KDVITSEIGSQAVHVNVLNNPNL	66	32	gi|67474326	hyp. protein [Entamoeba histolytica HM-1:IMSS]
3072.6	STARANGSVGGEITKRLVAMAEQSLGGFHK	59	31	gi|23015100	COG3210: Large exoproteins involved in heme utilization or adhesion [Magnetospirillum magnetotacticum MS-1]
3086.7	ATARANGSVGGEITKRLVSLAEQQLGGFQK	78	15	gi|68231582	Amidohydrolase [Frankia sp. EAN1pec]
2214.1	AQQVRQQNQSAEQNKQQNS	71	36	gi|71984913	F29D10.3 [Caenorhabditis elegans]
3098.5	TTSRANGSVGGEITKRLVSFAQQNMGGGQF	39	21	gi|244239	YCR592 [Saccharomyces cerevisiae]

Table 3.4: MASCOT Search results for formic acid digest products from *B. anthracis* str. Sterne and *B. subtilis* str. 168 against all entries, for the next nearest match.

Pseudo-Top Down Approach

Even after much empirical optimization (i.e. reaction time, acid concentration, etc.), it was observed that the acid digestion was less efficient than that of trypsin, where the cleavage reaction goes quickly to completion.[37, 48, 49] However, this proved to be advantageous for species identification, since both the intact proteins and the peptide digestion products can be observed simultaneously. **Figure 3.6** presents a spectrum of *B. anthracis* str. Sterne digested with 12 % formic acid for 90 seconds with the mass range expanded to 7,000 Da. In this spectrum, one can clearly observe not only the peptide digestion products, but also the SASP (gi|49183839) and SASP-B (gi|49187538) proteins from which they originated. An additional protein was observed at m/z 5257 (**Figure 3.6**), which matched the mass of a spore coat protein that contains no aspartic acid residues. Partial fragmentation (data not shown), along with the mass match allowed a tentative assignment of this peak to a spore coat protein (gi|49186413) from the small acid-soluble spore protein O family (Pfam: PF08175). The observation of both digest products and their precursor proteins suggests that this combination of markers may offer greater specificity than either set of markers alone.

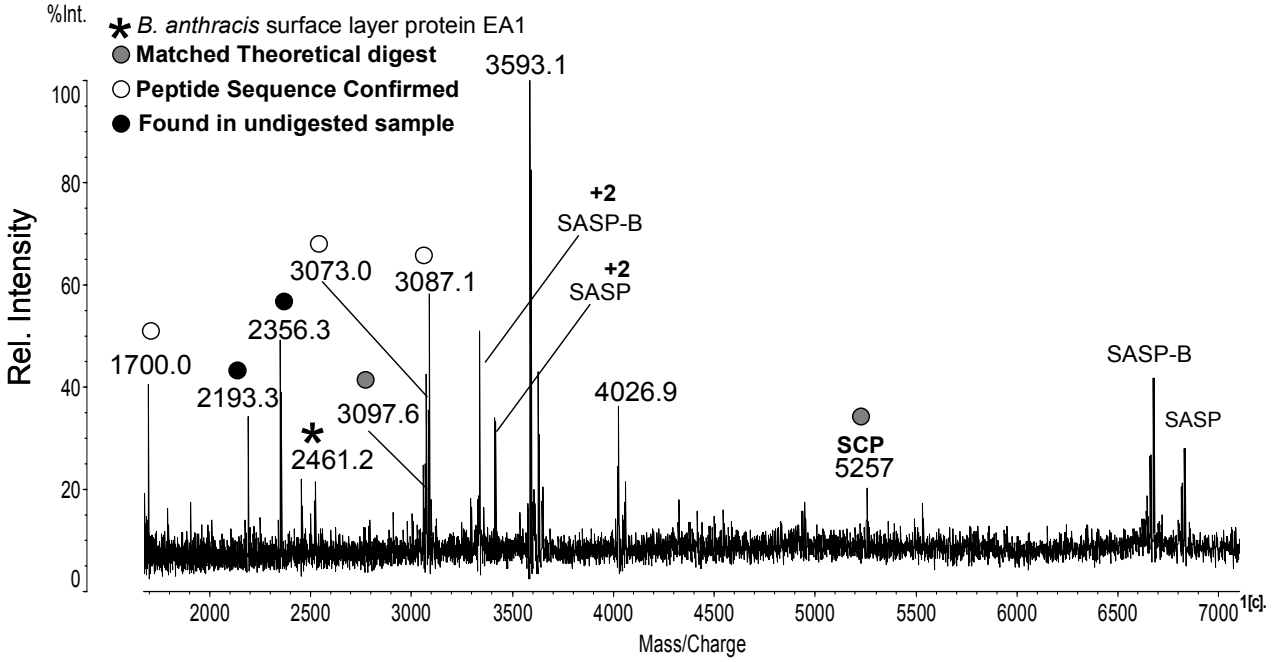


Figure 3.6: Mass spectrum of a 90 second microwave assisted formic acid digestion from *B. anthracis* str. Sterne.

For example, the peptide sequences of the ions at m/z 1700.0 and m/z 3087.1, from *B. anthracis* str. Sterne, are not specific within the cereus group. However, the mass of the identified *B. anthracis* SASP protein (assuming the likely N-terminal methionine loss) is specific for *B. anthracis*, with respect to other SASP proteins, when observed with a mass accuracy of 500 ppm or better. Similarly, the peptide sequence of the ion at 2213.8 is found in *Geobacillus stearothermophilus* in addition to *B. subtilis*, but the mass of the identified *B. subtilis* SASP protein is specific for *B. subtilis,* with respect to other SASP proteins, when observed with a mass accuracy of 1000 ppm or better. These results suggest that the parallel analysis of digest products and intact protein masses may result in significantly improved specificity for biomarker and microorganism identification.

Observation of a surface layer protein regulated by pXO1

In addition to the SASP peptides, the microwave digestion method produced an additional product from *B. anthracis* str. Sterne that appears to have originated from the extracellular antigen 1 (EA1) surface layer protein. The peptide product from this protein, with m/z 2461.7, was isolated in HPLC fraction 18. The LID sequence data (**Figure 3.7**) was matched with the EA1 S-layer protein (gi|47526173) with a MASCOT score of 66 and an expect value of 5.0 x 10^{-4}, when searched against all entries in NCBI's nr protein sequence database. Again the C-terminal peptide was observed, with one missed cleavage.

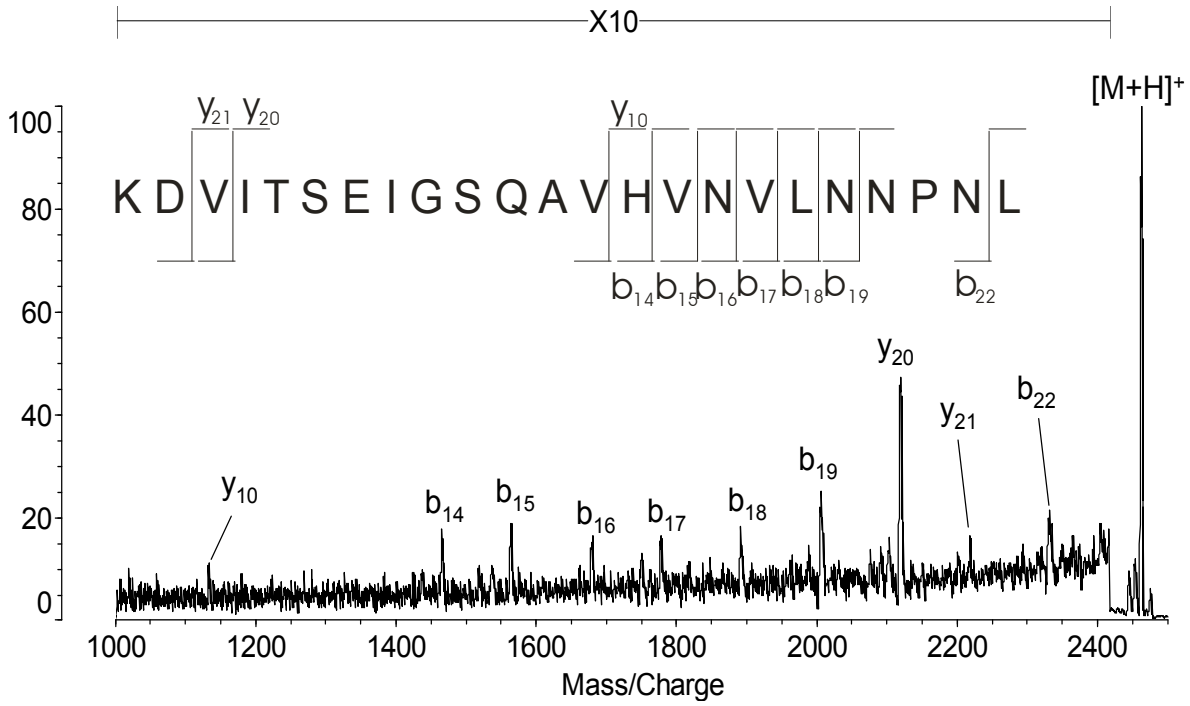


Figure 3.7: Fragmentation spectra (LID) of peptide m/z = 2164.7 from the microwave assisted formic acid digestion of *B. anthracis* str. Sterne spores

EA1 has been reported as the major cell-associated antigen for *B. anthracis*.[50] While this protein is highly expressed in the vegetative cell state, it has been reported to be difficult to remove during spore harvest and is commonly observed in spore preparations.[51] Although EA1 is encoded by the chromosomal gene *eag*, its expression is known to be controlled by the *PagR* regulator gene found on the pXO1 plasmid.[52] Observation of EA1 along with the spore coat protein in *B. anthracis* str. Sterne provides preliminary evidence that a new family of surface proteins were targeted by the acid digestion.

Analysis of spore mixtures

After the level of specificity of the peptide products had been established, an analysis of a mixture of *B. anthracis* str. Sterne and *B. subtilis* str. 168 was carried out. This was done to demonstrate the ability of the acid digestion method to identify *B. anthracis* str. Sterne in a complex biological background of *B. subtilis* str.168. **Figure 3.8a** shows a spectrum of peptides from the 1:1 mixture. SASP peptide products from *B. anthracis* str. Sterne were readily identified as cereus group specific peptides at m/z 1700.2, 3073.1, and 3087.1. In addition, the *B. anthracis* specific EA1 peptide was observed at m/z 2462.1. Digestion products from the SASPs of *B. subtilis* str. 168 were observed at m/z 2213.8 and 3099.1. The utility of the pseudo-top down approach was further demonstrated by expanding the mass window of the mixture mass spectrum (**Figure 3.8b**).

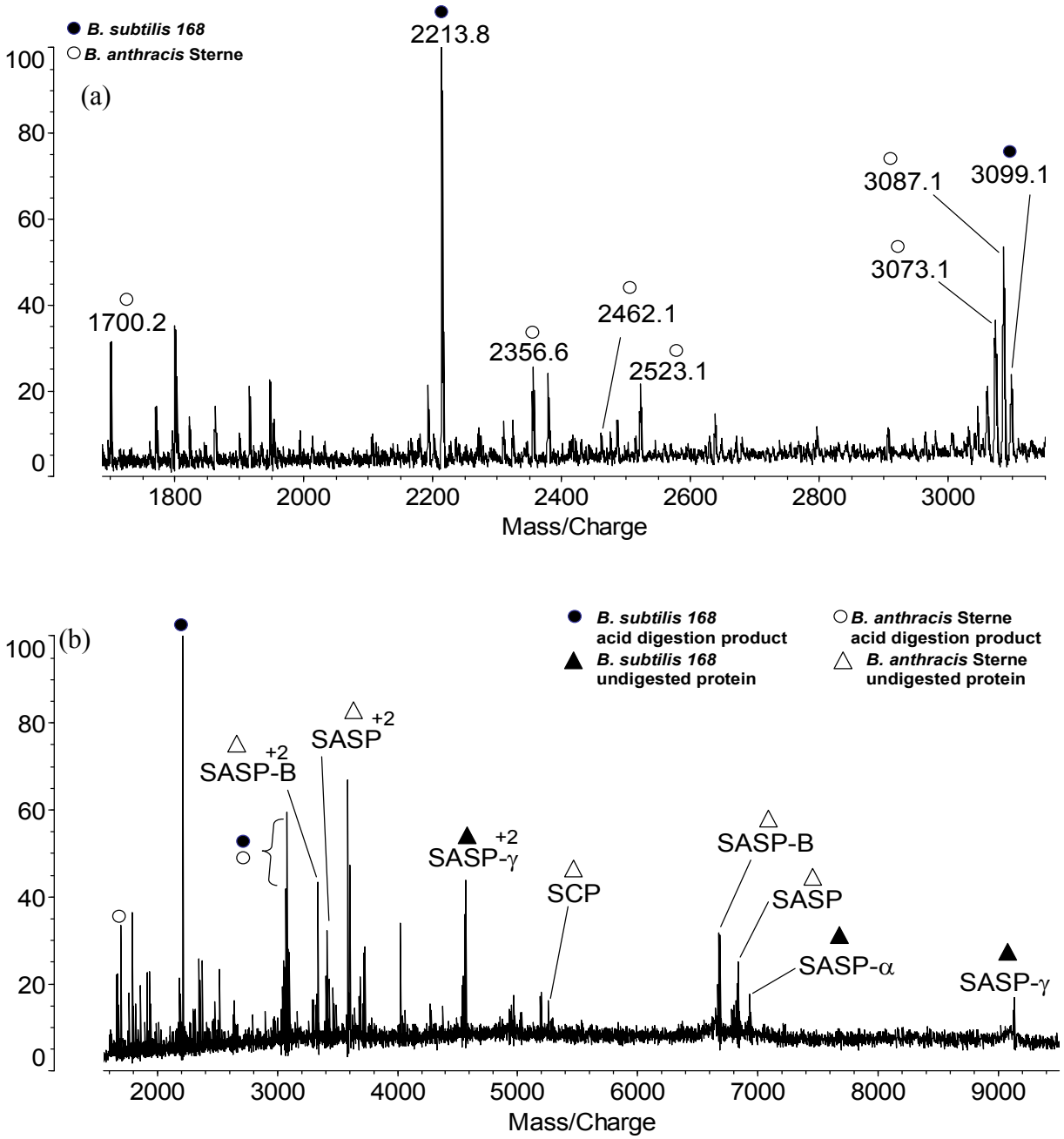


Figure 3.8: Microwave assisted formic acid digestion carried out for 90 seconds on 1:1 mixture. (a) A representative spectrum showing the acid digest peptide products of *B. anthracis* str. Sterne and *B. subtilis* str. 168 and (b) expanded mass window showing observed intact proteins and peptide products.

Protein peaks were observed for SASP-B (gi|49187538) and SASP (gi|49183839) from *B. anthracis* str. Sterne at m/z 6680.3 (unique mass at 500 ppm with respect to SASP proteins) and 6836.6, respectively. Doubly charged peaks for SASP-B and SASP proteins were also observed at m/z, 3340.3 and, 3418.4 respectively. Protein peaks from SASP-α (gi|16080009) and SASP-γ (gi|16077932) of *B. subtilis* str. 168 were also observed at m/z 6941.5, and 9138.3 (unique mass at 1000 ppm with respect to SASP proteins), respectively, accompanied by doubly charged ions at m/z 4569.5, and 3471.1, respectively.

Chapter 4: Analysis of Bacteriophage MS2

This chapter has been reproduced in part with permission from: Swatkoski, S.; Russell, S. C.; Edwards, N.; Fenselau, C. *Rapid analysis of a model virus using residue specific chemical cleavage and MALDI TOF mass spectrometry.* Analytical Chemistry **2007**, 79, 654-658.

Introduction

In this chapter, the application of microwave-accelerated acid hydrolysis to the identification of the virus, bacteriophage MS2, is discussed. Bacteriophage MS2 is a well studied virus in the field of biological warfare defense research because it is used as a simulant for actual viral threat agents. Bacteriophage MS2 is a single stranded RNA virus which has a very simple genome. Four genes encode 4 proteins which include a coat protein, a replicase, a lysis protein, and an assembly protein.[53-55] Of these four proteins, only the coat protein is present in a high enough copy number (180 copies/viron) for easy detection by mass spectrometry.

Previous studies have show that addition of acid to virus samples results in the release and solubilization of major structural proteins for rapid analysis by mass spectrometric based methods.[56] However, in identification of viruses such as bacteriophage MS2, relying on the direct analysis of a single protein biomarker introduces the possibility of misidentifications if the measured molecular weight does not accurately match the predicted mass. Furthermore, high sequence homology between closely related viral species can further complicate identification. Due to instrumentation limitations in mass accuracy, resolution, and sensitivity at higher masses, a method for the identification of the

bacteriophage MS2 based on the detection of peptides originating from the coat protein is desirable. In this study, microwave-accelerated acid hydrolysis was used to denature the virus and cleave the capsid protein of bacteriophage MS2. Reactions were carried out in a scientific microwave to enable precise control of experimental conditions.

Materials and Methods

Source of Samples. Bacteriophage MS2

Bacteriophage MS2 was cultured in a F^+ *E. coli* host (ATCC # 25404) using previously reported protocols.[56, 57] The E. coli host was grown until an optical density of 0.125-0.188 was reached. E. coli was infected with the bacteriophage MS2 stock suspension (ATCC # 15597-B1). The bacteriophage MS2/E. Coli suspension was then centrifuged at 10,000 g following a 3-4 hour incubation period at 37°C and a lysis of the E. coli host was observed. The supernatant was then collected and passed through a 0.2 μm syringe filter to remove any *E. coli* cell debris. Upon harvesting bacteriophage MS2, the titer was typically measured at 8×10^{10} pfu/mL. All bacteriophage MS2 analyses were carried out on this culture. No further purification was performed to remove any *E. coli* proteins or low molecular weight constituents from the complex growth media.

Microwave-assisted acid digestions of Bacteriophage MS2

All microwave-accelerated acid digestions of bacteriophage MS2 were carried out using a Discover Benchmate (CEM Corp., Matthews, NC) microwave

system. A 200 µL aliquot of bacteriophage MS2 suspension was mixed with 200 µL of 50% acetic acid and 50 µL of 1% Triton X-100 in a 10 mL Pyrex glass sample holder (CEM Corp., Matthews, NC). A small stir bar was added to enable mixing during the digestion process. The sample holder was then sealed with a rubber septum provided by CEM Corp. and inserted in the microwave apparatus. Microwave irradiation was applied to five different samples at powers of 100 W, 130 W, 160 W, 190 W, and 220 W. When 190 W was used, irradiation was applied to the samples for less than 30 s. Less than 15 s of 190 W microwave irradiation was required to achieve a targeted temperature of 108 °C. Once the required reaction temperature was achieved, only an additional 15 s of irradiation was applied.

MALDI-TOF Mass Spectrometry of Bacteriophage MS2

All samples, following microwave heating, were applied to the MALDI sample plate using a "sandwich" method.[58] One µL of saturated sinapinic acid in 50%/50% (v/v) acetonitrile/deionized water containing 1% trifluoroacetic acid was applied to the sample plate and allowed to air dry. Next, 1.0 µL of sample mixture was applied to the sample plate and allowed to air dry. In a final step, 1.0 µL of saturated sinapinic acid solution was added and allowed to air dry.

All MALDI mass spectra were acquired using an Axima CFR Plus MALDI-TOF mass spectrometer (Shimadzu Biotech., Columbia, MD). Spectra were acquired with a 337 nm N_2 laser in linear mode, which accommodates the larger peptides provided by cleavage at Asp.[59] All mass spectra were collected as an

average of 150 profiles. For the analysis of the undigested bacteriophage MS2 sample, a 3-point internal calibration was used. The bacteriophage MS2 culture, treated with 50% acetic acid and 1% Triton X-100, was spiked with insulin ($[M+H]^+_{av}$ = 5,734.5), cytochrome C ($[M+H]^+_{av}$ = 12361.9) and apomyoglobin ($[M+H]^+_{av}$ = 16952.2). The mass spectra of the digested bacteriophage MS2 was calibrated using the intact molecular weight of the MS2 coat protein and its doubly charged signal as internal calibrants for a 2 point calibration.

Database Searching

An "in house" version of MASCOT was used to carry out all database searches following data collection. All masses labeled in the mass spectrum shown as Figure 1C (m/z range 6600-14250), excluding the intact coat protein and its doubly charged signal at m/z 13729 and 6866, respectively, were submitted to MASCOT as a peptide mass fingerprint, with a peptide mass tolerance of ±3.0 Da. In the MASCOT search menu, the edited form of the "formic acid" enzyme option was used which accounts for all possible cleavage pathways. Peptide mass fingerprints were then searched against all entries in the NCBInr protein database allowing for up to 9 missed cleavages.

Results and Discussion

Optimization of Microwave-assisted Chemical Digestion

For the analysis of bacteriophage MS2, microwave power and reaction times were carefully coupled to achieve optimal conditions for digestion efficiency and specificity as evaluated by mass spectrometry. By performing these

reactions in the laboratory microwave system, time, temperature and microwave power could be controlled. The effect of microwave irradiation power on the digestion efficiency of the bacteriophage MS2 coat protein was monitored by performing a 30 W step-wise power study. At each microwave power setting, the sample was irradiated for 15 s after a temperature of 108 °C was achieved.

Figure 4.1 (a-d) shows the mass spectra of bacteriophage MS2 treated with 50% AA and 1% Triton X-100 obtained after irradiation with 130, 160, 190, and 220 W of microwave power. As microwave power was increased, a gradual decrease was observed in the ion signal from the coat protein precursor, while the chemical digest product ion intensities increased. The signal intensity of chemical digestion products resulting from cleavage on both the N- and C-terminal sides of the aspartic acid residues also increased with increasing microwave power, which suggested an increase in digestion efficiency. Mass spectra of the digestion products obtained following treatment of the virus sample with 190 W provided the best overall signal-to-noise (s/n) ratios for the digestion products. Therefore, 190 W was designated as the optimal power setting for this analysis. When the microwave power was increased to 220 W, the intensities of the protein precursor ion signal and other digest products were significantly weakened. Consequently, about 22.5% of the sequence coverage of the coat protein was lost.

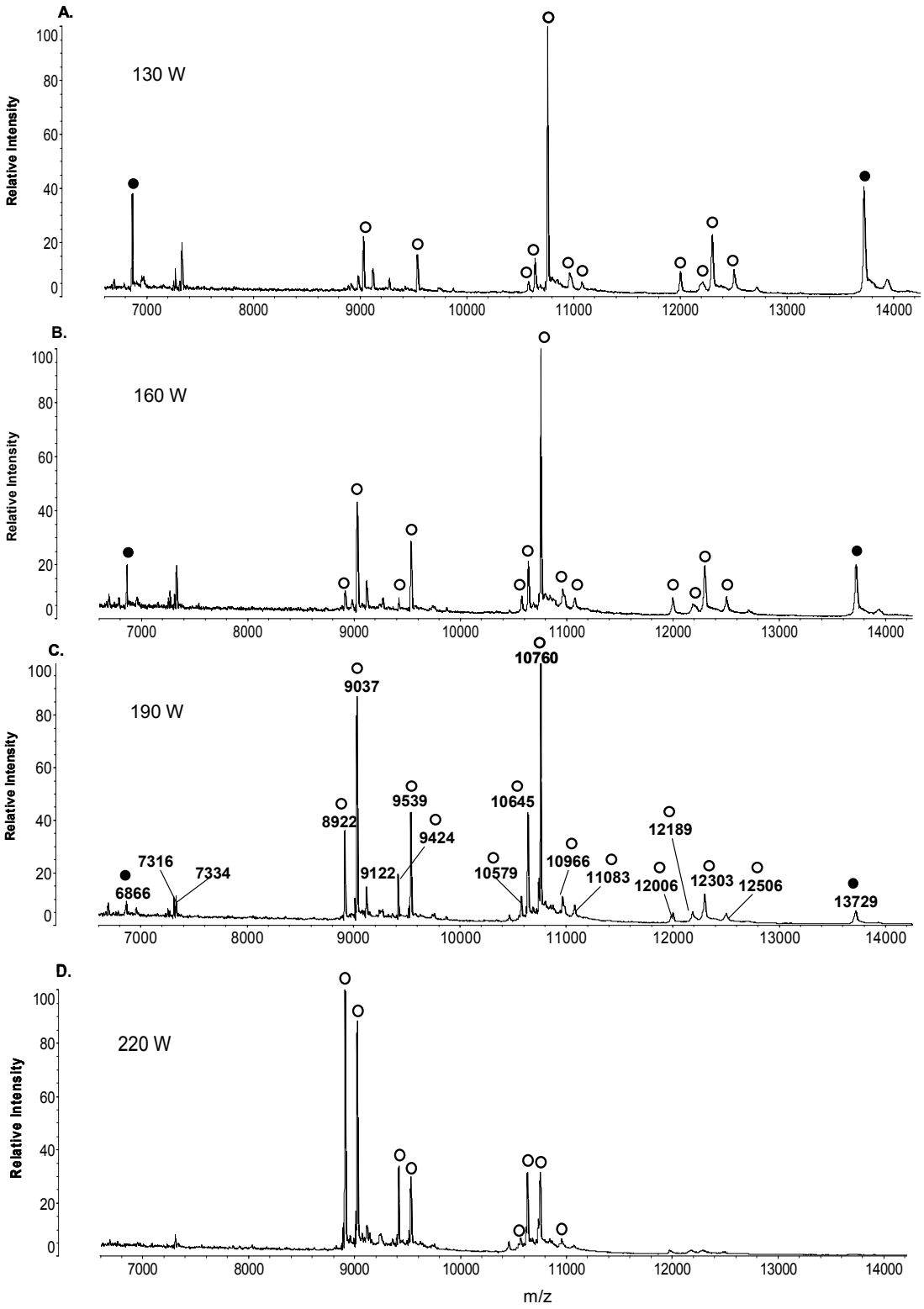


Figure 4.1: Mass spectrum of the bacteriophage MS2 suspension following 130 W (A) 160 W (B) 190 W (C) 220 W (D) of microwave irradiation for 15 sec. above 108° C. Closed circles indicate bacteriophage MS2 coat protein signals and open circles indicate acid digestion products from the bacteriophage MS2 coat protein.

Identification of Chemical Digestion Products

The bacteriophage MS2 coat protein is a 129 amino acid protein which has a molecular weight of 13728 Da. It contains four potential aspartic acid cleavage sites at amino acid positions, 11, 17, 100, and 114. **Figure 4.2** is a representative spectrum of an undigested bacteriophage MS2 sample treated with 50% acetic acid and 1% Triton X-100. The intact bacteriophage MS2 coat protein and its doubly charged signals were observed at m/z 13729 and 6864, respectively. As evaluated by mass spectrometry, bacteriophage MS2 samples treated with 50% acetic acid and Triton X-100 provided improved s/n and decreased broadening in intact coat protein signal compared to samples treated with 50% acetic acid alone.

The peptides observed in the representative mass spectra of bacteriophage MS2, treated with 50 % acetic acid and 1% Triton X-100, obtained after the samples were subjected to 190 W of microwave irradiation (**Figure 4.1C**) show that cleavage occurred at each of the four aspartic acid residues. In addition, four peptide pairs separated by 115 Da. were observed, which indicates that cleavage occurred on both the N- and C- terminal sides of the Asp residues at each cleavage site. **Table 4.1** is a list of all m/z values observed in the mass spectrum (**Figure 4.1C**) attributed to digestion products. The masses in parentheses are peptides that differ by one Asp residue, which corresponds to a difference of 115 Da. Parentheses also indicate that the termini of the internal peptides that retained the Asp residue cannot be determined without MS/MS analysis.

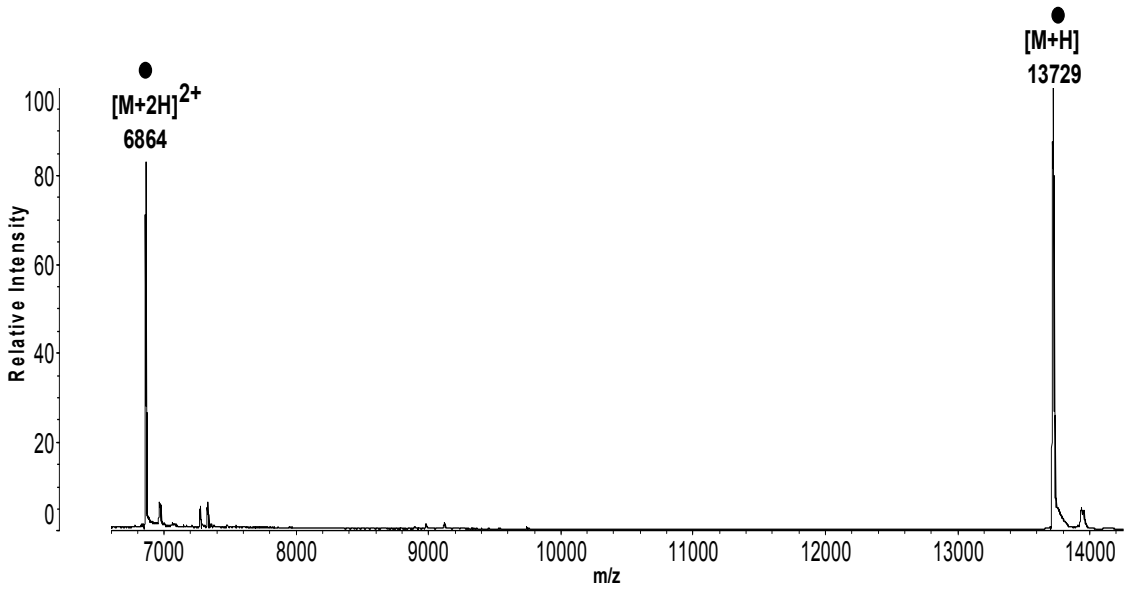


Figure 4.2: Mass spectrum of bacteriophage MS2 suspended in 50% acetic acid/1% triton X-100. Closed circles indicate bacteriophage MS2 coat protein signals.

Triton X-100 is reported to hinder spectral interpretation in the low mass range of peptide mass spectra. However, even without Triton X-100, chemical digestion products were not observable below m/z = 3000 because of the low molecular weight constituents of the *E.coli* growth media of the crude bacteriophage MS2 sample. The larger mass peptides that were observed provided 100% sequence coverage of the bacteriophage MS2 coat protein and overlapping peptides provided a reliable identification.

$[M+H]^+_{Obs.}$	Position	Amino Acid Sequence
13729	**1-129**	ASNFTQFVLVDNGGTGDVTV APSNFANGVAEWISSNSRSQ AYKVTCSVRQSSAQN RKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKD GNPIPS AIAANSGIY
12506	**12-129**	NGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVA TQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY
12303, (12189)	**1-114, (1-113)**	ASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQN RKYTI KVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK**(D)**
12006	**18-129**	VTVAPSNFANGVAEWISSN SRSQAYKVTCSVRQSSAQNR KYTIKVEVPKVATQTV GGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN PIPSAIAANSGIY
11083, (10966)	**11-113, (12-113)**	**(D)**NGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIK VEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQG LLK**(D)**
10760, (10645)	**1-100, (1-99)**	ASNFTQFVLVDNGGTGDVTV APSNFANGVAEWISSNSRSQ AYKVTCSVR QSSAQNRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNS**(D)**
10579	**17-113 or 18-114**	**(D)**VTVAPSNFANGVAEWISSN SRSQAYKVTCSVRQSSAQNR KYTIKVEV PKVATQTVGGVE LPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK**(D)**
9539, (9424)	**11-99,(12-99)**	**(D)**NGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSA QNRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNS**(D)**
9037, (8922)	**17-99, (18-99) or 18-100**	**(D)**VTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRK YTIKVEVPKVATQTVGGVE LPVAAWRSYLNMELTIPIFATNS**(D)**

Table 4.1: List of experimentally observed peptide masses (Figure 4.1C) and the corresponding amino acid sequences assigned.

MASCOT Search Results

The identification of bacteriophage MS2 was accomplished by submitting all masses in the mass spectrum (Figure 4.1C), excluding the coat protein precursor ion and its doubly charged ion. These masses were not included since they were used as internal calibrants. When submitted with a peptide mass tolerance of ±3.0 Da., MASCOT provided a score of 226 and an expect value of 9.9e-17 based on 13 peptide matches to the sequence of the bacteriophage MS2 coat protein. In addition, the bacteriophage ZR coat protein, which has identical amino-acid sequence, was also identified.

Two other protein sequences recorded MASCOT scores above the 95% identity threshold (MASCOT score > 78). A MASCOT score of 143 and expect value of 2e-8 based on 9 peptide matches was reported for the coat protein sequence of Enterobacteria phage MS2 strain bacteriophage R17 and for a second version of the coat protein sequence from bacteriophage ZR. A MASCOT score of 86 and expect value of 0.0088 based on 7 peptide matches was reported for the coat protein sequence of bacteriophage BO1.

All protein sequences with significant MASCOT scores were from the Enterobacteria phage MS2 species. Furthermore, coat protein sequences from other strains - f2, fr, M12, and JP501 - did not match with significant MASCOT scores.

Figure 4.3 shows a multiple sequence alignment of the Enterobacteria phage MS2 strains' coat protein sequences (gi|83642771, gi|15083, gi|83642775, gi|116695, gi|59799312, gi|6630977, gi|7689122, gi|15073). Sequences of strains

ZR, MS2, B01, M12, JP501, and fr represent the results of recent resequencing efforts [60].

The recent results cast doubt on the correctness of the Asp/Asn flip observed in the f2 and R17 strains, as this flip was present in older versions of the ZR sequence and corrected. [60] Each of the other differences between the strains' sequences is found in at least one of the resequenced strains.

Given their identical sequences, the MS2 strain coat protein cannot be distinguished from the ZR coat protein strain. Nevertheless, despite considerable sequence homology, the MS2 strain coat protein can be distinguished from the remainder of the strains. The bacteriophage MS2 coat protein sequence and the coat protein from bacteriophage R17 differ in intact molecular weight by a single Da. Therefore, based on intact coat protein mass measurements alone, bacteriophage MS2 could not have been distinguished from bacteriophage R17 at the mass accuracy used. However, the high sequence coverage and specificity provided by this chemical digestion method enables such discrimination.

```
f2/R17                      ND    N
M12                               X
MS2/ZR/B01      MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYT
JP501                E       E  N                   D
fr                  EE             K                                    N
                ***** :****::* **:*.* **************:******************:*****

BO1                                                                     P
f2                                        L
M12                                                   A
MS2/ZR/R17      IKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP
JP501              A                                  A
fr              V          VQ             M       V    D  A     L  TF T    A
                :** *******. *************:*:*****:****.** *****:** :* ****.

R17/M12/JP501
MS2/ZR/B01/f2   SAIAANSGIY
fr              T
                :*********
```

Figure 4.3: Multiple sequence alignment of all coat protein sequences from Enterobacteria phage MS2 strains. Identical amino-acids not shown.

It can be seen that the amino acid sequences of MS2 and R17 differ by only a sequence reversal at positions 11 and 12 and an Asp/Asn replacement at position 17. The Asp/Asn substitution, also observed in the resequenced JP501 strain, allowed for their discrimination. This example demonstrates that the Asp specificity of this acid digestion method provides a means of differentiating between the presence of an Asp or Asn residue in an amino acid sequence, even when low accuracy measurements are used. [61] In the case of bacteriophage BO1, a single amino-acid change is sufficient to disrupt 6 of the peptide mass matches.

These observations make two things abundantly clear. First, even complete amino-acid sequence determination may fail to discriminate two strains or species if the observable proteins have exactly the same sequence.[39] Second, our ability to use amino-acid sequences as a reference to determine strain or species is heavily dependent on the quality of the amino-acid sequence itself and its taxonomic classification.

Chapter 5: Analysis of Human Adenovirus Type 5

Introduction

In the preceding chapter, microwave-accelerated acid digestion combined with MALDI TOF MS was utilized for the reliable identification of a model virus, RNA bacteriophage MS2. The simplicity of the bacteriophage MS2 proteome markedly facilitated its identification. In this analysis, we evaluated the use of this strategy for identification of a virus sample with a more complex proteome in the presence of its host cell lysate. The DNA virus, human adenovirus Type 5, was characterized using both MALDI TOF LID and LC ESI MS/MS experiments.

Human Adenovirus is a non-enveloped virus that contains 11 major structural proteins. *In silico* digestion of adenovirus proteins, produce 229 Asp-specific peptides. To facilitate rapid characterization and identification of digestion products, the Rapid Microorganism Identification Database (RMIDB)[62] was used to match observed digestion product masses to predicted masses of Asp-specific cleavage products from all adenovirus proteins. In combination with peptide mass matching, we evaluated the use of MALDI TOF LID to identify peptides from a complex digestion mixture. Futher characterization of digest products was accomplished using LC tandem mass spectrometry.

Materials and Methods

Sources of Samples

Preparation of "crude", "enriched" and "pure" Virus Samples

Human Adenovirus Type 5 (ATCC # VR-5) was cultured in a HeLa cell host (ATCC # CCL-2). HeLa cells were grown until 80-90% confluence was achieved. HeLa cells were infected with adenovirus type 5 stock and incubated for 1.5 hours with occasional rocking. The virus stock was removed and 2% FBS cell maintenance media was added. Human Adenovirus Type 5 was allowed to grow in the HeLa cell host for 2.0-2.5 days until cytopathic effect was observed. The media was removed and 0.25% trypsin-EDTA solution was added to detach infected cells from the flasks. The cells were collected and centrifuged at 8000 g for 20 min. The cells were washed twice with 0.1 M PBS pH 7.1 and resuspended in 10mM Tris-HCl buffer. Analysis was performed directly on this sample without any further purification. This sample will be referred to as "crude" in this chapter.

For the preparation of an "enriched" adenovirus sample, the same growth procedure for the preparation of the crude adenovirus type 5 samples was carried out. However, upon harvesting the infected cells, 3 freeze/thaw cycles were performed followed by 3, 30 s sonication periods. The cell debris was pelleted by centrifugation at 8000 g for 20 min and the supernatant was

collected. The supernatant was loaded on top of a CsCl gradient and spun at 26,000 rpm for 2 hr. The virus band was recovered and stored at -80 C until use.

A “pure” adenovirus was prepared by performing an additional high speed centrifugation step on the once banded, enriched sample. Cesium Chloride was added to the enriched sample and spun at 33000 rpm fro 16.5 hr. The virus band was recovered and stored at -80 C until use.

Microwave-accelerated acid digestion of crude adenovirus samples

An aliquot of the adenovirus sample was acidified to 12.5% with glacial acetic acid in a 200 µl glass tube. Digestions were carried out using a CEM Benchmate laboratory microwave system (CEM Corp., Matthews, NC) equipped with 45 ml digestion vessel and a fiber optic temperature probe. This microwave setup allowed the digestions to be performed on a microscale. Digestions were carried out at a constant temperature of 140 ± 5 ˚C. for 90 sec and 2 min. for MALDI TOF analyses and LC-MS/MS analyses, respectively. For MALDI TOF LID experiments, reaction times were extended to 5 min to improve peptide ion intensities.

MALDI-TOF Mass Spectrometry of Human Adenovirus Type 5

Before digestion, intact proteins from crude, enriched and pure human adenovirus were released and solubilized directly on the MALDI target and analyzed by MALDI-TOF mass spectrometry. For MALDI TOF analyses, 0.5 µL of Sinapinnic acid 50%/50% (v/v) acetonitrile/deionized water containing 1.0% trifluoroacetic acid was applied to the sample plate and allowed to air dry. A half

µL of virus suspension was combined with 0.5 µL of a 1:1 mixture of 1% Triton X-100 and 50% acetic acid on the sample plate and allowed to air dry. A half µL of sinapinnic acid was applied to the sample plate and allowed to air dry. For the analysis of the enriched and pure adenovirus sample, drop dialysis was performed on the virus sample to assist in the removal of residual CsCl from the purification step. All MALDI mass spectra were acquired using an Axima CFR Plus MALDI-TOF mass spectrometer. Spectra were acquired with a 337 nm N_2 laser in linear mode.

For the analysis of chemical digestion products, 0.3 µL of the digested virus samples were applied to the sample plate and allowed to air dry, 0.3 µL of CHCA matrix solution was added (10 mg/ml in 70%/30% (v/v) acetonitrile/deionized water containing 0.1% trifluoroacetic acid) was applied to the sample plate and allowed to air dry. Peptide mass spectra were collected in linear mode on an Axima CFR Plus MALDI-TOF mass spectrometer. For LID analyses, spectra were collected in reflectron mode as averages of 500 profiles.

LC ESI MS/MS Analysis

For LC ESI MS/MS experiments, the digestion product suspension was centrifuged to remove the host cell debris. The supernatant was desalted using a C_{18} ziptip, dried, and then resuspended in 40 µl of 0.1% formic acid. Ten µl was loaded onto a 15 cm C_{18} reverse phase column (LC Packings) at a flow rate of 0.20 µl/min. The peptides were then gradient eluted using a 3 to 97% ACN

gradient over 90 min. MS and MS/MS spectra were acquired on an Applied Biosystems Q-Star Q-TOF mass spectrometer.

Database Searching

The Rapid Microorganism Identification Database (RMIDB)

All m/z values from 900-2500 obtained using MALDI TOF ms were submitted to the RMIDB with a tolerance of 1.5 Da. Searches were performed against Asp-specific peptide masses (up to 1 missed cleavage) generated *in silico* from all adenovirus protein sequences compiled from Swiss-Prot, Genbank, TIGR, and TrEMBL for preliminary peak assignments.

MASCOT Searches

LID spectra were submitted to MASCOT with mass tolerance's of +/- 1.0 Da and +/- 1.5 Da for parent and fragment ions, respectively. Spectra were searched against all NCBInr entries, using the “formic acid” cleavage option and “other viruses” as the taxonomic search parameter.

For LC MS/MS data, spectra were searched with mass tolerances of +/- 1.0 and +/- 0.8 Da. for parent and fragment ions, respectively. Spectra were searched against all proteins entries in the Swiss-Prot database under the “other viruses” taxonomic classification using our modified in-house “formic acid” option as the specified cleavage agent.

Results and Discussion

Microwave-accelerated acid digestions were carried out on “crude” adenovirus samples to demonstrate that purification or fractionation was not required for detection of peptide biomarkers. Prior to digestion, the intact proteins from the crude adenovirus were analyzed by MALDI TOF MS, and compared to a spectrum obtained from the purified sample. **Figure 5.1** is a superimposed mass spectrum of the crude and purified adenovirus samples. While many additional host protein signals were detected in the crude sample as expected, a very good correlation between the intact virus protein fingerprints of the purified and crude samples was observed.

Following only a 90 sec digestion, a multitude of peptide signals was observed by MALDI TOF analysis. A representative MALDI TOF mass spectrum of human adenovirus obtained following a 90 sec. digestion in the presence of 12.5% acetic acid is shown as **Figure 5.2a**. Several peptide pairs, separated by 115 Da., were observed in the mass spectrum, which again provides further evidence Asp-specific cleavage. All observed peptide masses (**Figure 5.2a**) were searched against predicted masses generated by an Asp-specific *in silico* digestion using the RMIDB to obtain tentative qualitative information about the observed peptide products.

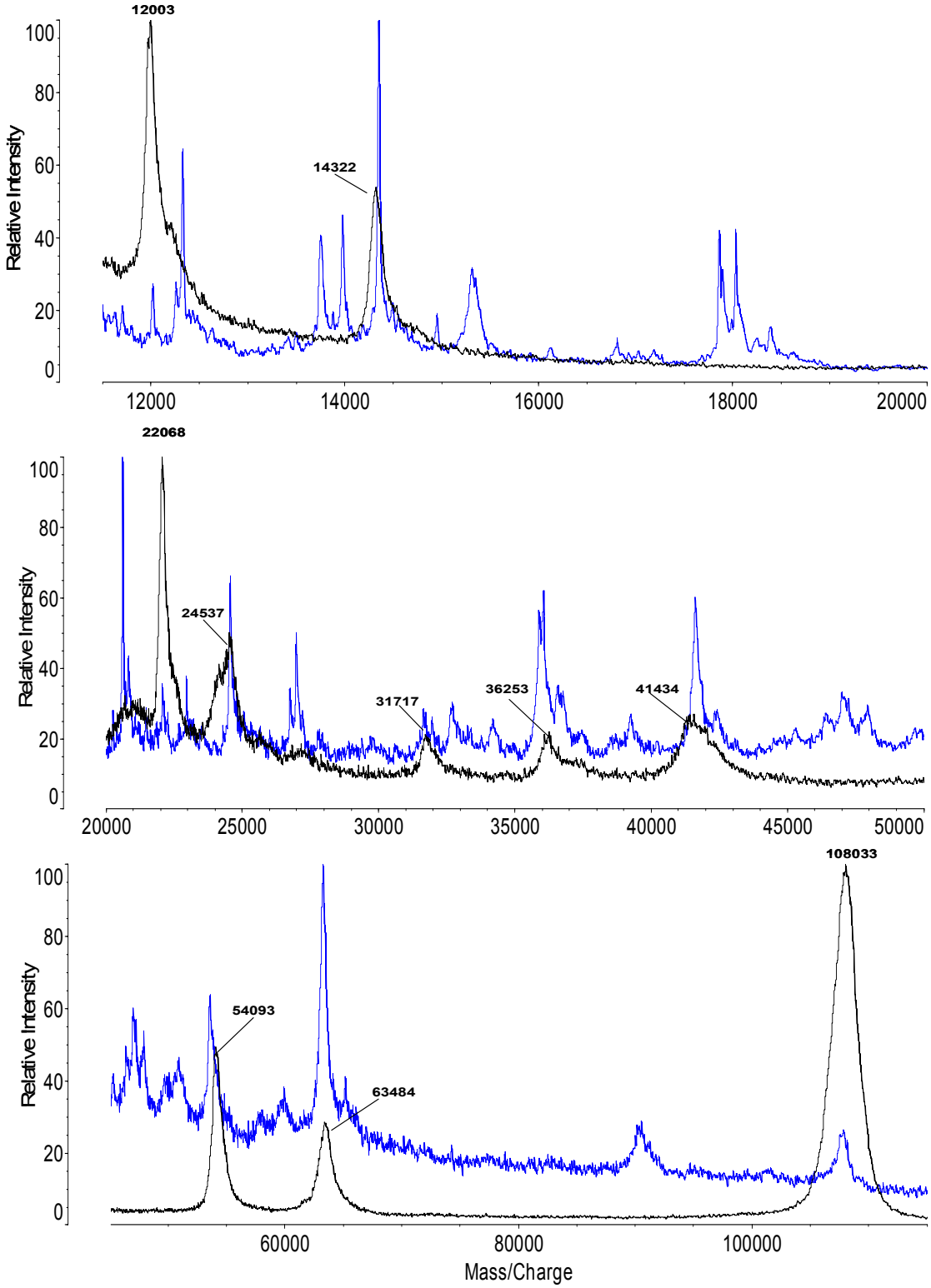


Figure 5.1: Superimposed mass spectra of "crude" Human adenovirus Type 5 (blue) and "pure" Human adenovirus Type 5 (black) both treated with a 1:1 mixture of 1% Triton X-100 and 50% Acetic acid from m/z 11500-20000 (A) 20000-50000 (B) and 50000-110000. Structural proteins are labeled in pure spectrum.

Interestingly, twenty-one peptides produced matches to predicted masses from the adenovirus hexon protein. The hexon protein is a capsid protein present at 720 copies per viron, making it the second most abundant structural protein in the adenovirus viron. [63] While it is expected that the most abundant structural proteins would be best represented in a mass spectrum of the digestion products, this observation also correlates well with previous reports that capsid proteins are readily released and solubilized from intact viruses with the addition of organic acids. [56, 64] Additionally, a lower number of peptides (2-6) were tentatively assigned to the penton base, IVa2, fiber, core (VII), hexon-associated, and terminal proteins from adenovirus.

The feasibility of using LID to confirm the peptide identities was demonstrated for the peptide at m/z 1183.3. When searched in MASCOT, this peptide was confidently identified with a score 30 and an e-value of 0.023. This peptide was found to originate from the human adenovirus type 5 fiber protein. The peptide at 1298.0 corresponds to the mass of the same peptide with an aspartic acid residue retained at the C-terminal. **Figure 5.3** is an LID fragmentation spectrum of the peptide at m/z 1183.3.

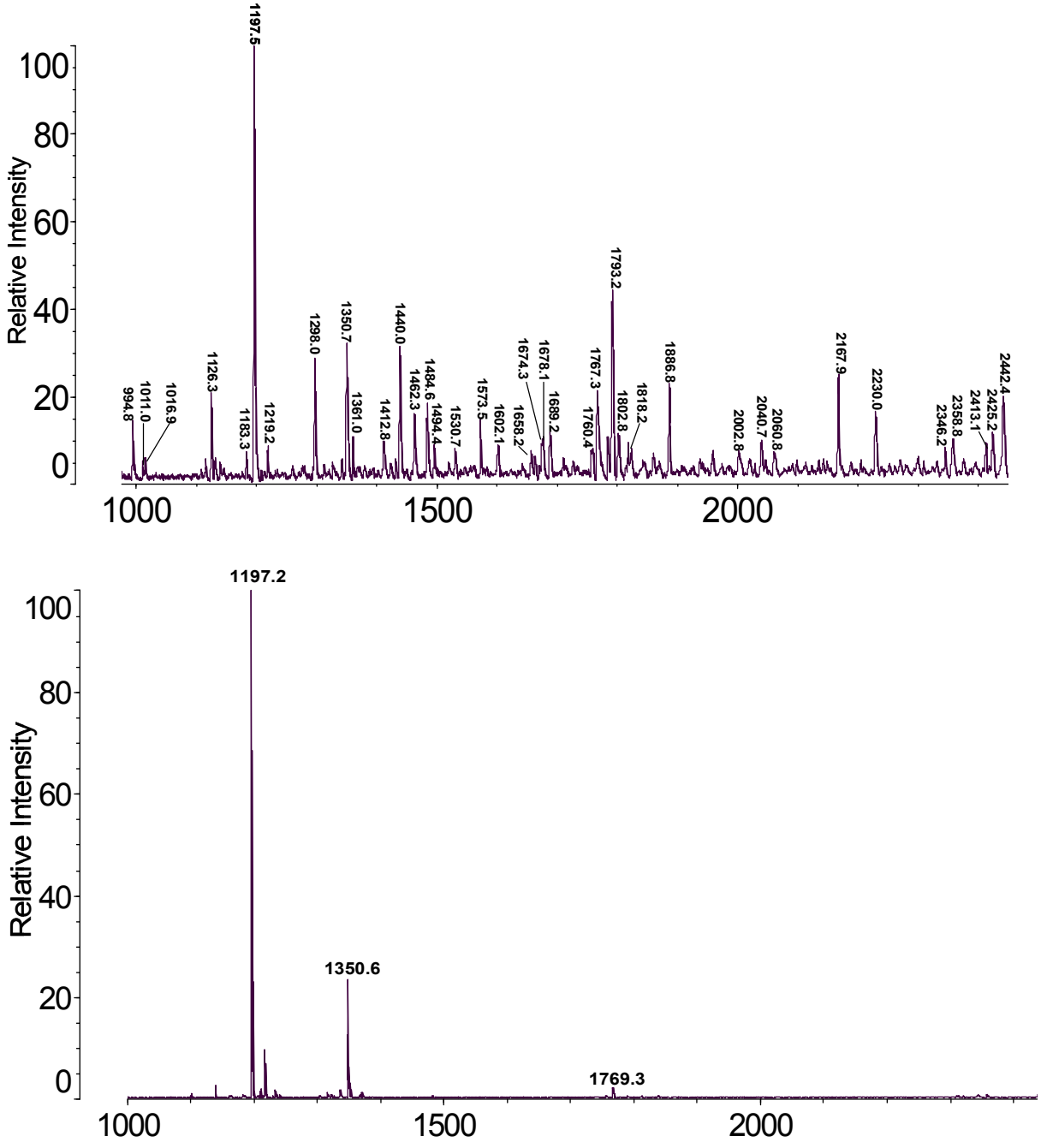


Figure 5.2: Mass spectrum of digestion products from adenovirus proteins following a 2 min. digestion (A) with no digestion (B).

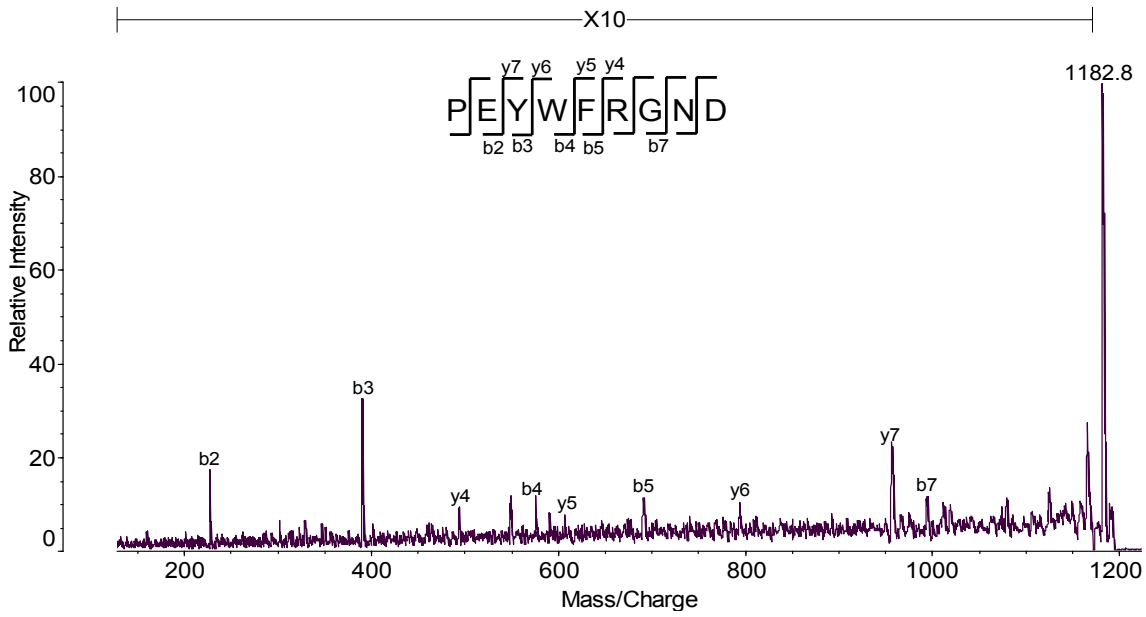


Figure 5.3: Fragmentation spectrum (LID) of m/z= 1182.8

As a control, a mass spectrum of the undigested sample was obtained. **Figure 5.2b** is a mass spectrum of adenovirus prior to digestion (m/z 1000-2500). Only two intense peptide signals were detected in the undigested sample at m/z 1197.2 and 1350.6. We attempted to deduce the identity of the peptide at 1197.2 using MALDI TOF LID and database searching. **Figure 5.3** is a MALDI TOF LID mass spectra of m/z 1197.2.

When the LID spectrum was submitted to MASCOT, the top peptide hit was for the peptide sequence, GLRFPSKMFGG, which originates from the human adenovirus major core protein VII. While this peptide produced a sufficient fragmentation spectrum for database submission, a MASCOT score of 22 for this peptide was found to be insignificant because it was below the 95% identity threshold. This can be attributed to the inability to provide enzyme specificity since this peptide is not a product of the microwave-accelerated acid digestion. However, upon further investigation of the major capsid protein VII primary sequence, this peptide was found to be part of the signal peptide, MSILISPSNNTGWGLRFPSKMFGG.[65, 66] Several reports have shown that prior to incorporation of the major capsid protein VII into the mature viron, this peptide is cleaved by the Adenovirus Protease.[67] Cleavage of adenovirus proteins by the Adenovirus protease is known to be specific for the motif, GZ/GXG/Z, where Z = any amino acid, and X indicates the position of cleavage. [67]

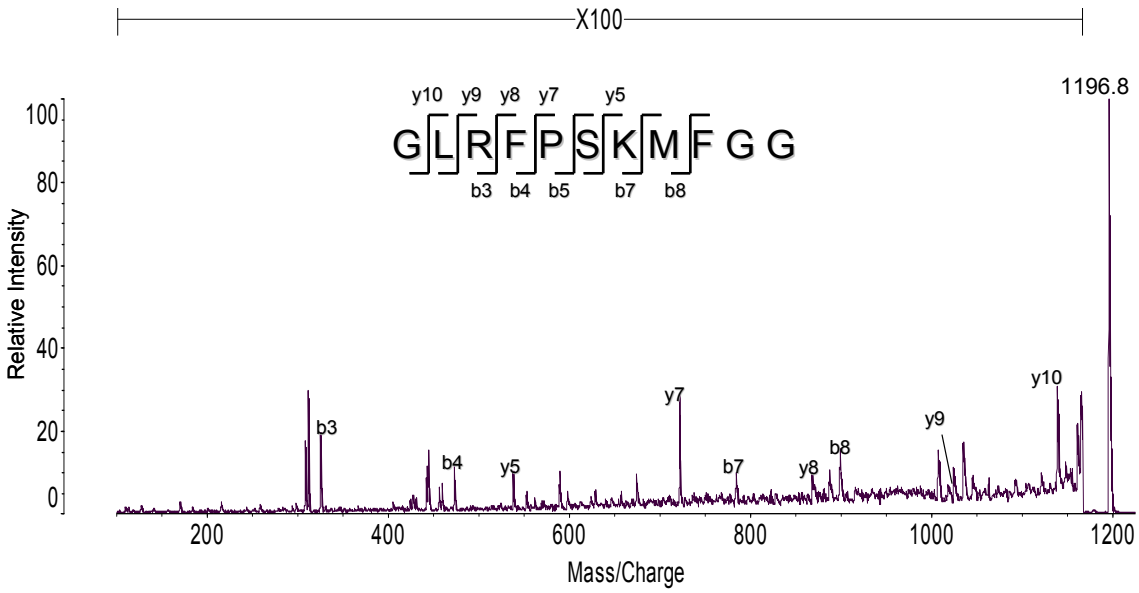


Figure 5.4: Fragmentation (LID) Spectrum of m/z 1196.8.

These observations lead us to hypothesize that the additional peak at 1350.6 in the control spectrum may also be a product of Adenovirus Protease cleavage. The Human adenovirus minor capsid protein VI is also known undergo processing by the Adenovirus Protease at the C-terminus at position 239. Cleavage on the c-terminus of ^{239}L results in the production of the peptide, GVQSLKRRRCF, which has a calculated molecular weight of 1351.0 Da.

We conclude that both the peptides detected in Figure 5.2b are signal peptides removed by normal adenovirus protease activity. It is important to note that the activity of the adenovirus protease is required for production of infectious virus particles. The ability to rapidly detect these peptide products may be of importance to researchers interested in surveying newly engineered protease inhibitors for prevention of viral infections.

LC ESI MS/MS of Adenovirus digestion products

In addition to evaluating Asp-specific chemical cleavage for rapid adenovirus identification, the peptide products of a 2 min microwave-accelerated acid digestion of adenovirus were separated using HPLC and analyzed by tandem mass spectrometry. In this experiment, five proteins were identified from Human Adenovirus Type 5 with significant scores (>28) based on 9 peptide identifications with MASCOT e-values below 0.05. **Figure 5.5a** displays the resulting Total Ion Chromatogram from the peptide mixture. Also displayed is an example a TOF MS survey scan, XIC, and tandem mass spectrum **(Figures 5.5**

b-d). **Table 5.1** summarizes the MASCOT search results of the LC MS/MS analysis.

In this analysis, microwave-accelerated acid hydrolysis served to denature the adenovirus and subsequently digest its major structural proteins. The strength of MALDI TOF LID for rapid protein identification from a complex mixture of peptides was demonstrated. Sequence information obtained from a peptide originating from the adenovirus fiber protein, without prior fractionation, provided a confident identification of the virus when combined with bioinformatics. Additionally, we have shown that after only a 2 min digestion, chemically generated Asp-specific peptides enabled identification of 5 major structural proteins when analyzed using LC MS/MS.

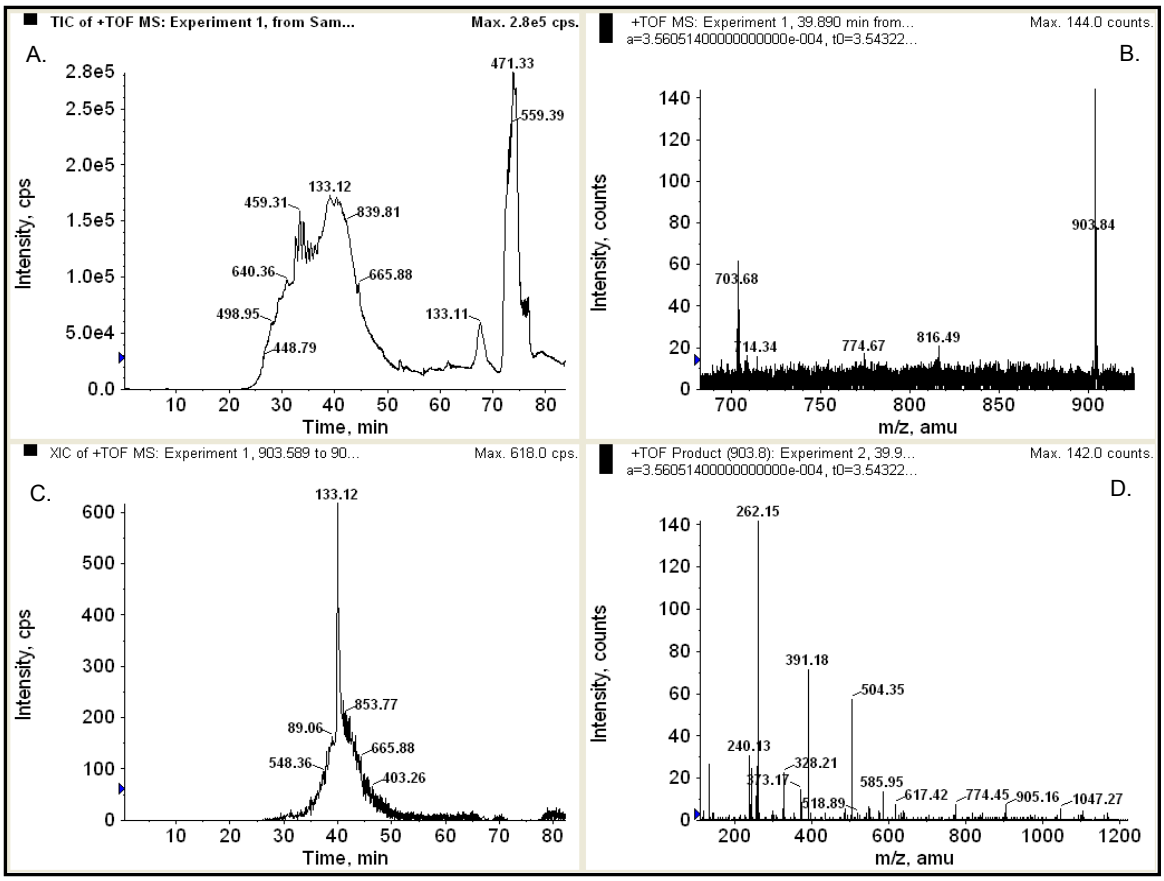


Figure 5.5: TIC (A) TOF Mass Spectrum (B) XIC (C) and MS/MS mass spectrum of m/z 903.84 (D) from HPLC Q-TOF analysis.

Precursor m/z	Peptide mass	Peptide score	Peptide e-value	Peptide Sequence	Protein Source
769.92	1537.82	60	3.50E-05	D.PYYTYSGSIPYLD.G	Hexon Protein
722.75	2165.22	32	0.018	D.NPNTYDYMNKRVVAPGLVD.C	Hexon Protein
749.92	1497.82	55	0.00012	D.SLTAPSEFATTASTD.A	Late 100 kDa Protein
865.51	2593.5	52	0.00019	D.AANAPTTFPVEAPPLEEEEVIIEQ.D	Late 100 kDa Protein
903.84	2708.5	74	1.20E-06	D.AANAPTTFPVEAPPLEEEEVIIEQD.P	Late 100 kDa Protein
600.89	1199.77	38	0.0062	D.DKLTALLAQLD.S	Hexon-assoc. Protein IX
589.37	1765.09	32	0.029	D.LRQQVSALKASSPPNAV.-	Hexon-assoc. Protein IX
647.72	1940.15	29	0.43	D.FAFLSPLASSAASRSSARD.D	Hexon-assoc. Protein IX
649.33	1296.65	40	0.0039	D.PEYWNFRNGD.L	Fiber Protein
682.05	2043.11	36	0.0093	D.PVTGLVMPGVYTNEAFHPD.I	Penton Protein

Table 5.1: MASCOT search results from LC MS/MS analysis of Human Adenovirus Type 5 digestion products.

Chapter 6: Analysis of *Saccharomyces cerevisiae* Ribosomal Proteins

Introduction

Previously, we demonstrated the usefulness of Asp-specific peptides generated by microwave-accelerated acid hydrolysis for the identification and characterization of protein biomarkers released from microorganisms. In each of these studies, the number of protein precursors, and therefore, constituent peptide products was limited. Consequently, protein identification, and thus identification of the microorganism was feasible using manual mass spectrometric based methods.

To demonstrate the suitability of integrating this method into an automated, high throughout proteomics workflow, a larger protein pool was studied. Ribosomes, isolated from *Saccharomyces cerevisiae* were directly processed to proteotypic peptides by microwave-assisted acid hydrolysis and analyzed by LC ESI and MALDI tandem mass spectrometry. The ribosome of *Saccharomyces cerevisiae* consists of a large and small subunit, termed 60S and 40S, respectively, based on their sedimentation coefficients. The ribosome is made up of ribosomal RNA (rRNA) and a large protein component, which contains 79 proteins present in an equimolar amount. Of these 79 proteins, 46 are found in the large subunit and 33 in the small subunit.[68, 69]

Asp-specific digestion of the ribosome proteome *in silico* produces 387 peptides in the m/z range of 500-5000. This peptidome has been theoretically shown to contain a set of peptides, that are larger and contain more basic residues on average than tryptic peptides. **Figure 6.1** displays a size distribution

comparison of Asp-specific and tryptic peptides. A comparison of the distribution of basic residues between Asp-specific and tryptic peptides is shown in **Figure 6.2**.

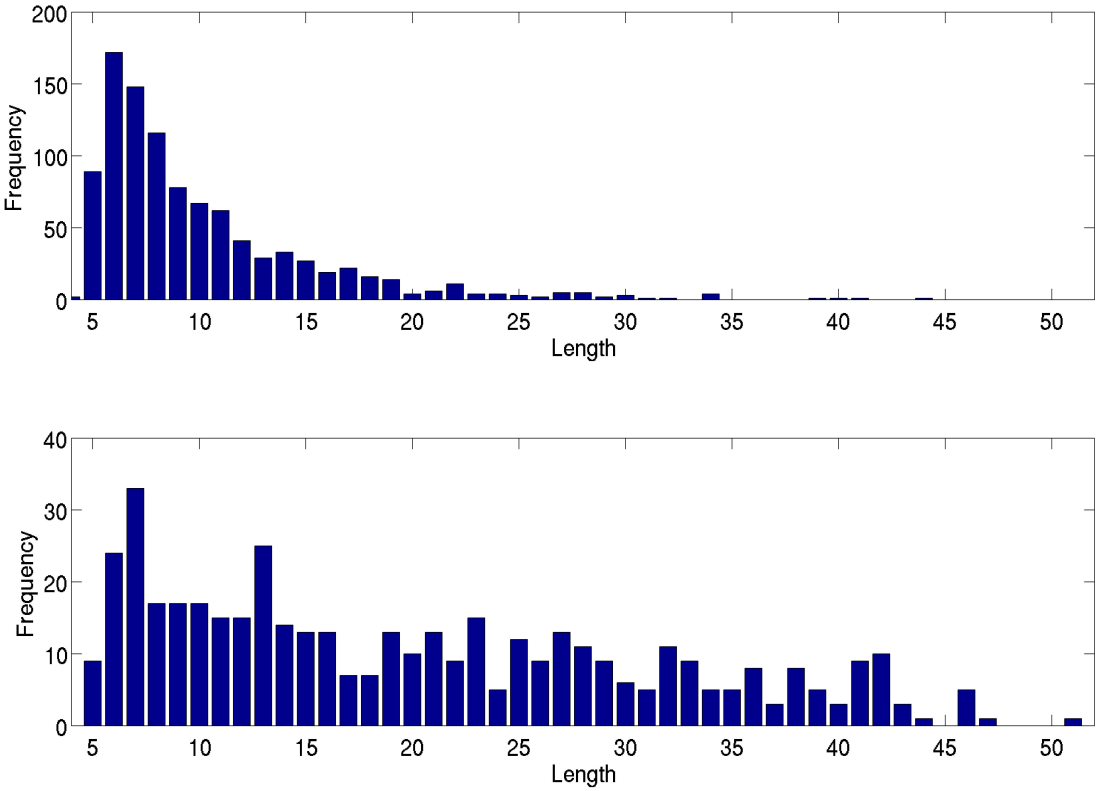


Figure 6.1: Distribution of peptide lengths from tryptic cleavage (A) Asp-specific cleavage (B).

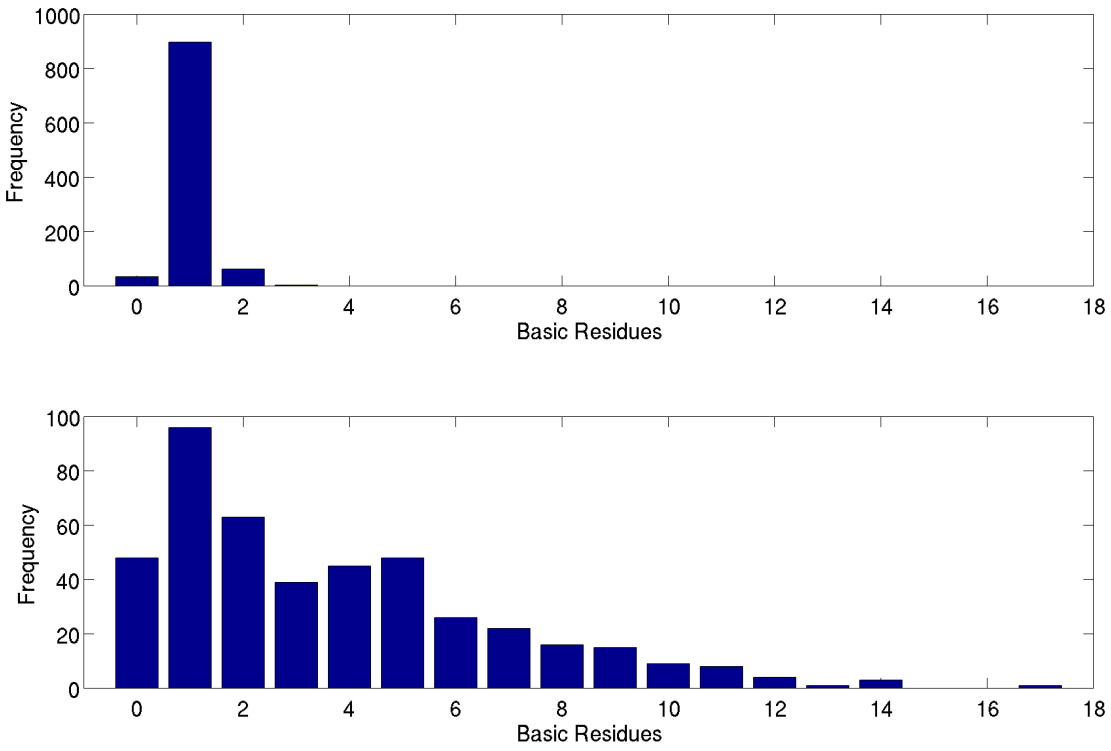


Figure 6.2: Distribution of basic residues from tryptic cleavage (A) Asp-specific cleavage (B).

Materials and Methods

Materials.

Ribosomes isolated from *Saccharomyces cerevisiae* were obtained from the laboratory of Dr. Jon Dinman at the University of Maryland. Ribosomes were prepared at 4.6 pmol/µl. The ribosomes were salt washed during the purification process to remove ribosome-associated factors.

Methods.

Microwave-Accelerated Acid Hydrolysis

Ten µL of intact ribosomes was diluted ten fold with Milli-Q grade water. A 100 µL aliquot of glacial acetic acid was added to the ribosome suspension. To isolate the precipitated RNA[68], samples were spun at 14,000 rpm for 10 min on a benchtop microfuge. The supernatant was collected and the acid concentration was adjusted to 12.5% acetic by the addition of 600µL of Mill-Q grade water. A 100 µL aliquot was added to a 200 µl glass sample holder and placed in a 45 ml digestion vessel. The digestion vessel was closed and placed in a Discover Benchmate (CEM Corp., Matthews, NC) microwave system equipped with a fiber optic temperature probe. After initial optimization, all digestions were carried out at a constant temperature of 140 ± 5 ˚C for 20 min.

MALDI TOF Analysis of Ribosomes from Saccharomyces cerevisiae

For MALDI-TOF analysis of ribosomal protein digestion products, 0.3 µL of the digested sample was applied to the MALDI sample plate and allowed to air dry. 0.3 µL of CHCA matrix solution (10 mg/ml in 70%/30% (v/v)

acetonitrile/deionized water containing 0.1% trifluoroacetic acid) was then added to the sample plate and allowed to air dry. All mass spectra were acquired with a 337 nm N_2 laser on a Shimadzu Axima CFR Plus mass spectrometer. Peptide mass spectra were acquired in Linear mode.

LC LTQ-Qrbitrap Analysis of Ribosomes from Saccharomyces cerevisiae

Following digestion of the ribosomal proteins, the digestion products were dried and resuspended in 40 µl of 0.1% formic acid. The peptide mixture was separated using nano-LC system (Microtech, Vista, CA) and subsequently analyzed using ESI mass spectrometry. Analyses were carried out using an LTQ-Orbitrap mass spectrometer (Thermo Electron, San Jose, CA). Peptide mixtures were injected onto a 50 cm C_{18} reverse phase column and then gradient eluted using 5-60% ACN over 130 min. Full MS scans were carried out in the Orbitrap and MS/MS experiments were performed in the LTQ. Precursor ions were automatically selected and fragmented using collisionally induced dissociation.

LC MALDI TOF-TOF Analysis of Ribosomes from Saccharomyces cerevisiae

Ribosomal protein digests were separated using HPLC and subsequently spotted on the MALDI target using a Shimadzu Accuspot. The Accuspot provided automatic mixing of LC elutents and the CHCA matrix solution prior to deposition onto the MALDI surface. All MALDI survey spectra were acquired automatically, on an Axima-TOF2™ (Shimadzu Biotech, Kratos, Manchester, UK) mass

spectrometer equipped with a 337 nm N_2 laser. Only peptides with m/z values > 3 kDa. were selected for MS/MS experiments.[70]

Database Searching

Tandem mass spectra were submitted to MASCOT for peptide and protein identifications. LTQ-Orbitrap MS/MS searches were performed with mass tolerances of ± 0.05 Da. and ± 0.8 Da. for precursor and fragment ions, respectively, allowing for up to 5 missed cleavages. Database searches were performed using the modified in-house "formic acid" option for cleavage specificity which accommodates all possible peptides generated by acid cleavage. Searches were performed against a restricted database constructed in-house that includes only proteins from the *Saccharomyces cerevisiae* ribosome.[71]

For MASCOT searches of Axima TOF^2 MS/MS spectra, mass tolerances of ± 1.0 Da. and ± 1.5 Da. were chosen for precursor and fragment ions, respectively. As with the LTQ-Orbitrap data, the modified "formic acid" option was chosen for cleavage specificity and searches were performed against the in-house *Saccharomyces cerevisiae* ribosomal protein database.

Results and Discussion

Following digestion of the ribosome samples, MALDI TOF MS was used as a rapid method for confirmation of cleavage of ribosomal proteins by microwave-accelerated acid hydrolysis. Mass spectra were collected without any prior separation of digestion products. **Figure 6.3a** is a MALDI TOF mass

spectrum (m/z 800-8000) of ribosomal protein digestion products produced by 20 min of microwave heating in the presence of 12.5% acetic acid. **Figure 6.3b** displays an enlarged region of the mass spectrum from (m/z 1800-2900) to illustrate that peptide pairs separated by 115 Da. were observed, which again provided additional evidence that cleavages were Asp-specific.

The ribosomal protein digestion products were then analyzed using LC ESI and MALDI MS/MS for large scale protein identification. **Figure 6.4** is the total ion chromatogram of the ribosomal protein digestion products following 20 min of microwave heating. A total of 387 peptides were predicted by Asp-specific digestion, and 207 peptides were identified by LC ESI MS/MS and bioinformatics. It must be noted, however, that 50 peptide pairs were identified. Therefore, the number of distinct peptides identified was 157. Peptide identifications were considered to be significant if MASCOT expect values were less than 0.05. These peptide identifications enabled identification of 58 of 79 ribosomal proteins. **Figure 6.5** displays an example of a tandem mass spectrum of a doubly charged precursor ion with m/z 600.85. This peptide was identified as, IKAVVESVGAEV, which originates from RPP2B. **Table 6.1** is a list of the peptides identified by LC MS/MS analysis.

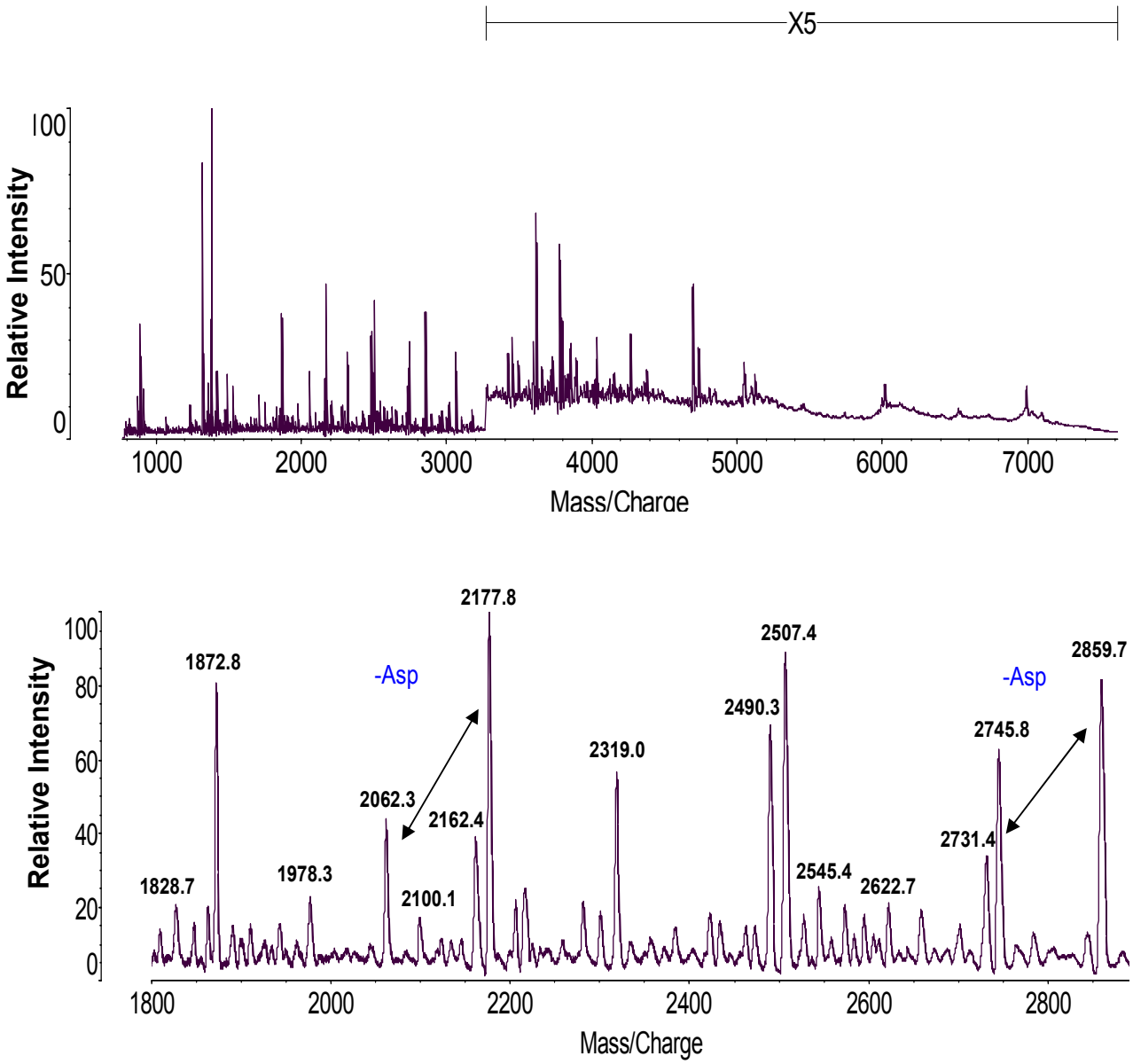


Figure 6.3: MALDI TOF mass spectrum (m/z 800-8000) of ribosomal protein digestion products (A). MALDI TOF mass spectrum (m/z 1800-2900) of ribosomal protein digestion products (B).

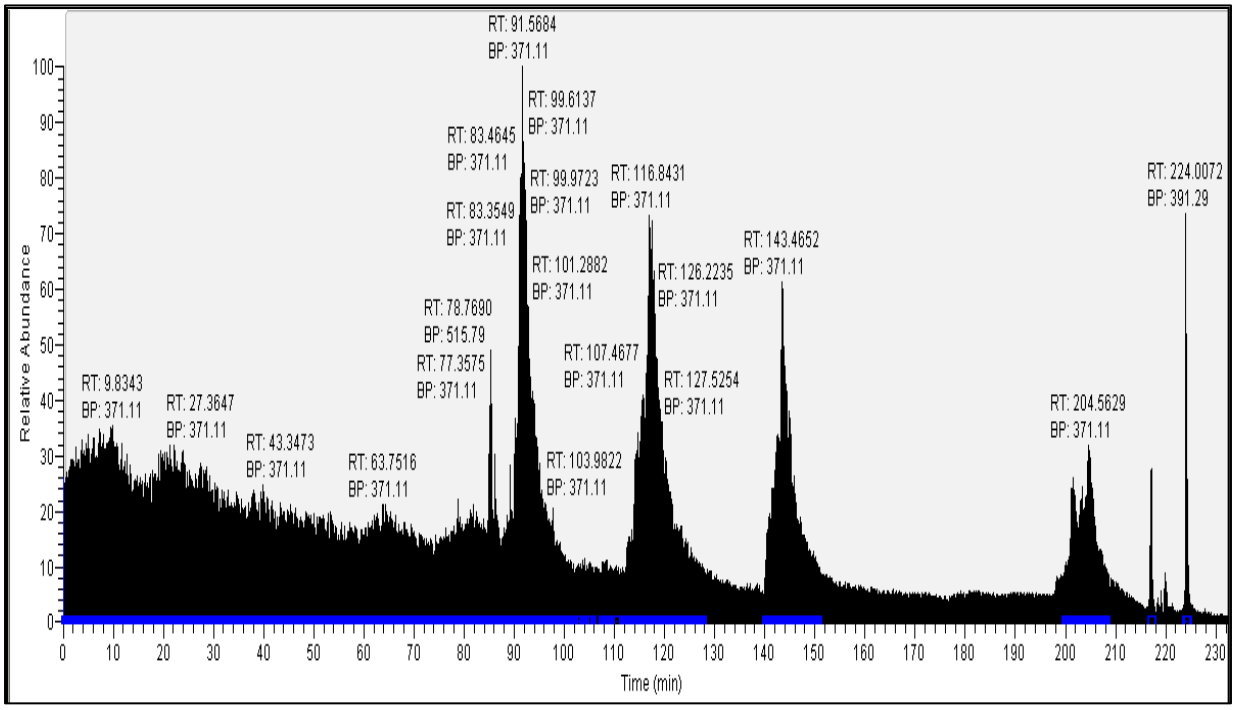


Figure 6.4: LTQ-Orbitrap TIC of ribosomal protein digestion products.

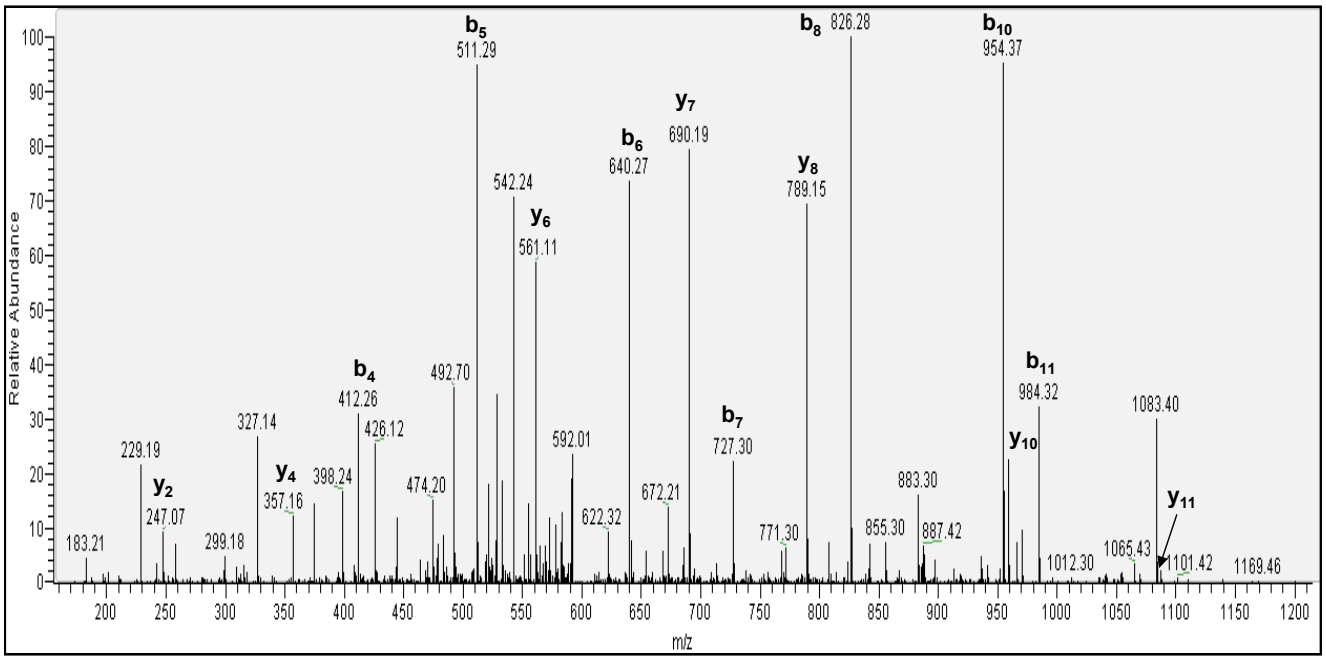


Figure 6.5: LTQ-Orbitrap tandem mass spectrum of a doubly charged peptide (m/z 600.85)

1. RPP1B

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Peptide
671.33	670.32	670.31	0.01	0	11	**0.019**	D.MGFGLF.D
690.38	689.37	689.36	0.01	0	25	**0.00044**	D.AGLEITS.D
786.36	785.35	785.34	0.01	1	12	**0.031**	D.MGFGLFD.-
786.36	785.35	785.34	0.01	1	15	**0.016**	D.DMGFGLF.D
805.40	804.40	804.39	0.01	1	19	**0.0026**	D.AGLEITSD.N
489.79	977.57	977.55	0.01	0	50	**4.8e-007**	D.VYAKALEGK.D
365.21	1092.60	1092.58	0.01	1	39	**1.2e-005**	D.VYAKALEGKD.L
643.39	1284.76	1284.74	0.02	0	78	**7.5e-010**	D.NLLTITKAAGANV.D
700.90	1399.78	1399.77	0.02	1	78	**1.6e-009**	D.NLLTITKAAGANVD.N
782.42	1562.83	1562.81	0.02	2	73	**2.3e-009**	D.NVWADVYAKALEGK.D
811.35	1620.69	1620.66	0.02	0	58	**1.7e-007**	D.AAAEEEKEEEAAEES.D
560.29	1677.86	1677.84	0.02	3	10	**0.01**	D.NVWADVYAKALEGKD.L
936.02	1870.02	1869.99	0.02	2	37	**3.2e-005**	D.NLLTITKAAGANVDNVWA.D
808.10	2421.27	2421.24	0.03	0	73	**5.1e-009**	D.LKEILSGFHNAGPVAGAGAASGAAAAGG.D
846.44	2536.30	2536.27	0.03	1	105	**2.9e-012**	D.LKEILSGFHNAGPVAGAGAASGAAAAGGD.A
700.18	3495.86	3495.81	0.05	2	9	**0.0065**	D.VYAKALEGKDLKEILSGFHNAGPVAGAGAASGAAAAGG.D

2. RPS19A

307.69	613.36	613.35	0.01	0	36	**1.2e-005**	M.PGVSVR.D
365.20	728.39	728.38	0.01	1	32	**3e-005**	M.PGVSVRD.V
515.79	1029.56	1029.55	0.01	0	42	**1.7e-005**	D.RIAAQTLEE.D
549.81	1097.60	1097.58	0.01	2	49	**6.3e-007**	M.PGVSVRDVAAQ.D
573.30	1144.59	1144.57	0.02	1	43	**1.9e-005**	D.RIAAQTLEED.E
607.32	1212.62	1212.61	0.02	3	41	**3.7e-006**	M.PGVSVRDVAAQD.F
629.84	1257.67	1257.66	0.02	2	48	**2.9e-006**	D.LDRIAAQTLEE.D
687.36	1372.70	1372.68	0.02	3	56	**1.1e-006**	D.LDRIAAQTLEED.E
694.36	1386.70	1386.68	0.02	0	83	**9.4e-010**	D.IVKTSSGNEMPPQ.D
751.87	1501.73	1501.71	0.02	1	43	**2.6e-005**	D.IVKTSSGNEMPPQD.A
800.77	2399.29	2399.26	0.03	0	41	**7.4e-006**	D.FINAYASFLQRQGKLEVPGYV.D
839.12	2514.33	2514.29	0.03	1	18	**0.003**	D.FINAYASFLQRQGKLEVPGYVD.I
538.17	3760.17	3760.12	0.05	0	8	**0.017**	D.ASGSINRKVLQALEKIGIVEISPKGGRRISENGQR.D

3. RPS19B

352.20	702.38	702.37	0.01	1	46	**7.4e-006**	M.AGVSVRD.V
515.79	1029.56	1029.55	0.01	0	42	**1.7e-005**	D.RIAAQTLEE.D
536.80	1071.58	1071.57	0.01	2	48	**7.8e-007**	M.AGVSVRDVAAQ.D
573.30	1144.59	1144.57	0.02	1	43	**1.9e-005**	D.RIAAQTLEED.E
594.31	1186.61	1186.59	0.01	3	52	**3.5e-007**	M.AGVSVRDVAAQD.F
629.84	1257.67	1257.66	0.02	2	48	**2.9e-006**	D.LDRIAAQTLEE.D
687.36	1372.70	1372.68	0.02	3	56	**1.1e-006**	D.LDRIAAQTLEED.E
694.36	1386.70	1386.68	0.02	0	83	**9.4e-010**	D.IVKTSSGNEMPPQ.D
751.87	1501.73	1501.71	0.02	1	43	**2.6e-005**	D.IVKTSSGNEMPPQD.A
800.77	2399.29	2399.26	0.03	0	41	**7.4e-006**	D.FINAYASFLQRQGKLEVPGYV.D
839.12	2514.33	2514.29	0.03	1	18	**0.003**	D.FINAYASFLQRQGKLEVPGYVD.I
538.17	3760.17	3760.12	0.05	0	8	**0.017**	D.ASGSINRKVLQALEKIGIVEISPKGGRRISENGQR.D

4. RPS3

408.25	1221.74	1221.72	0.02	0	25	**0.00018**	D.PAKSRTGPKALP.D
642.33	1282.65	1282.63	0.02	1	60	**9.3e-008**	D.GFLIHSGQPVND.F
442.64	1324.91	1324.89	0.02	0	26	**0.00012**	M.VALISKKRKLVA.D
480.99	1439.94	1439.92	0.02	1	15	**0.0014**	M.VALISKKRKLVAD.G
829.92	1657.83	1657.81	0.02	3	20	**0.0011**	D.GFLIHSGQPVNDFID.T
895.43	1788.84	1788.82	0.02	0	68	**8.1e-009**	D.YRPAEETEAQAEPVEA.-
688.40	2062.18	2062.16	0.03	0	50	**4.8e-007**	D.AVTIIEPKEEEPILAPSVK.D
726.74	2177.21	2177.18	0.03	1	52	**5.7e-007**	D.AVTIIEPKEEEPILAPSVKD.Y
482.30	2406.45	2406.42	0.03	0	36	**1.3e-005**	D.TATRHVLMRQGVLGIKVKIMR.D

5. RPL38

344.47	1373.83	1373.81	0.02	0	27	**0.00011**	D.IKQFLELTRRA.D
497.29	1488.86	1488.84	0.02	1	19	**0.0012**	D.IKQFLELTRRAD.V
412.84	2059.18	2059.15	0.03	2	33	**2.7e-005**	M.AREITDIKQFLELTRRA.D
435.85	2174.21	2174.18	0.03	3	24	**0.00019**	M.AREITDIKQFLELTRRAD.V
435.89	2174.39	2174.36	0.03	0	42	**3e-006**	D.AGKAKKLIQSLPPTLKVNRL.-

6. RPS0A

665.32	664.31	664.31	0.01	0	15	**0.0035**	D.SPSEFV.D
381.21	760.40	760.39	0.01	0	19	**0.00059**	D.LYFYR.D
780.35	779.34	779.33	0.01	1	20	**0.0022**	D.SPSEFVD.V
558.29	1114.57	1114.56	0.01	0	12	**0.012**	D.RTQPWSIMP.D
615.81	1229.60	1229.59	0.02	1	15	**0.011**	D.RTQPWSIMPD.L
951.05	1900.09	1900.07	0.02	0	60	**9.9e-008**	D.AQAIKEASYVNIPVIALT.D
1008.57	2015.12	2015.09	0.03	1	66	**5.5e-008**	D.AQAIKEASYVNIPVIALTD.L
602.56	3007.78	3007.73	0.04	0	9	**0.012**	D.GVHVINVGKTWEKLVLAARIIAAIPNPE.D
646.36	3226.75	3226.71	0.04	0	43	**4.9e-006**	D.AQLLLAANTHLGARNVQVHQEPYVFNARP.D
669.36	3341.78	3341.74	0.04	1	30	**0.00018**	D.AQLLLAANTHLGARNVQVHQEPYVFNARPD.G

7. RPL12A

379.19	756.36	756.36	0.01	1	18	**0.0029**	D.FKNPHD.I
413.24	1236.71	1236.69	0.02	0	17	**0.0018**	D.KNVKHSGNIQL.D
420.24	1257.70	1257.69	0.02	0	9	**0.024**	D.EIIEIARQMR.D
451.59	1351.74	1351.72	0.02	1	17	**0.004**	D.KNVKHSGNIQLD.E
519.29	2591.43	2591.40	0.04	2	18	**0.0017**	D.KNVKHSGNIQLDEIIEIARQMR.D
770.65	3848.20	3848.16	0.05	0	75	**6.4e-009**	D.PNEVKYLYLRAVGGEVGASAALAPKIGPLGLSPKKVGE.D
793.65	3963.23	3963.18	0.05	1	59	**2.9e-007**	D.PNEVKYLYLRAVGGEVGASAALAPKIGPLGLSPKKVGED.I

8. RPS12

604.38	1206.75	1206.73	0.01	0	41	**4.3e-006**	D.PENKVPLIKVA.D
330.72	1318.86	1318.84	0.02	0	31	**4e-005**	D.ALKVVLRTALVH.D
441.60	1321.78	1321.76	0.02	1	46	**2.4e-006**	D.PENKVPLIKVAD.A
685.90	1369.79	1369.77	0.02	0	68	**8.9e-009**	D.AKQLGEWAGLGKI.D
359.48	1433.89	1433.87	0.02	1	40	**9.4e-006**	D.ALKVVLRTALVHD.G
495.95	1484.82	1484.80	0.02	1	46	**2.5e-006**	D.AKQLGEWAGLGKID.R
607.34	2425.35	2425.32	0.03	0	11	**0.0038**	D.REGNARKVVGASVVVVKNWGAET.D

9. RPP2A

671.33	670.32	670.31	0.01	0	11	**0.019**	D.MGFGLF.D
786.36	785.35	785.34	0.01	1	12	**0.031**	D.MGFGLFD.-
786.36	785.35	785.34	0.01	1	15	**0.016**	D.DMGFGLF.D
811.35	1620.69	1620.66	0.02	0	58	**1.7e-007**	D.AAAEEEKEEEAAEES.D
549.97	1646.89	1646.87	0.02	1	8	**0.016**	D.EKVSSVLSALEGKSVD.E
851.12	2550.32	2550.29	0.03	0	70	**4.8e-009**	D.ELITEGNEKLAAVPAAGPASAGGAAAASG.D

10. RPS5

787.47	786.46	786.45	0.01	1	24	**0.00044**	D.IINVLTD.Q
479.76	957.50	957.49	0.01	0	53	**7.2e-007**	D.AITNTGPRE.D
501.25	1000.48	1000.47	0.01	0	38	**1.7e-005**	D.TEAPVEVQE.D
537.27	1072.53	1072.51	0.01	1	40	**9e-006**	D.AITNTGPRED.T
401.23	1200.67	1200.66	0.02	0	11	**0.0039**	D.ELERVAKSNR.-
486.28	1455.82	1455.80	0.02	0	25	**0.00016**	D.TTRVGGGGAARRQAV.D
393.72	1570.85	1570.83	0.02	1	22	**0.00063**	D.TTRVGGGGAARRQAVD.V
628.59	2510.34	2510.31	0.03	2	9	**0.02**	D.AITNTGPREDTTRVGGGGAARRQAV.D
657.35	2625.37	2625.33	0.03	3	8	**0.031**	D.AITNTGPREDTTRVGGGGAARRQAVD.V

11. RPP0

671.33	670.32	670.31	0.01	0	11	**0.019**	D.MGFGLF.D
786.36	785.35	785.34	0.01	1	12	**0.031**	D.MGFGLFD.-
786.36	785.35	785.34	0.01	1	15	**0.016**	D.DMGFGLF.D
531.29	1060.57	1060.56	0.02	0	15	**0.0017**	D.NGQVFPSSIL.D
766.82	1531.63	1531.62	0.02	0	84	**2.1e-010**	D.AAPAEEAAAEEEEES.D
682.69	2045.04	2045.02	0.03	0	50	**4.9e-007**	D.RIENPEKYAAAAPAATSAASG.D
721.03	2160.07	2160.04	0.03	1	85	**5.8e-010**	D.RIENPEKYAAAAPAATSAASGD.A
791.75	2372.23	2372.20	0.03	2	44	**2.2e-006**	D.LVDRIENPEKYAAAAPAATSAASG.D

12. RPS21B

515.77	1029.53	1029.52	0.01	1	36	**7.2e-005**	D.SLNRLAQND.G
583.35	1164.68	1164.66	0.01	0	65	**2.9e-008**	D.HASVQINVAKV.D
427.58	1279.70	1279.69	0.02	1	26	**0.00054**	D.HASVQINVAKVD.E
661.87	1321.73	1321.71	0.02	0	46	**5.3e-006**	D.GLLKNVWSYSR.-
465.92	1394.73	1394.72	0.02	2	35	**6.2e-005**	D.DHASVQINVAKVD.E
690.85	2759.39	2759.35	0.04	0	54	**2.2e-007**	D.EEGRAIPGEYITYALSGYVRSRGES.D

13. RPS7A

602.32	601.32	601.31	0.01	1	19	**0.0023**	D.VQQID.Y
409.23	816.45	816.43	0.01	2	34	**4.4e-005**	D.SKDVQQI.D
466.74	931.47	931.46	0.01	3	30	**0.0002**	D.SKDVQQID.Y
490.81	2938.83	2938.78	0.04	0	5	**0.014**	D.LVFPTEIVGKRVRYLVGGNKIQKVLL.D
793.18	3168.69	3168.65	0.04	0	40	**5.3e-006**	D.YKLESFQAVYNKLTGKQIVFEIPSETH.-
656.90	3935.34	3935.29	0.05	0	7	**0.011**	D.VAGGKKALAIFVPVPSLAGFHKVQTKLTRELEKKFQ.D

14. RPL23A

874.45	873.44	873.43	0.01	0	24	**0.00081**	**D.GVFLYFE.D**
692.40	1382.78	1382.77	0.02	0	48	**2.2e-006**	**D.LWPRVASNSGVVV.-**
714.66	2854.63	2854.59	0.04	0	38	**1.5e-005**	**D.NSGARNLYIIAVKGSGSRLNRLPAASLG.D**
743.42	2969.65	2969.62	0.04	1	34	**8.5e-005**	**D.NSGARNLYIIAVKGSGSRLNRLPAASLGD.M**

15. RPL25

460.28	918.54	918.53	0.01	0	35	**1.6e-005**	**D.IANRIGYI.-**
609.85	1217.69	1217.68	0.02	2	62	**3.3e-008**	**D.ALDIANRIGYI.-**
727.73	2180.16	2180.14	0.02	0	56	**1.1e-007**	**D.SYKVIEQPITSETAMKKVE.D**
574.81	2295.19	2295.17	0.03	1	14	**0.0041**	**D.SYKVIEQPITSETAMKKVED.G**
489.11	2440.49	2440.46	0.03	0	14	**0.0019**	**D.VLKVNTLVRPNGTKKAYVRLTA.D**

16. RPP2B

671.33	670.32	670.31	0.01	0	11	**0.019**	**D.MGFGLF.D**
786.36	785.35	785.34	0.01	1	12	**0.031**	**D.MGFGLFD.-**
786.36	785.35	785.34	0.01	1	15	**0.016**	**D.DMGFGLF.D**
600.85	1199.69	1199.68	0.01	0	51	**4e-007**	**D.IKAVVESVGAEV.D**
658.37	1314.72	1314.70	0.02	1	57	**1.8e-007**	**D.IKAVVESVGAEVD.E**
536.58	1606.71	1606.68	0.02	0	37	**9.6e-006**	**D.AAEEEKEEEAKEES.D**
708.06	2121.16	2121.13	0.03	0	36	**1.3e-005**	**-.MKYLAAYLLLVQGGNAAPSAA.D**

17. RPS7B

602.32	601.32	601.31	0.01	1	19	**0.0023**	**D.VQQID.Y**
409.23	816.45	816.43	0.01	2	34	**4.4e-005**	**D.SKDVQQI.D**
466.74	931.47	931.46	0.01	3	30	**0.0002**	**D.SKDVQQID.Y**
530.79	1059.56	1059.54	0.01	0	53	**2.6e-007**	**D.LESSSPELKA.D**
588.30	1174.59	1174.57	0.02	1	54	**3.8e-007**	**D.LESSSPELKAD.L**
489.32	1464.93	1464.91	0.02	0	26	**0.00012**	**D.LRPLQIKSIREI.D**
527.66	1579.96	1579.94	0.02	1	12	**0.0067**	**D.LRPLQIKSIREID.V**

18. RPS10A

558.29	1114.56	1114.54	0.02	0	37	**4.3e-005**	**D.FNQAKHEEI.D**
615.80	1229.58	1229.57	0.02	1	28	**0.00058**	**D.FNQAKHEEID.T**
540.07	2156.25	2156.22	0.03	0	37	**1.1e-005**	**D.RNKIHQYLFQEGVVVAKK.D**
455.26	2271.28	2271.25	0.03	1	25	**0.00031**	**D.RNKIHQYLFQEGVVVAKKD.F**

19. RPS14B

352.20	702.38	702.37	0.01	0	25	**0.00085**	**D.TFVHVT.D**
714.38	713.37	713.36	0.01	1	40	**1.9e-005**	**D.VTPVPSD.S**
409.71	817.41	817.40	0.01	1	40	**3.1e-005**	**D.TFVHVTD.L**
443.75	885.48	885.47	0.01	2	41	**4e-006**	**M.ANDLVQAR.D**
462.28	1845.07	1845.05	0.02	0	18	**0.0016**	**D.LSGKETIARVTGGMKVKA.D**
491.03	1960.10	1960.08	0.03	1	35	**6.2e-005**	**D.LSGKETIARVTGGMKVKAD.R**

20. RPL8B

805.40	804.40	804.45	-0.05	1	9	**0.026**	**D.KRAKTSD.S**
414.70	827.39	827.38	0.01	2	55	**3e-007**	**D.ANFADKY.D**
472.22	942.42	942.41	0.01	3	43	**9.1e-006**	**D.ANFADKYD.E**
558.34	1114.67	1114.66	0.02	0	49	**1.3e-006**	**D.EAALAKLVSTI.D**
615.86	1229.70	1229.69	0.02	1	35	**7e-005**	**D.EAALAKLVSTID.A**
699.41	3492.03	3491.99	0.04	1	9	**0.026**	**D.ASPKPYAVKYGLNHVVSLIENKKAKLVLIAND.V**

21. RPL8A

409.71	817.41	817.44	-0.03	1	8	**0.048**	**D.KRAKNSD.S**
414.70	827.39	827.38	0.01	2	55	**3e-007**	**D.ANFADKY.D**
472.22	942.42	942.41	0.01	3	43	**9.1e-006**	**D.ANFADKYD.E**
558.34	1114.67	1114.66	0.02	0	49	**1.3e-006**	**D.EAALAKLVSTI.D**
615.86	1229.70	1229.69	0.02	1	35	**7e-005**	**D.EAALAKLVSTID.A**

22. RPL2A

427.94	1280.78	1280.77	0.02	0	24	**0.00045**	**D.ARGVIGVIAGGGRV.D**
453.28	1356.82	1356.80	0.02	0	42	**7e-006**	**D.SGRGAPLAKVVFR.D**
466.28	1395.81	1395.79	0.02	1	24	**0.00079**	**D.ARGVIGVIAGGGRVD.K**
495.54	1978.11	1978.09	0.03	0	18	**0.0018**	**D.RGALARASGNYVIIIGHNP.D**
431.64	2153.18	2153.15	0.03	1	15	**0.0063**	**D.YAERHGYIRGIVKQIVHD.S**
575.68	3448.03	3447.99	0.04	0	12	**0.0068**	**M.GRVIRNQRKGAGSIFTSHTRLRQGAAKLRTL.D**
578.76	4044.29	4044.25	0.05	2	14	**0.0044**	**D.RGALARASGNYVIIIGHNPDENKTRVRLPSGAKKVISS.D**

23. RPL22A

513.29	1024.57	1024.56	0.01	0	20	**0.00047**	**D.PASYAKYLI.D**
518.76	1035.50	1035.49	0.01	0	45	**2.9e-006**	**D.VSSPTENGVF.D**
570.81	1139.60	1139.59	0.01	1	43	**4.8e-006**	**D.PASYAKYLID.H**
597.35	1789.04	1789.02	0.02	0	31	**4.4e-005**	**M.APNTSRKQKIAKTFTV.D**
635.70	1904.07	1904.05	0.02	1	25	**0.00015**	**M.APNTSRKQKIAKTFTVD.V**
731.40	2921.56	2921.52	0.04	2	15	**0.0018**	**M.APNTSRKQKIAKTFTVDVSSPTENGVF.D**
509.74	3561.13	3561.08	0.05	0	12	**0.0033**	**D.GTVVTVVSTAKFSGKYLKYLTKKYLKKNQLR.D**
526.17	3676.16	3676.11	0.05	1	6	**0.026**	**D.GTVVTVVSTAKFSGKYLKYLTKKYLKKNQLRD.W**

24. RPL3

599.85	2395.37	2395.34	0.03	0	44	**2.1e-006**	**D.RSKPVALTSFLGYKAGMTTIVR.D**
628.61	2510.40	2510.37	0.03	1	59	**1.1e-007**	**D.DRSKPVALTSFLGYKAGMTTIVR.D**

25. RPS16A

304.68	607.34	607.33	0.01	0	6	**0.039**	**D.KFSNI.D**
343.74	685.46	685.45	0.01	0	25	**0.00029**	**D.RTLLIA.D**
362.19	722.37	722.36	0.01	1	10	**0.021**	**D.KFSNID.I**
401.25	800.49	800.48	0.01	1	16	**0.0056**	**D.RTLLIAD.S**
545.51	2722.53	2722.50	0.03	0	7	**0.02**	**D.SRRPEPKKFGGKGARSRFQKSYR.-**
607.19	3637.10	3637.05	0.05	0	20	**0.00091**	**D.IRVRVTGGGHVSQVYAIRQAIAKGLVAYHQKYV.D**
626.36	3752.12	3752.07	0.05	1	27	**0.00043**	**D.IRVRVTGGGHVSQVYAIRQAIAKGLVAYHQKYVD.E**

26. RPS22A/22B

367.22	732.42	732.41	0.01	0	29	**0.00012**	**M.TRSSVLA.D**
424.73	847.45	847.44	0.01	1	29	**0.00012**	**M.TRSSVLAD.A**
532.30	2125.18	2125.15	0.03	0	5	**0.03**	**D.HEEARRKHVSGKILGFVY.-**

27. RPS17A

529.81	1057.60	1057.59	0.01	0	15	**0.0033**	**D.LSRSNGVLNV.D**
587.32	1172.63	1172.61	0.01	1	33	**0.00011**	**D.LSRSNGVLNVD.N**
534.59	2134.35	2134.32	0.03	0	40	**1e-005**	**D.LVKSLGLKLPLSVINVSAQR.D**
563.35	2249.38	2249.35	0.03	1	31	**0.00017**	**D.LVKSLGLKLPLSVINVSAQRD.R**

28. RPS20

735.45	2203.32	2203.29	0.03	0	24	**0.0002**	**D.LEAPVQIVKRITQITIEPGV.D**
773.79	2318.35	2318.32	0.03	1	39	**1.4e-005**	**D.LEAPVQIVKRITQITIEPGVD.V**

29. RPL30

631.41	630.40	630.40	0.01	0	28	**0.00043**	**D.ILTTLA.-**
417.24	832.47	832.45	0.01	2	49	**6.4e-007**	**D.SDILTTLA.-**

30. RPS15

596.82	1191.62	1191.61	0.01	0	52	**3.3e-007**	**D.LEKLLEMSTE.D**
654.33	1306.65	1306.63	0.02	1	66	**2.3e-008**	**D.LEKLLEMSTED.F**

31. RPL5

354.22	1059.64	1059.62	0.01	0	22	**0.0003**	**D.GPRPFKVFL.D**
411.22	1230.65	1230.64	0.02	0	15	**0.0017**	**D.IYTSAHEAIRA.D**
559.29	1674.84	1674.82	0.02	2	43	**2.5e-006**	**D.SLEDIYTSAHEAIRA.D**
572.62	1714.84	1714.82	0.02	0	13	**0.0024**	**D.GGLYVPHSENRFPGW.D**
501.80	2003.15	2003.13	0.03	0	9	**0.007**	**D.IGLQRTTTGARVFGALKGAS.D**

32. RPL4A/4B

538.57	2150.23	2150.21	0.03	0	30	**9.4e-005**	**D.LESIQKTKEAVAALKAVGAHS.D**
567.32	2265.26	2265.23	0.03	1	38	**4e-005**	**D.LESIQKTKEAVAALKAVGAHSD.L**

33. RPS2

741.46	740.45	740.44	0.01	0	22	**0.00069**	**D.TLLPGLQ.D**
514.94	1541.80	1541.78	0.02	2	30	**5.6e-005**	**D.IYSDEASAQKKRF.-**

34. RPS4A/4B

400.22	798.43	798.42	0.01	1	33	**0.0001**	**D.PNIKVND.T**
431.75	861.49	861.48	0.01	0	11	**0.0089**	**D.GRTIRYP.D**
613.82	1225.62	1225.61	0.02	0	29	**0.00012**	**D.ATNENFRLVY.D**
321.70	1282.78	1282.76	0.02	0	8	**0.027**	**D.VKGRFAVHRIT.D**
350.46	1397.81	1397.79	0.02	1	13	**0.02**	**D.VKGRFAVHRITD.E**

35. RPP1A

331.72	661.42	661.42	0.01	0	15	**0.0015**	**D.IFAKAL.D**
671.33	670.32	670.31	0.01	0	11	**0.019**	**D.MGFGLF.D**
389.23	776.45	776.44	0.01	1	22	**0.00063**	**D.IFAKALD.G**
786.36	785.35	785.34	0.01	1	12	**0.031**	**D.MGFGLFD.-**
786.36	785.35	785.34	0.01	1	15	**0.016**	**D.DMGFGLF.D**

36. RPL1A/1B

340.71	679.40	679.39	0.01	0	15	**0.0029**	**D.LYGKVT.D**
398.22	794.43	794.42	0.01	1	10	**0.018**	**D.DLYGKVT.D**
398.22	794.43	794.42	0.01	1	23	**0.00097**	**D.LYGKVTD.V**

37. RPL9A

338.73	675.45	675.44	0.01	0	16	**0.0023**	**D.IRKFL.D**
675.86	1349.71	1349.70	0.02	0	45	**4.8e-006**	**D.GIYVSHKGFITE.D**
450.93	1349.76	1349.75	0.02	0	11	**0.0076**	**D.GAKFIEVRNFLG.D**

38. RPL16B

542.30	1082.59	1082.58	0.01	0	62	**3e-008**	**D.VAKQLASFGY.-**

39. RPS29B

307.51	919.52	919.50	0.01	0	16	**0.0026**	**D.IGFHKYR.-**

40. RPL16A

517.31	1032.61	1032.60	0.01	0	47	**1.9e-006**	**D.VAKQLAALGY.-**

41. RPL6B

703.35	1404.68	1404.67	0.02	0	44	**2e-006**	**M.TAQQAPKWYPSE.D**

42. RPS26B

423.60	1267.78	1267.76	0.02	0	42	**3.1e-006**	**D.NKVSPAAAAKKAL.-**

43. RPL11A/11B

821.43	820.43	820.42	0.01	0	19	**0.0013**	**D.PSIGIFGM.D**
396.22	1185.63	1185.62	0.02	0	12	**0.0094**	**D.TVSWFKQKY.D**

44. RPS28A/28B

457.27	1825.05	1825.03	0.02	0	11	**0.0078**	**D.TSRTIVRNVKGPVREN.D**
486.03	1940.08	1940.06	0.02	1	16	**0.0063**	**D.TSRTIVRNVKGPVREND.I**
633.03	3792.15	3792.10	0.05	2	8	**0.015**	**D.TSRTIVRNVKGPVRENDILVLMESEREARRLR.-**

45. RPL26A/26B

374.99	1495.95	1495.93	0.02	0	16	**0.0026**	**D.RKALIQRKGGKLE.-**
511.49	3062.87	3062.83	0.04	0	9	**0.014**	**D.KVTKEKVNGASVPINLHPSKLVITKLHL.D**

46. RPL13B

548.08	2188.27	2188.25	0.03	0	23	**0.00028**	**D.ANVQRLKEYQSKIIVFPR.D**

47. RPS29A

441.26	880.50	880.49	0.01	0	29	**5.6e-005**	**D.IGFNKFR.-**

48. RPL20A/20B

385.24	1536.93	1536.91	0.02	0	11	**0.0074**	**D.VKRQYVKQFLTK.D**
414.00	1651.96	1651.94	0.02	1	20	**0.0019**	**D.VKRQYVKQFLTKD.L**

49. RPL10

426.50	1701.97	1701.95	0.02	0	18	**0.00072**	**D.RPEYLKKREAGEVK.D**

50. RPS18A/18B

386.96	1543.83	1543.81	0.02	0	25	**0.00033**	**D.YHTLANNVESKLR.D**

51. RPS8A/8B								
	474.88	2369.34	2369.31	0.03	0	13	**0.0053**	D.GYILEGEELAFYLRRLTAKK.-
52. RPS1A								
	570.33	2846.60	2846.56	0.03	0	21	**0.00042**	D.IKAPSTFENRNVGKTLVNKSTGLKSAS.D
53. RPL19A								
	425.94	2549.57	2549.54	0.03	0	11	**0.008**	M.ANLRTQKRLAASVVGVGKRKVWL.D
54. RPL6A								
	300.85	899.52	899.51	0.01	0	13	**0.0026**	D.KPHMLKF.-
55. RPS9A/9B								
	457.75	2740.48	2740.44	0.03	0	13	**0.0055**	M.PRAPRTYSKTYSTPKRPYESSRL.D
56. RPS27A/27B								
	517.45	3615.09	3615.04	0.05	2	12	**0.0064**	M.VLVQDLLHPTAASEARKHKLKTLVQGPRSYFL.D
57. RPL14A/14B								
	422.27	842.52	842.52	-0.01	1	10	**0.022**	I.DQKKVLI.D
	490.63	2937.76	2937.72	0.04	0	5	**0.03**	D.FERFQVMVLRKQKRYTVKKALAKA.-
58. RPS24A								
	525.73	2623.60	2623.56	0.03	1	6	**0.048**	D.AVTIRTRKVISNPLLARKQFVVD.V

Table 6.1: Table of identified peptides from the *Saccharomyces cerevisiae* ribosome.

In addition to LC ESI MS/MS analysis, ribosomal protein digestion products were analyzed using LC MALDI combined with collisional activation. MALDI can provide an attractive alternative to ESI as an ionization method because it produces primary singly charged ions. This proves to be advantageous when analyzing high mass peptides that carry a high number of charges when produced by ESI. High charge state peptides traditionally produce fragmentation spectra that can complicate database searching. In this analysis, we targeted singly charged peptides greater than 3 kDa. following automated sample preparation. The fragmentation spectra of m/z = 3172.3 is displayed in **Figure 6.6**. The MASCOT search results from the identification of this peptide identification and others at m/z = 3852.1, 3967.4, 3345.9, and 3164.1 are displayed in **Table 6.2**.

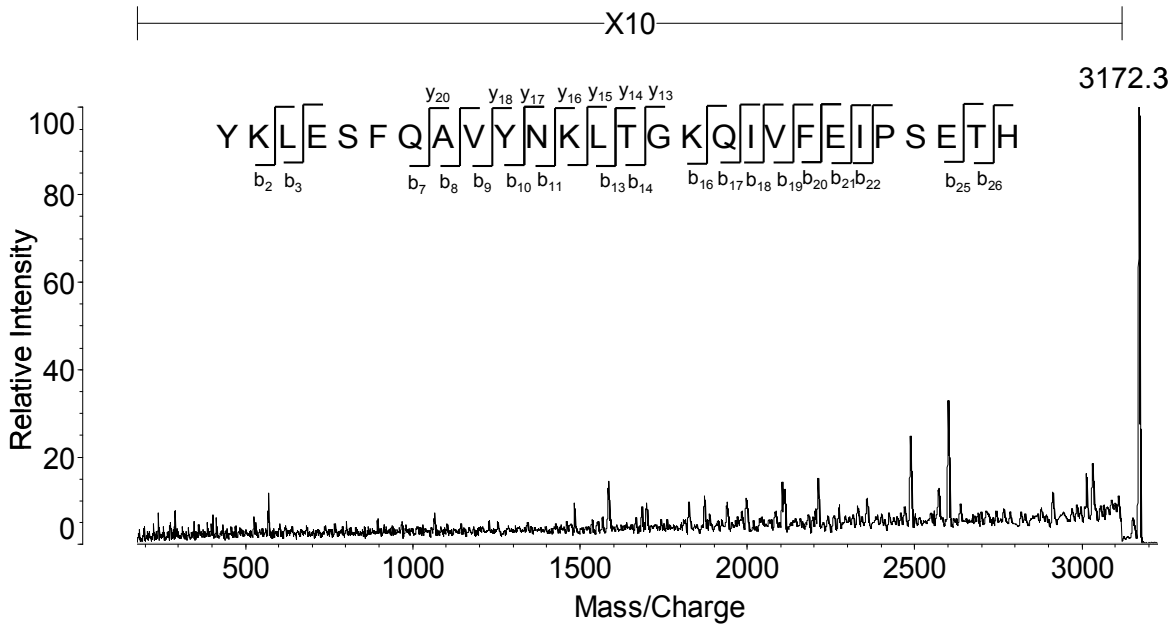


Figure 6.6: MALDI TOF-TOF fragmentation spectrum of m/z = 3172.3.

Precursor M/Z	Peptide Score	Peptide E-value	Peptide Sequence	Protein Source	Protein Score
3172.3	126	6.80E-14	D.YKLESFQAVYNKLTGKQIVFEIPSETH.-	RPS7A	121
3852.1	81	2.60E-09	D.PNEVKYLYLRAVGGEVGASAALAPKIGPLGLSPKKVGE.D	RPL12A/B	81
3967.4	16	0.0064	D.PNEVKYLYLRAVGGEVGASAALAPKIGPLGLSPKKVGED.I	RPL12A/B	11
3345.4	49	2.80E-06	D.AQLLLAANTHLGARNVQVHQEPYVFNARPD.G	RPS0A/B	45
3164.1	24	0.00065	D.GAFVKFLSKKGSLENNIREFPEYFAAQA.-	RPL10	24

Table 6.2: MASCOT search results from LC MALDI TOF-TOF analysis of high mass peptides (>3 kDa.)

Chapter 7: Relative Quantitation of Proteins

Introduction

This chapter discusses the integration of microwave-accelerated acid hydrolysis into quantitative proteomics. Much success in the field of comparative proteomics has evolved out of the development of proteolytic ^{18}O labeling strategies for protein quantitation. However, current ^{18}O labeling protocols can be labor-intensive and involve lengthy digestion with enzymes. For example, proteins can be digested with trypsin (18 hr.) in ^{16}O water followed by a 10 hr incubation in the presence of ^{18}O and immobilized trypsin.[4] Under these conditions, two ^{18}O atoms are fully incorporated on the carboxy- termini of the tryptic peptides. Incorporation of two ^{18}O atoms provides an increase in mass of 4 Da. and thus affords isotope ratio measurements of two identical peptides from two separate peptide pools, unlabeled and labeled with ^{18}O.

In this study, the feasibility of rapidly labeling Asp-specific peptides via microwave-accelerated acid hydrolysis was assessed. The final step in the proposed reaction mechanism for the Asp-X and X-Asp cleavages of proteins shown in **Chapter 1** (**Figure 1.7**), involves the potential introduction of a single oxygen atom when the reaction is carried in H_2O. Based on this chemistry, we have demonstrated incorporation of an ^{18}O atom into the C-terminus of Asp-X digestion products when microwave-assisted acid hydrolysis is carried out in highly enriched $H_2{}^{18}O$. As part of this study, the effects of several experimental parameters, such as acid type and concentration and digestion time, on ^{18}O labeling efficiency were assessed. In addition, we confirmed the position of the

^{18}O label using laser induced dissociation. ^{18}O labeling was first optimized on the model proteins, myoglobin and ovalbumin, and then extended to the Bacillus spore SASPs and the fiber protein from human adenovirus Type 5.

Materials and Methods

Myoglobin and ovalbumin were prepared at a concentration of 1 mg/ml in deionized $H_2{}^{16}O$. For samples that were digested in $H_2{}^{18}O$, a 100 µl aliquot was dried to remove all $H_2{}^{16}O$ and resuspended in 100 µl of commercial heavy water, > 95% enriched. Samples digested in $H_2{}^{16}O$ were taken directly from the stock suspension. For digestion, an aliquot of the protein solutions were acidified to 12.5% or 6 % (final concentration) with glacial acetic acid or 96% formic acid. Digestions in $H_2{}^{16}O$ and $H_2{}^{18}O$ were carried out in parallel using a CEM Benchmate laboratory microwave system equipped with 45 ml digestion vessel and a fiber optic temperature probe. All digestions were carried out at a constant temperature of 140 ± 5 ˚C for 90 sec, 3 min, 5 min, 15 min or 30 min. For the relative quantitaion studies, labeled and unlabeled samples were combined 1:1 and 2:1 (v:v, labeled: unlabeled)

Bacillus Spores

Bacillus cereus (ATCC E33L) spore suspensions were prepared at 2 mg/ml in deionized $H_2{}^{16}O$, whereas *Bacillus anthracis* Sterne suspensions were prepared in $H_2{}^{18}O$. *Bacillus cereus* and *Bacillus anthracis* Sterne spores were grown and harvested in house using methods previously described.[72] Spore suspensions were acidified with glacial acetic acid and the digestions were carried out in parallel for 90 sec at 120 ˚C. As with the digestion of pure protein standards, twenty seconds were added to the digestion time to allow temperatures to exceed 108˚C. For the relative quantitaion studies, labeled and unlabeled samples were combined 1:1 and 2:1 (v:v, labeled *B.anthracis* Sterne: unlabeled *B.cereus*).

Human Adenovirus Type 5

Human Adenovirus Type 5 samples were grown in HeLa cells and cultured 24 and 48 hours after infection. The infected HeLa cells were collected and washed twice with 0.01 M PBS. The washed cell pellet was then resuspended in 10 mM Tri-HCl buffer prior to digestion. A 100 µl aliquot of the 48 hr culture was dried to remove all $H_2{}^{16}O$ and resuspended in 100 µl of commercial heavy water, > 95% enriched. The 24 hr adenovirus samples, digested in $H_2{}^{16}O$, were taken directly from the virus stock suspension. For digestion, an aliquot of each virus sample was acidified to 12.5% with glacial acetic acid and heated to 140 ˚C for 90 sec. Once again, twenty seconds were added to the digestion time to allow temperatures to exceed 108˚C. For relative quantitation of the adenovirus fiber

protein, the peptide pools from the 24 and 48 hr adenovirus samples were combined 1:1 (v:v). Ziptip (C_{18}) was used as a quick cleanup technique prior to analysis of the 1:1 peptide mixture.

Results and Discussion

The effects of acid type and concentration in addition to reaction time on ^{18}O labeling were addressed in this study. After extensive optimization of experimental conditions, using myoglobin and ovalbumin as model proteins, 12.5% acetic acid was shown to provide the most efficient and reproducible labeling of Asp-specific peptides. **Figure 7.1** shows the results of incorporation of an ^{18}O label when ovalbumin was subjected to parallel chemical digestions for 90 sec in $H_2{}^{16}O$ and in $H_2{}^{18}O$ acidified with acetic acid. A single ^{18}O atom was found to be incorporated following a 90 second digestion in $H_2{}^{18}O$. Moreover, no evidence of non-specific labeling was observed. The theoretical isotope cluster for the peptide, STRTQINKVVRFD, from ovalbumin (**Figure 7.2**) correlates well with the measured isotope clusters shown in Figure 7.1 A and B. These spectra indicate that >90 % of the peptides carry an ^{18}O atom, consistent with high incorporation from the commercial heavy water, > 95% enriched.

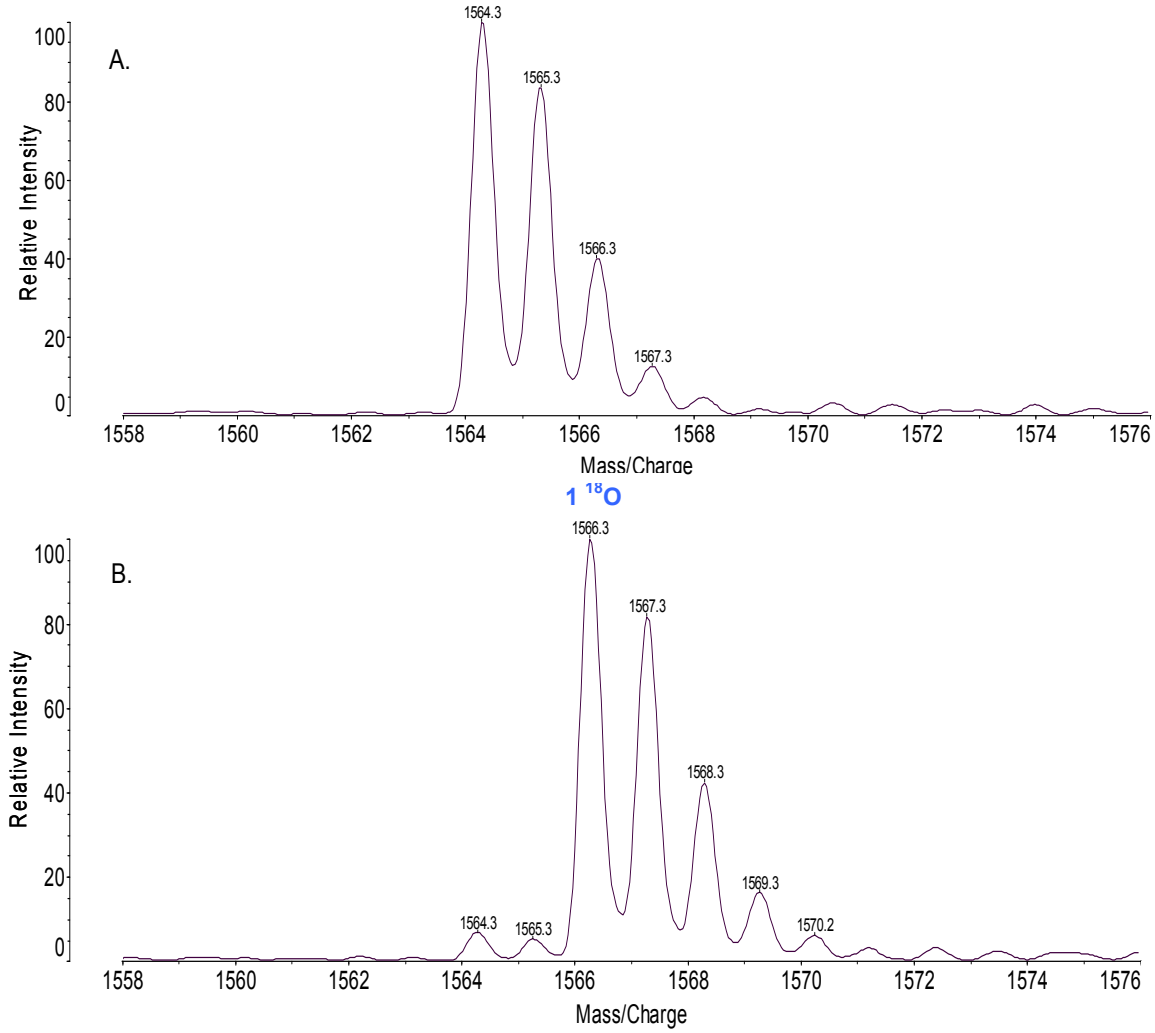
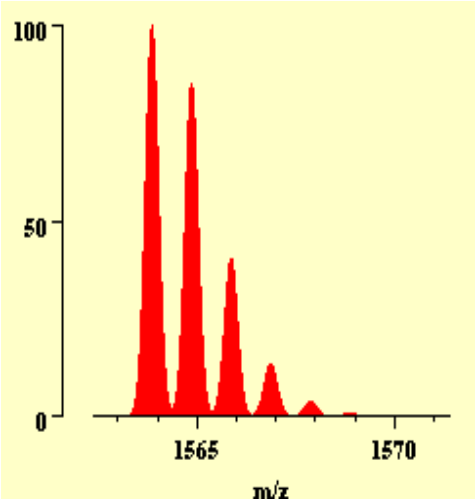


Figure 7.1: Mass spectrum of a the peptide, STRTQINKVVRFD, $[M+H]^{+}_{calc.}$ = 1563.9 after digestion with 12.5% acetic acid in $H_2{}^{16}O$ (A) and $H_2{}^{18}O$ (B) for 90 sec.



Isotope Number	m/z	Percent Total	Percent Maximum
0	1563.86071	40.88	100.00
1	1564.86354	34.99	85.60
2	1565.86621	16.55	40.49
3	1566.86880	5.62	13.75
4	1567.87133	1.52	3.71
5	1568.87388	0.34	0.83
6	1569.87634	0.06	0.15
7	1570.87938	0.01	0.02

Figure 7.2: Theoretical isotope cluster for the peptide, STRTQINKVVRFD, $[M+H]^+_{calc.}$ = 1563.9 from ovalbumin.

In order to confirm that the ^{18}O atom is incorporated on the carboxyl-terminal ends of peptides, as we hypothesized, tandem mass spectrometry experiments were performed. **Figure 7.3** is a fragment ion mass spectrum of the ^{18}O-peptide from myoglobin at m/z = 1667.1. The fragment ion spectrum of m/z = 1667.1 confirms that the ^{18}O atom was incorporated at the carboxy- terminus, based on the 2 Da. mass shift of y- ions, but no increase in mass of b- ions.

Incorporation of the ^{18}O into the carboxyl terminus of chemical digest products was also confirmed through the observation that the C-terminal peptide remained unlabeled after digestion. **Figure 7.4** is mass spectrum of the terminal peptide from myoglobin. The theoretical isotope cluster for this peptide is shown in **Figure 7.5**.

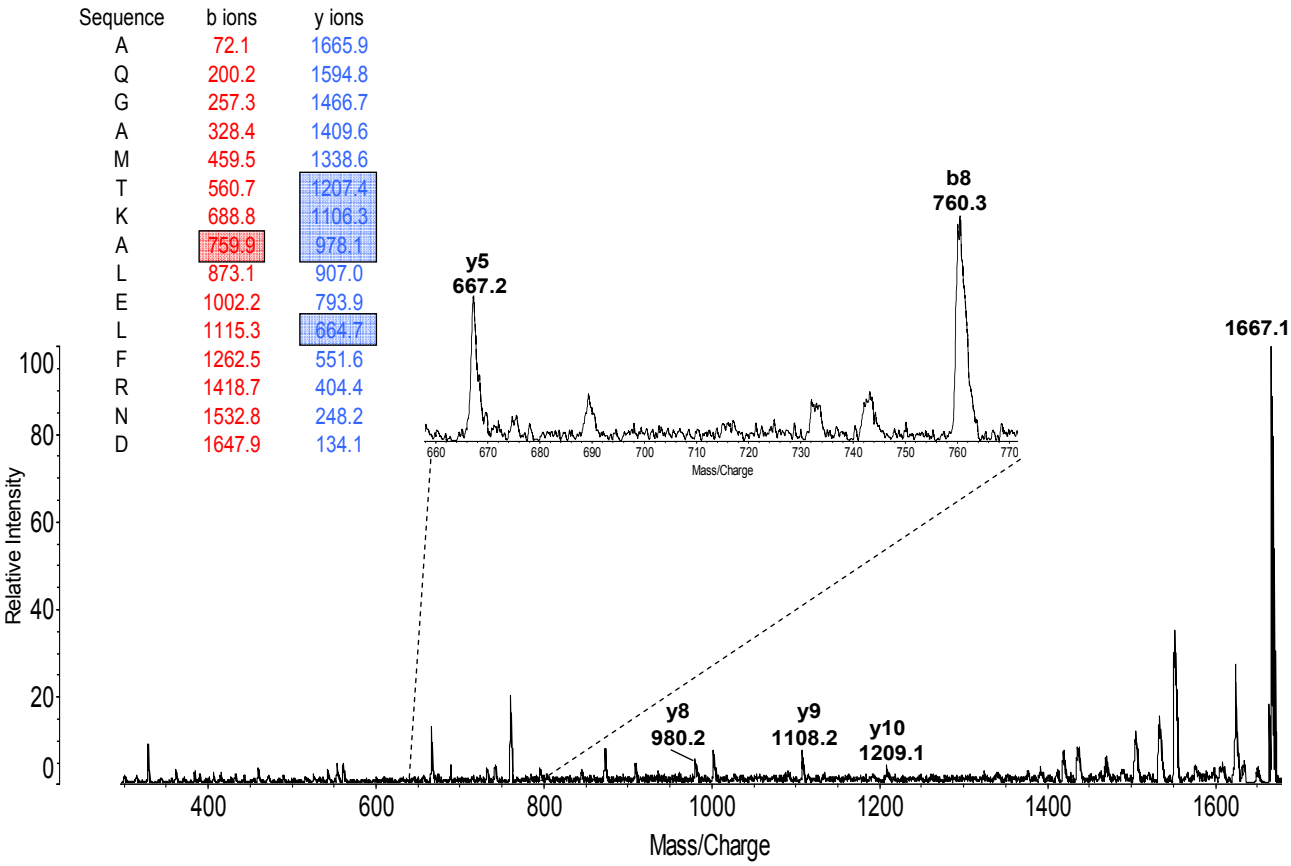


Figure 7.3: LID Mass spectrum of a the peptide, AQGAMTKALELFRND, $[M+H]^+_{calc.}$ = 1664.8 following 5 min digestion with 25% acetic acid. Inset is a list of all predicted fragment ion masses. Blue boxes are used to emphasize mass 2 Da. mass shifts in y-ions. Red box is used to show no mass shift in b-ions.

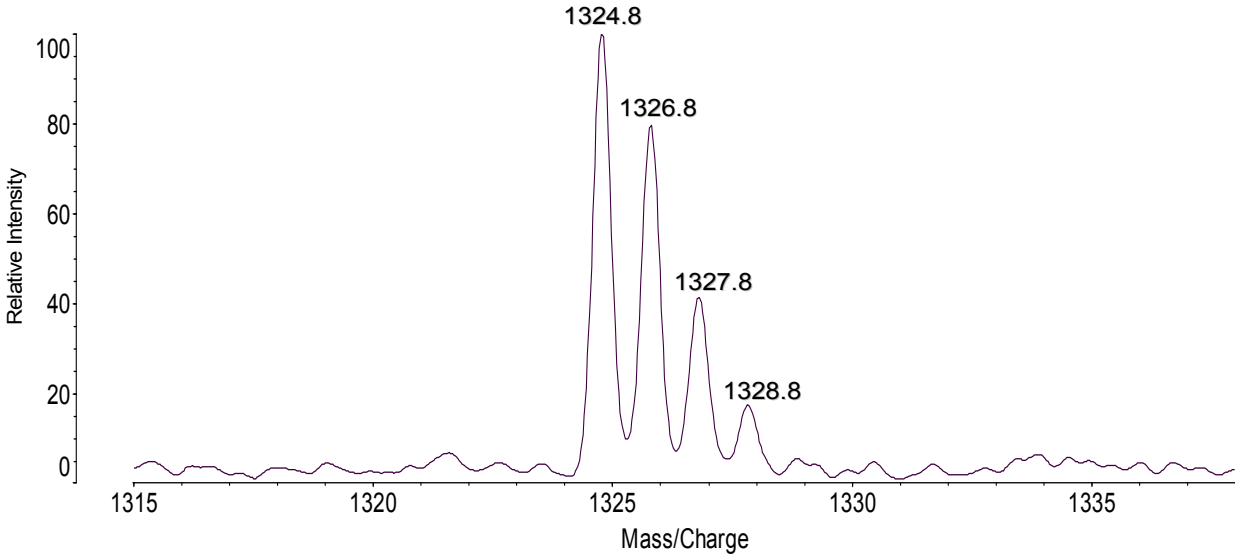
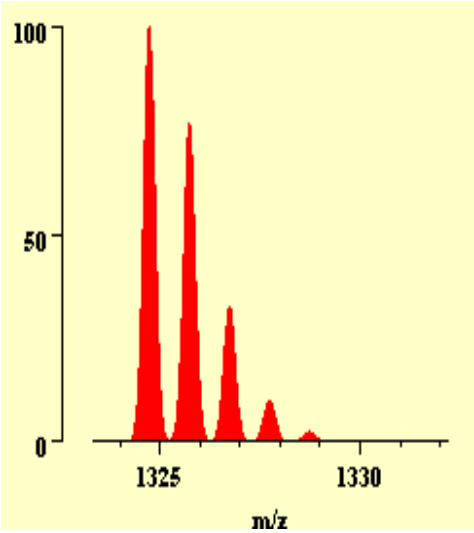


Figure 7.4: Mass spectrum of the C-terminal myoglobin peptide, IAAKYKELGFQG, $[M+H]^{+}_{calc.}$ = 1324.8



Isotope Number	m/z	Percent Total	Percent Maximum
0	1324.72651	44.89	100.00
1	1325.72948	34.58	77.02
2	1326.73226	14.69	32.73
3	1327.73494	4.48	9.98
4	1328.73758	1.08	2.41
5	1329.74018	0.22	0.48
6	1330.74306	0.03	0.08
7	1331.74647	0.00	0.01

Figure 7.5: Theoretical isotope cluster for the peptide, IAAKYKELGFQG, $[M+H]^+_{calc.}$ = 1324.8 from myoglobin.

As mentioned in the discussion of Asp-specific microwave-accelerated acid hydrolysis, four potential peptide products are possible based on the two pathways proposed in the reaction mechanism (**Figure 1.7**). However, based on our analysis of a variety of protein samples, we have found that peptide products that retain the Asp at the C-terminus are preferentially formed under short digestion times. Thus far, efficient incorporation of an ^{18}O label into peptides that retain the Asp at the C-terminus has been shown using this labeling strategy.

To determine the labeling pattern for those peptides that lose this C-terminal Asp, a peptide pair from ovalbumin was monitored. **Figure 7.6 A-C** shows mass spectra of a peptide pair from ovalbumin generated from 90 sec, 3 min, and 15 min incubations in 12.5 % acetic acid in the laboratory microwave device. Following a 90 sec digestion, the two peptides in the pair are found not to be labeled with the same efficiency. The peptide carrying an Asp residue, m/z = 1183.6, shows full incorporation of one ^{18}O atom. However, its peptide partner at m/z = 1068.7 shows only partial incorporation of a single ^{18}O atom. The ratio of the labeled to unlabeled peptides in the latter case is approximately 1 to 1, suggesting a different mechanism.

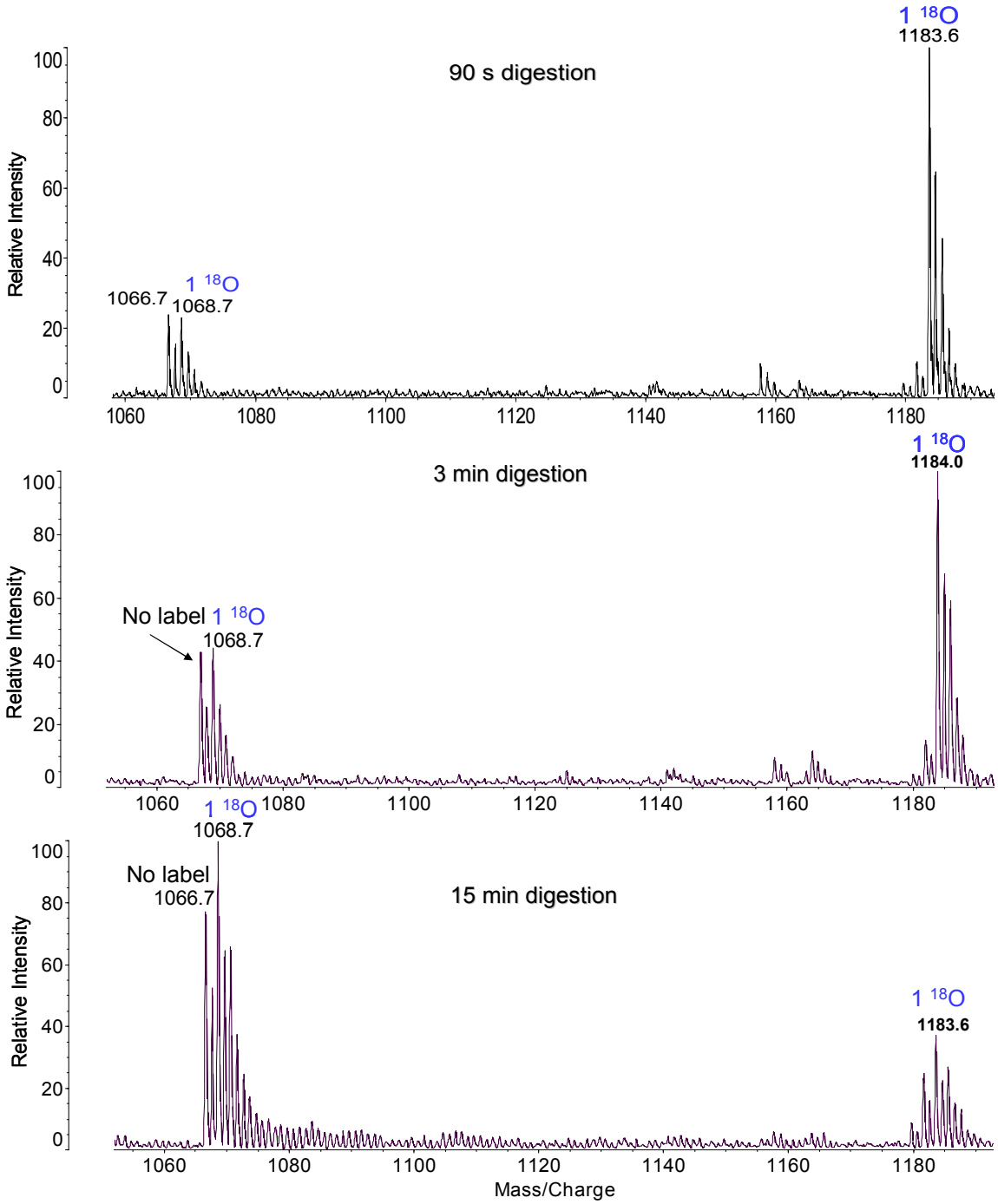


Figure 7.6: Mass spectrum of a peptide pair from ovalbumin generated from a 90 sec (A), 3min (B), and 5 min (C) digest with 12.5% acetic acid.

Attempts to induce full incorporation of a single ^{18}O atom by increasing the digestion times to 3 min and 15 min were unsuccessful. The results of the 3 min digestion were unchanged from those seen after 90 sec digestion. Increasing digestion time to 15 min did not improve labeling efficiency of the peptide without any Asp residue. Furthermore, at longer times, more of the Asp-peptide is converted to the Asp-free peptide with introduction of ^{16}O. This favors the use of short reaction time. Further reduction of digestion times (15 or 30 sec) was not advantageous because it resulted a significant decrease in peptide ion intensities.

The utility of this chemical digestion ^{18}O labeling strategy as a method for relative protein quantitation was evaluated by combining the 90 sec. ovalbumin digestion products in $H_2{}^{16}O$ and $H_2{}^{18}O$ in 1:1 and 2:1 ratios (labeled:unlabeled). **Figure 7.7 a-b** show the labeled and unlabeled molecular ion isotope clusters resulting from the 1:1 and 2:1 mixtures. Upon comparison of the intensities of the labeled and unlabeled peptides, these results correlate rather well with theoretical isotope clusters for 1:1 and 2:1 mixtures. Computer algorithms are available to deconvolute the contribution of the second ^{13}C isotope and other isotopes from the contribution of the ^{18}O label.

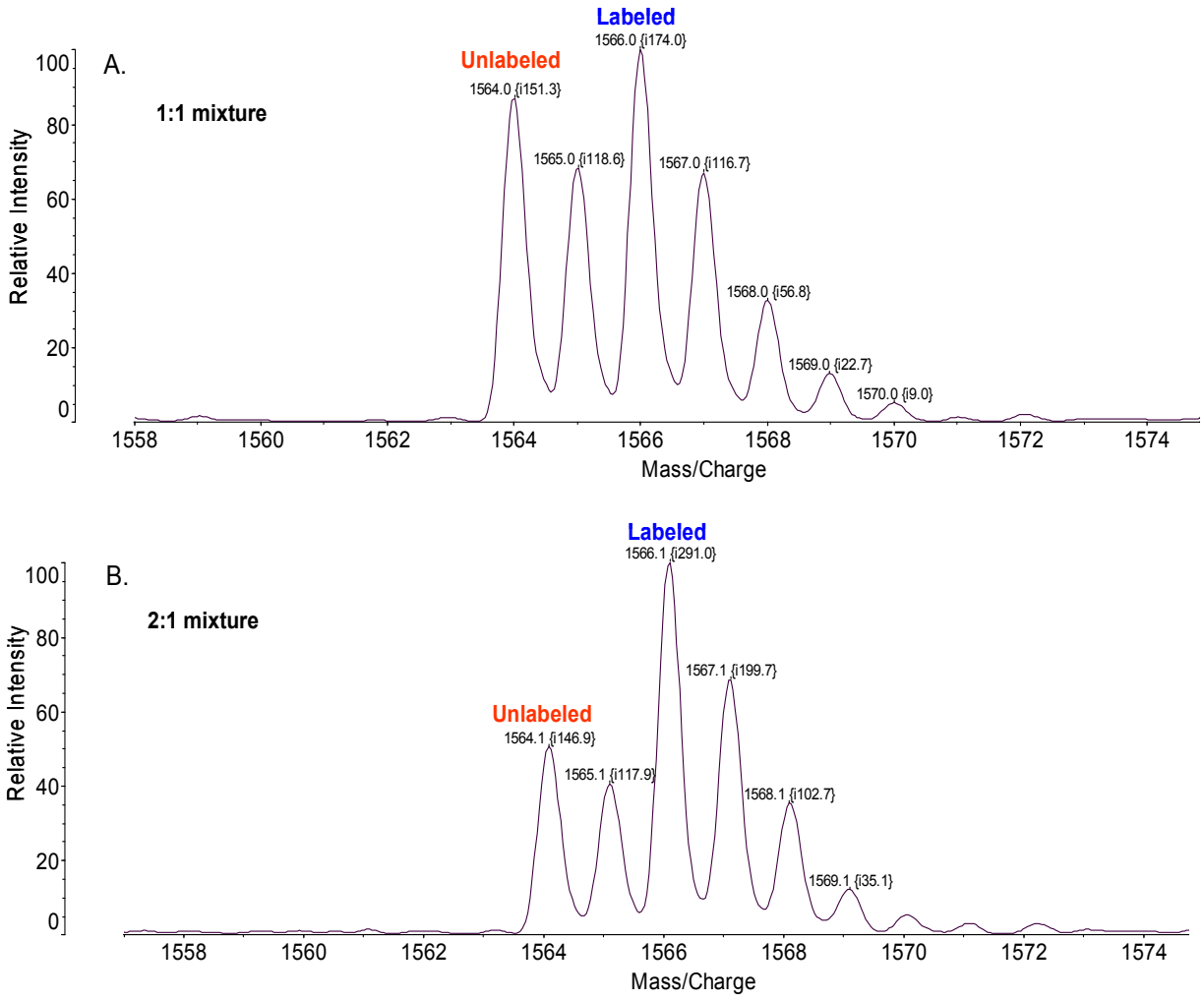


Figure 7.7: Mass spectrum of a the peptide, STRTQINKVVRFD, $[M+H]^+_{calc.}$ = 1563.9 from a 1:1 (A) and 2:1 (B) mixtures (labeled:unlabeled). Ion Intensities are included in brackets.

APPLICATION 1: Bacillus Spores

The objective of this study was to demonstrate the use of acid catalyzed ^{18}O labeling on a more complex biological sample. The small acid soluble proteins (SASPs) from bacillus spores within the *B.cereus* group are known to have highly conserved amino acid sequences. Following chemical digestion of either *B.cereus* (ATCC E33L) or *B.anthracis* Sterne, the peptide, ARSTNKLAVPGAESALD (m/z = 1699.9 Da.) is observed in both mass spectra. In this study, we labeled the peptides generated from chemical cleavage of the SASPs from *B.anthracis* Sterne using12.5% acetic acid in >95 % $H_2{}^{18}O$. Following, we combined this pool of peptides with a second pool of peptides generated by the chemical cleavage of *B.cereus* in $H_2{}^{16}O$. **Figure 7.8** is a schematic of the experimental procedure carried out for this analysis. **Figure 7.9 a** shows the molecular ion isotope clusters of the peptide, ARSTNKLAVPGAESALD, generated by chemical cleavage of the SASP B protein from *B.cereus* with 12.5% acetic acid in $H_2{}^{16}O$. Figure 7.9 b shows this same peptide from *B.anthracis* Sterne shifted by 2 Da. which is indicative of the incorporation of an ^{18}O atom.

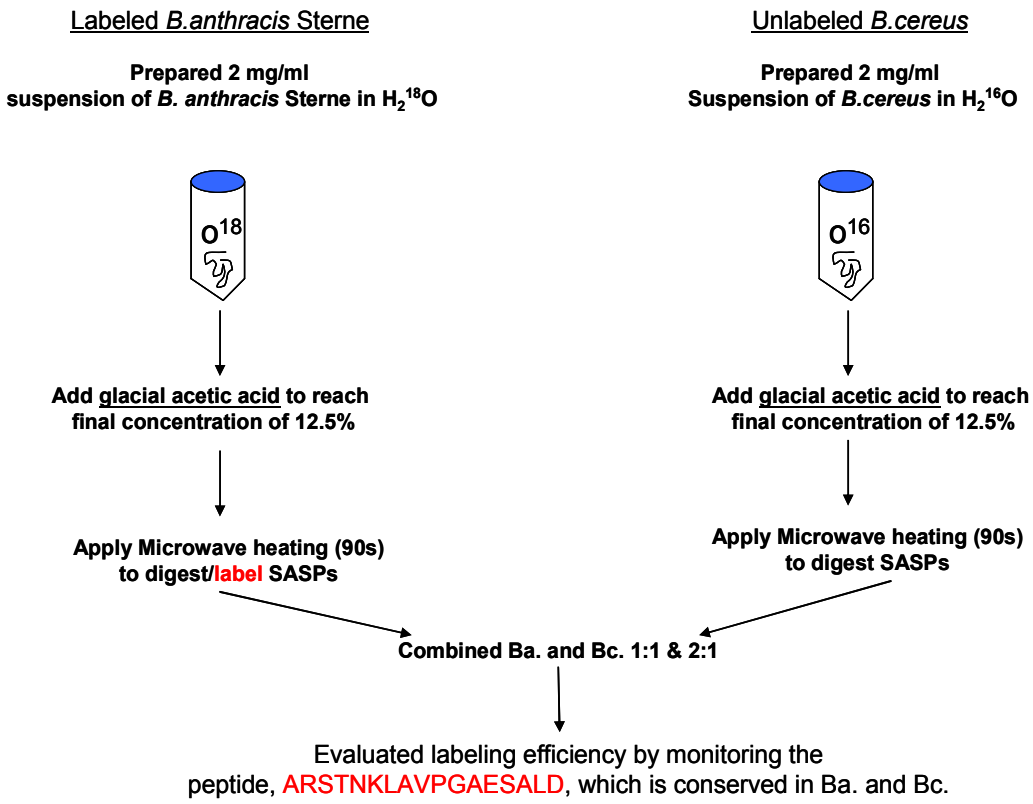


Figure 7.8: Schematic representation of experimental protocol for the chemical digestion and ^{18}O labeling of Bacillus spores.

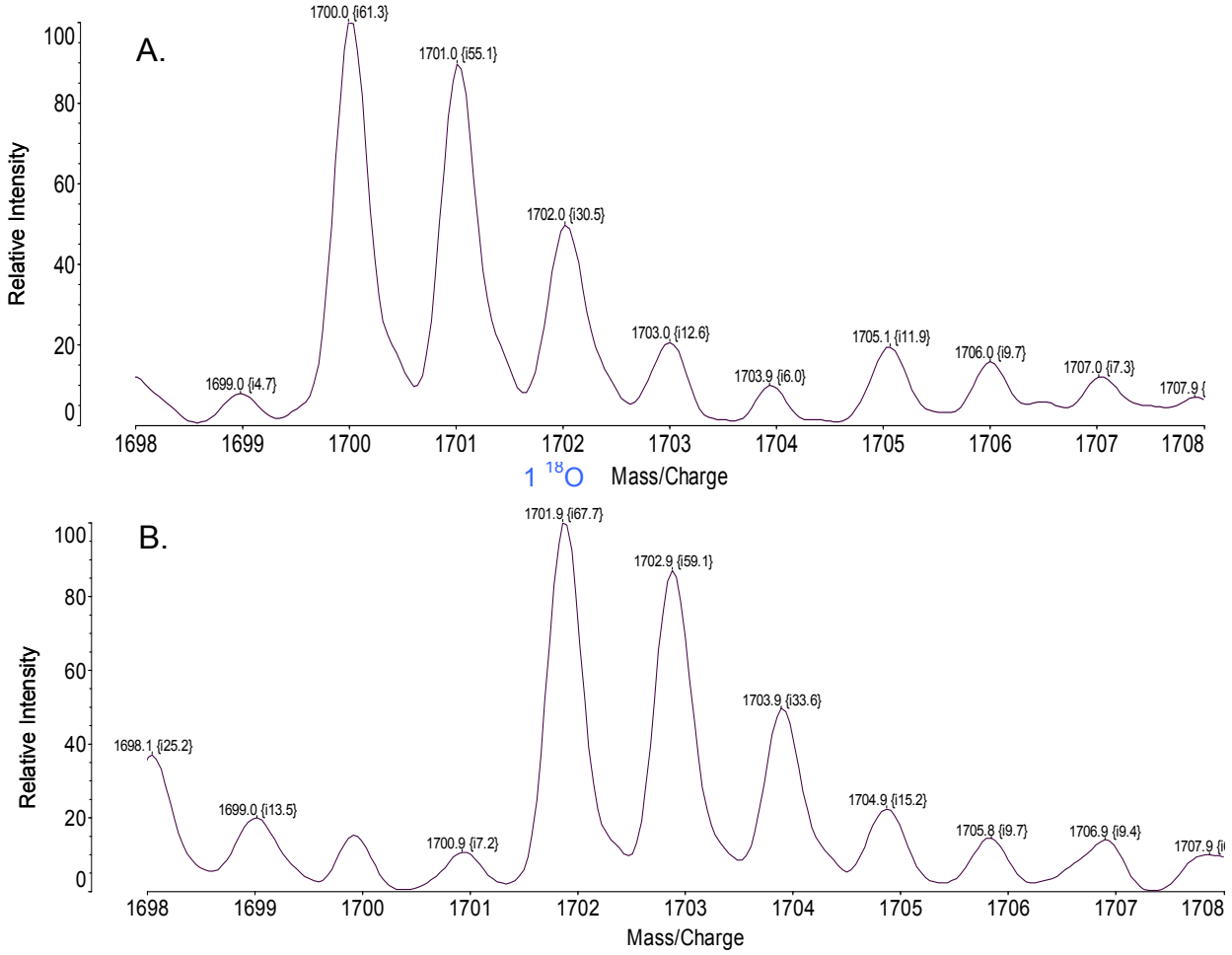


Figure 7.9: Mass spectrum of the peptide, ARSTNKLAVPGAESALD, $[M+H]^+_{calc.}$ = 1699.9, from SASP B from *B.cereus* following digestion in $H_2{}^{16}O$ (A). Mass spectrum of the same peptide from *B.anthracis* Sterne (B) following digestion in $H_2{}^{18}O$. Ion intensities are included in brackets

After demonstrating the feasibility of labeling SASP peptides generated by chemical cleavage in $H_2{}^{18}O$, the peptide pools from *B.anthracis* Sterne and *B.cereus* were combined 1:1 and 2:1. **Figures 7.10 a-b** are representative spectra of the 1:1 and 2:1 mixtures of labeled *B.anthracis* Sterne and unlabeled *B.cereus*, respectively. The relative intensities of the unlabeled and labeled peptides in **Figures 7.10 a-b** correlate well with 1:1 and 2:1 mixtures, respectively. Suitable computer algorithms are available to deconvolute the contribution of the second ^{13}C isotope and other isotopes from the contribution of the ^{18}O label. This experiment indicates the feasibility of using labeled *B. cereus* as an internal standard to quantitate *B. anthracis* in a sample.

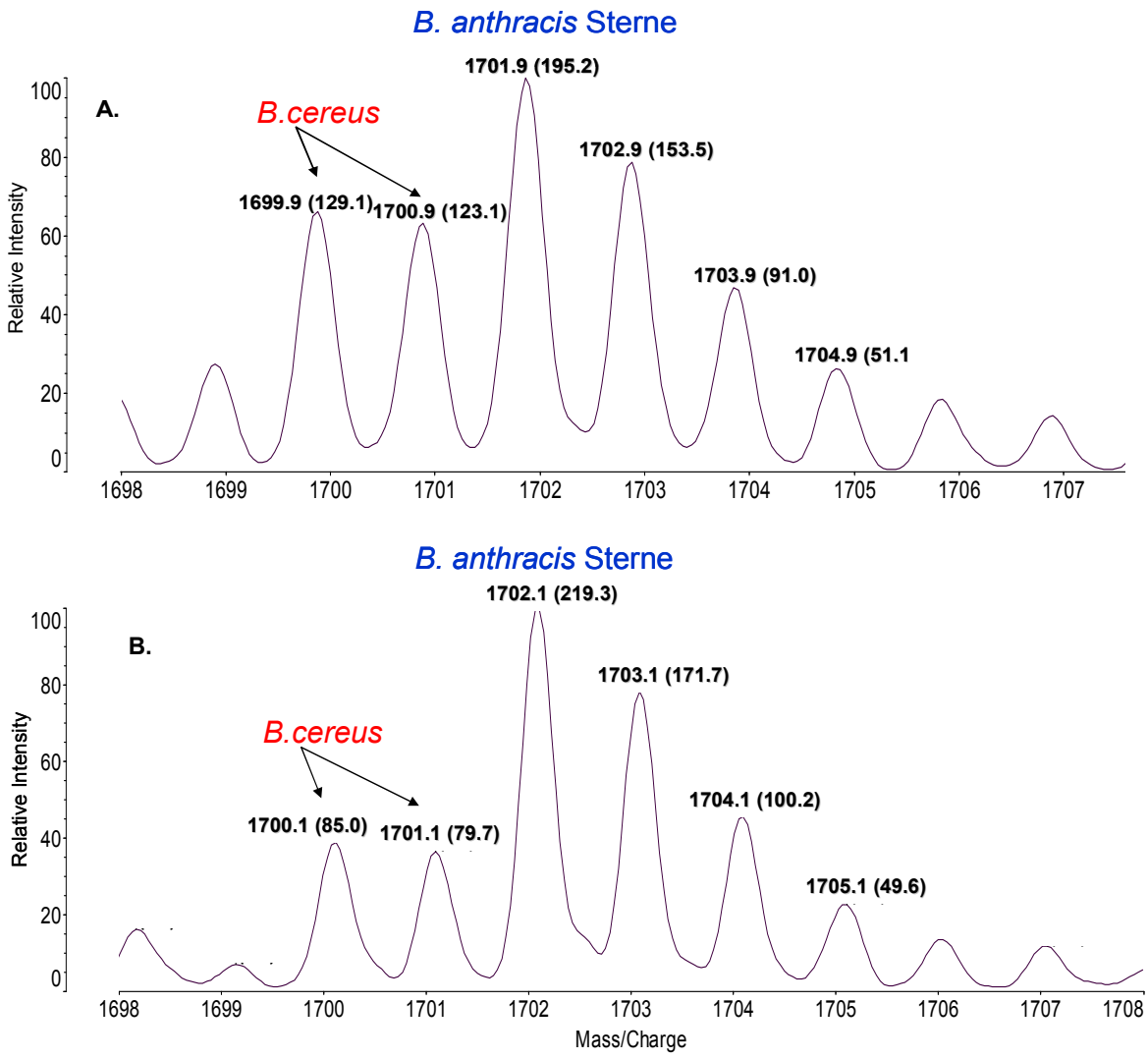


Figure 7.10: Mass spectrum of a SAPS-B peptide pair, from *B.cereus* and *B.anthracis* Sterne mixed 1:1(A) and 2:1 (B) (labeled: unlabeled). Ion intensities are included in parentheses.

Application 2: Qantitation of Adenovirus Fiber Protein

Chemical digestion ^{18}O labeling was also applied to two separate intact human adenovirus samples to quantify the production of fiber protein 24 hr. and 48 hr. post infection. The adenovirus sample, harvested 48 hr. post-infection, was digested using 12.5% acetic acid in $H_2{}^{18}O$, whereas the 24 hour culture was digested in $H_2{}^{16}O$. Figure 7.11 a shows the spectrum of the unlabeled adenovirus peptide, PEYWNFRNGD, from the fiber protein generated by 90 sec. digestion of the 24 hour culture of adenovirus. Figure 7.11 b is a mass spectrum of the same peptide, shifted by 2 Da., from a 48 hour culture of adenovirus labeled with ^{18}O.

The peptide pools from the 24 hr and 48 hr adenovirus cultures were combined 1:1 by volume and the relative intensities of the labeled and unlabeled peptide pairs were measured. Figure 7.12 is the mass spectrum of a fiber protein peptide pair from a 1:1 mixture of labeled and unlabeled adenovirus. The relative intensities of this peptide pair indicate that the fiber protein is present in an approximate two fold excess in the 48 hr culture compared to the 24 hr. culture. Figure 7.13 displays the theoretical isotopic distribution of a 1:2 mixture of the unlabeled:labeled peptides.

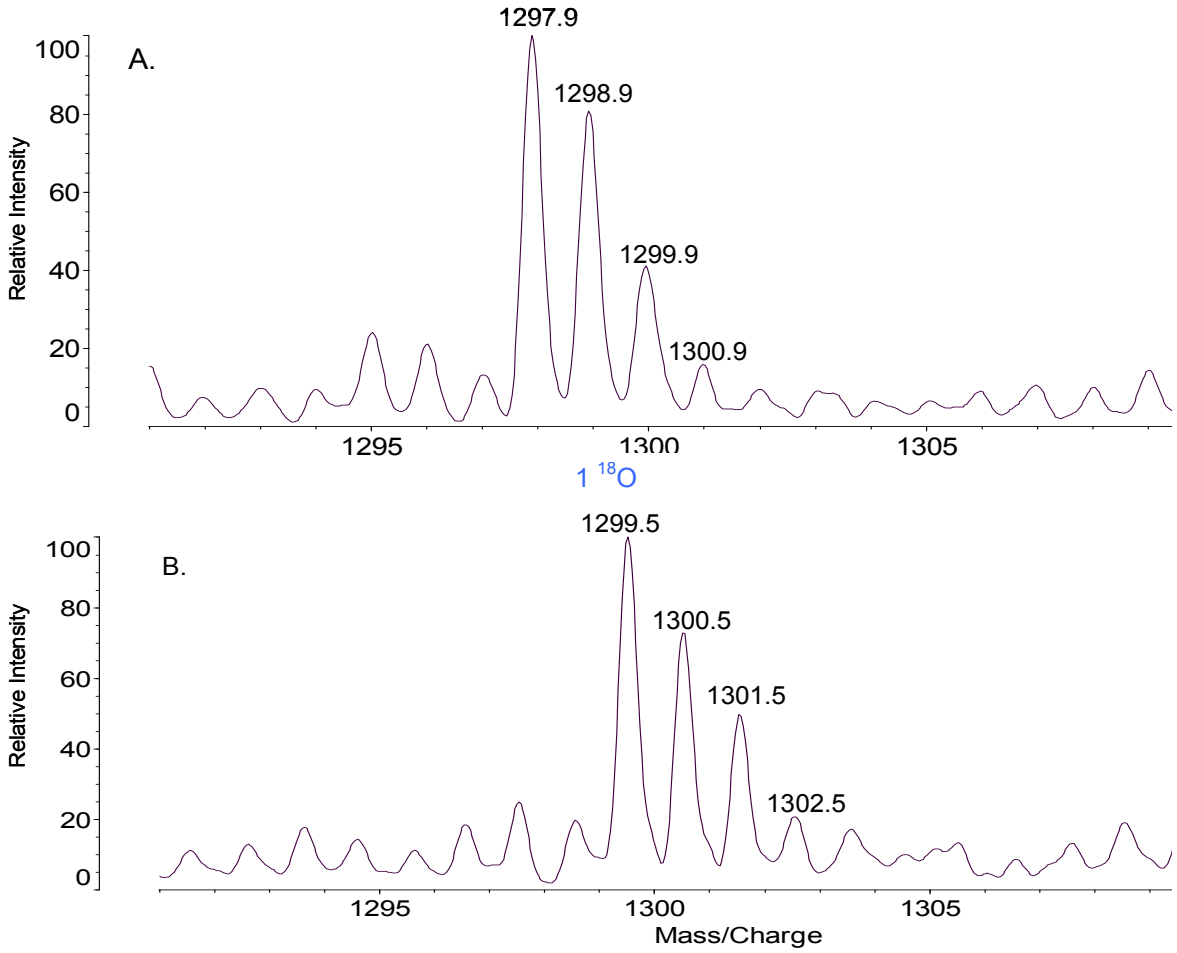


Figure 7.11: Mass spectrum of the peptide, PEYWNFRNGD, $[M+H]^+_{calc.}$ = 1297.6, from the chemical digestion of the fiber protein from a 24 hour culture of adenovirus (A) from a 48 hour culture of adenovirus (B).

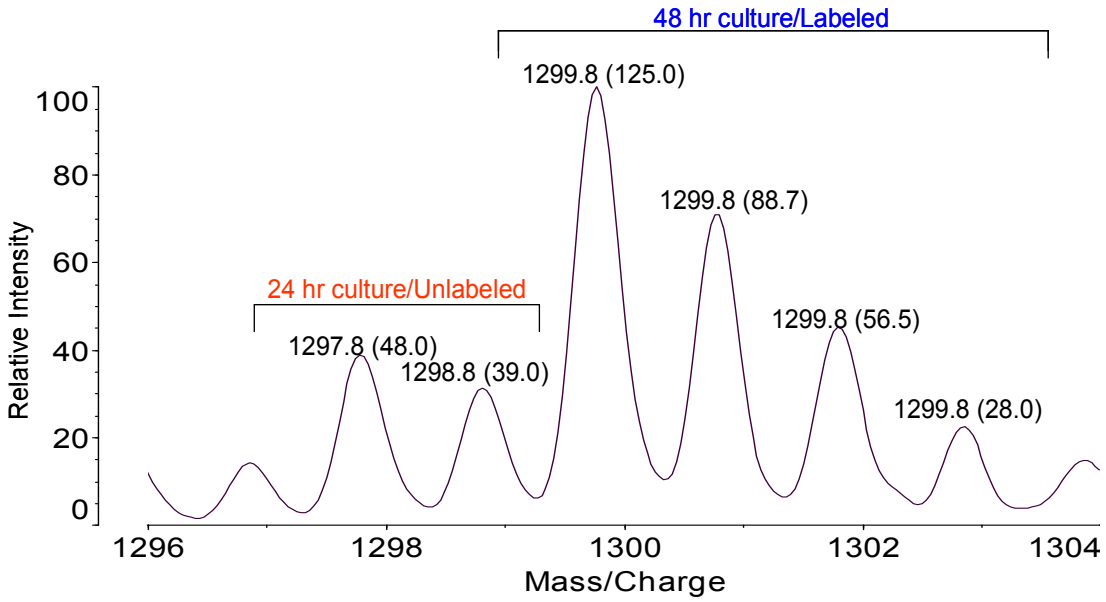


Figure 7.12: Mass spectrum of the peptide, PEYWNFRNGD, $[M+H]^{+}_{calc.}$ = 1297.6, from the adenovirus fiber protein from a 1:1 mixture of labeled and unlabeled adenovirus. Ion intensities are shown in parentheses.

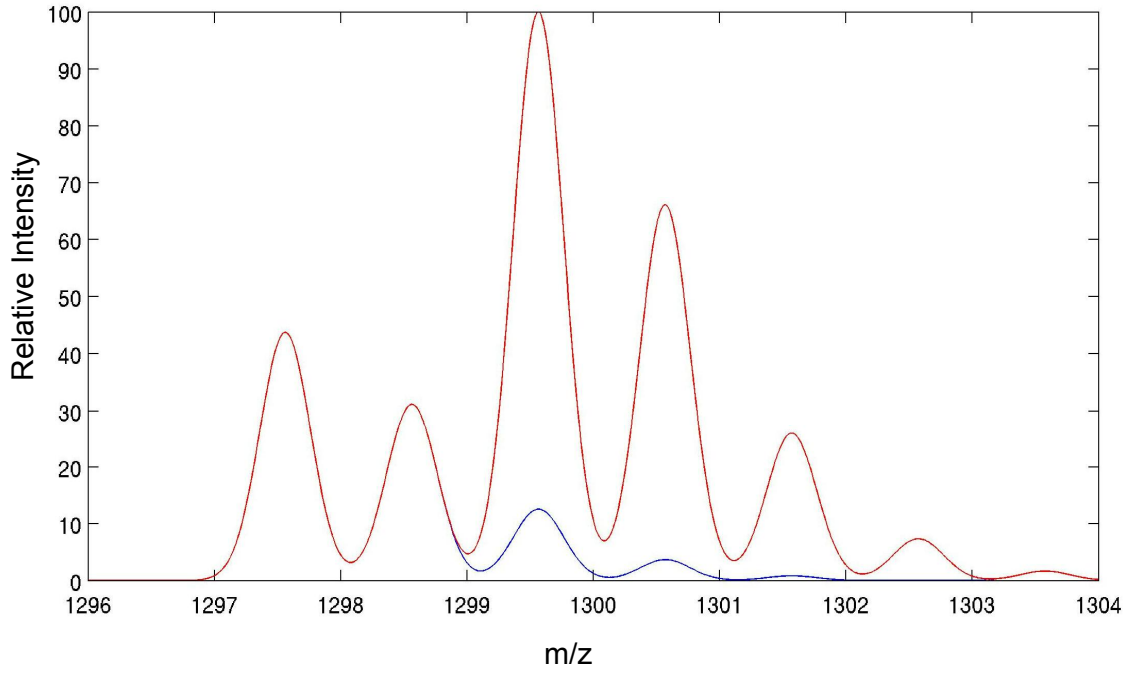


Figure 7.13: Theoretical isotope clusters of a 1:2 (unlabled:labeled) mixture of the peptide, PEYWNFRNGD, $[M+H]^+_{calc.}$ = 1297.6 from the adenovirus fiber protein.

Chapter 8: Conclusions and Prospectus

A novel strategy for the proteomic analysis of proteins has been developed and evaluated for biological samples with varying proteomic complexities. Digestion has been shown to exhibit high specificity that mimics enzymatic-based methods commonly used in proteomics. The work presented in this dissertation has shown that chemically generated, Asp-specific peptides are suitable for protein identification using both manual mass spectrometric analyses, and automated high-throughput workflows.

The efficiency of this cleavage method was demonstrated on a pure protein with various common modifications. Cleavage was shown to occur to at all aspartic residues; resulting in high sequence coverage of the protein. In addition, this digestion method enabled detection of acetylation and potential phosphorylation sites of the protein, which suggests that these modifications remain stable under the required experimental conditions for cleavage. Once cleavage conditions were optimized to achieve specificity, the necessary modifications were made to the MASCOT search parameter, “enzyme”, to accommodate all cleavage possibilities for protein identification using bioinformatics tools. This modified cleavage script has been submitted to MASCOT for use by other labs interested in using chemically generated Asp-specific peptides for protein identification.

Microwave-accelerated acid hydrolysis showed remarkable compatibility with rapid proteomics-based microorganism identification. The high temperature

and low pH conditions introduced by this digestion strategy resulted in the denaturation of both spores and viruses. As a result, selective solubilization and digestion of protein biomarkers were combined into a single step.[59] This enabled rapid analyses by reducing sample preparation time. Furthermore, this digestion procedure provided an attractive alternative to trypisn because of its robustness. We have demonstrated that proteins selectively released from microorganism samples can be successfully identified using either peptide mass fingerprinting, or MS/MS ion searching. The results of this work suggest that microwave-accelerated acid hydrolysis may simplify automated sample preparation for future applications in microorganism identification in the field.

Although a bioinformatic analysis of the yeast ribosome indicated that Asp-specific cleavage products would have a very different composition than tryptic peptides, our experimental results have shown that these peptides are suitable for use in a high throughput proteomic analysis. Asp-specific peptides have been shown to provide searchable CID spectra for peptide identification. Our results were consistent with recent reports in the literature, which have suggested that larger peptides can be advantageous for protein identification due to the high sequence coverage they provide.[73, 74] In addition, the analysis of longer peptides, results in fewer peptides, which can facilitate LC separations. While more than half of the yeast ribosome proteome was identified, it is believed that with optimization of digestion procedures and chromatography conditions, this coverage may be further improved.

In addition to the qualitative information microwave-accelerated acid digestions provides, we have demonstrated the feasibility of combining it with ^{18}O labeling for the relative quantitation of proteins. While this strategy is much more rapid that protease catalyzed ^{18}O labeling approaches, only a single ^{18}O atom was introduced into the longer Asp-specific peptides. Extensive overlap of ^{13}C isotopic species occurs with the ^{18}O pattern, and the development of novel computer algorithms will be required for deconvolution.

As mentioned in the introduction, the future development of proteomics relies heavily on the collaborative development of novel sample preparation techniques and analytical technologies. It is anticipated that with even further improvements in both of these arenas, microwave-accelerated acid hydrolysis will become integral part of proteomics analyses.

Bibliography

(1) Wasinger, V. C.; Cordwell, S. J.; Cerpapoljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humpherysmith, I. *Electrophoresis* **1995**, *16*, 1090-1094.

(2) Reid, G. E.; McLuckey, S. A. *Journal of Mass Spectrometry* **2002**, *37*, 663-675.

(3) Tao, W. A.; Aebersold, R. *Current Opinion in Biotechnology* **2003**, *14*, 110-118.

(4) Yao, X. D.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Analytical Chemistry* **2001**, *73*, 2836-2842.

(5) Miyagi, M.; Rao, K. C. S. *Mass Spectrometry Reviews* **2007**, *26*, 121-136.

(6) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Analytical Chemistry* **1991**, *63*, A1193-A1202.

(7) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; T., Y. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.

(8) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(9) Fenselau, C. *Analytical Chemistry* **1997**, *69*, A661-A665.

(10) Dass, C. *Principles and Practice of Biological Mass Spectrometry*; John Wiley & Sons, Inc: New York, 2001.

(11) Lay, J. O. *Mass Spectrometry Reviews* **2001**, *20*, 172-194.

(12) Cotter, R. J. *Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research*; American Chemical Society: Washington, 1997.

(13) Yates, J. R. *Journal of Mass Spectrometry* **1998**, *33*, 1-19.

(14) Beavis, R. C. a. F., D *Proteomics:A Trens Guide* **2000**, 22-26.

(15) Mann, M.; Wilm, M. *Analytical Chemistry* **1994**, *66*, 4390-4399.

(16) Olsen, J. V.; Ong, S. E.; Mann, M. *Molecular & Cellular Proteomics* **2004**, *3*, 608-614.

(17) Schultz, J. *Methods Enzymol.* **1967**, *11*, 255-263.

(18) Inglis, A. S. *Methods Enzymol.* **1983**, *91*, 324-332.

(19) Li, A. Q.; Sowder, R. C.; Henderson, L. E.; Moore, S. P.; Garfinkel, D. J.; Fisher, R. J. *Anal. Chem.* **2001**, *73*, 5395-5402.

(20) Zhong, H. Y.; Zhang, Y.; Wen, Z. H.; Li, L. *Nat. Biotechnol.* **2004**, *22*, 1291-1296.

(21) Zhong, H. Y.; Marcus, S. L.; Li, L. *J. Am. Soc. Mass. Spectrom.* **2005**, *16*, 471-481.

(22) Hauser, N.; Basile, F., Nashville, TN, May 23-27, 2004.

(23) Suzuki, T.; Kitajima, K.; Emori, Y.; Inoue, Y.; Inoue, S. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 6244-6249.

(24) Sheares, B. T. *Journal of Biological Chemistry* **1988**, *263*, 12778-12782.

(25) An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B. *Analytical Chemistry* **2003**, *75*, 5628-5637.

(26) Henderson, J. Y.; Moir, A. J. G.; Fothergill, L. A.; Fothergill, J. E. *European Journal of Biochemistry* **1981**, *114*, 439-450.

(27) Swatkoski, S.; Russell, S.; Edwards, N.; Fenselau, C. *Analytical Chemistry* **2007**, *79*, 654-658.

(28) Dick, L., Kim, Catherine, Qui, Difei, Cheng, Kuang-Chuan *Biotechnology and Bioengeneering* **2007**, *97*, 544-553.

(29) Geoghegan, K. F.; Hoth, L. R.; Tan, D. H.; Borzillerl, K. A.; Withka, J. M.; Boyd, J. G. *Journal of Proteome Research* **2002**, *1*, 181-187.

(30) Ribo, M.; Bosch, M.; Torrent, G.; Benito, A.; Beaumelle, B.; Vilanova, M. *European Journal of Biochemistry* **2004**, *271*, 1163-1171.

(31) Beck, A.; Bussat, M. C.; Klinguer-Hamour, C.; Goetsch, L.; Aubry, J. P.; Champion, T.; Julien, E.; Haeuw, J. F.; Bonnefoy, J. Y.; Corvaia, N. *Journal of Peptide Research* **2001**, *57*, 528-538.

(32) Nagle, G. T.; Knock, S. L.; Painter, S. D.; Blankenship, J. E.; Fritz, R. R.; Kurosky, A. *Peptides* **1989**, *10*, 849-857.

(33) Demirev, P. A.; Lin, J. S.; Pineda, F. J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 4566-4573.

(34) 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*, 1217-1223.

(35) Demirev, P. A.; Ho, Y. P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 2732-2738.

(36) Pineda, F. J.; Lin, J. S.; Fenselau, C.; Demirev, P. A. *Anal. Chem.* **2000**, *72*, 3739-3744.

(37) Pribil, P. A.; Patton, E.; Black, G.; Doroshenko, V.; Fenselau, C. *J. Mass Spectrom.* **2005**, *40*, 464-474.

(38) Warscheid, B.; Fenselau, C. *Proteomics* **2004**, *4*, 2877-2892.

(39) Fensealu, C. R., S.; Swatkoski, S.; Edwards, N. *European Journal of Mass Spectrometry* **2007**, *13*, 35-39.

(40) Hathout, Y.; Setlow, B.; Cabrera-Martinez, R. M.; Fenselau, C.; Setlow, P. *Appl. Environ. Microbiol.* **2003**, *69*, 1100-1107.

(41) Sonnhammer, E. L. L.; Eddy, S. R.; Durbin, R. *Proteins* **1997**, *28*, 405-420.

(42) Haft, D. H.; Selengut, J. D.; White, O. *Nucleic Acids Res.* **2003**, *31*, 371-373.

(43) Jensen, G. B.; Hansen, B. M.; Eilenberg, J.; Mahillon, J. *Environ. Microbiol.* **2003**, *5*, 631-640.

(44) Candelon, B.; Guilloux, K.; Ehrlich, S. D.; Sorokin, A. *Microbiology-Sgm* **2004**, *150*, 601-611.

(45) Sun, D. X.; Setlow, P. *J. Bacteriol.* **1987**, *169*, 3088-3093.

(46) Loshon, C. A.; Fliss, E. R.; Setlow, B.; Foerster, H. F.; Setlow, P. *J. Bacteriol.* **1986**, *167*, 168-173.

(47) Pain, A.; Renauld, H.; Berriman, M.; Murphy, L.; Yeats, C. A.; Weir, W.; Kerhornou, A.; Aslett, M.; Bishop, R.; Bouchier, C.; Cochet, M.; Coulson, R. M. R.; Cronin, A.; de Villiers, E. P.; Fraser, A.; Fosker, N.; Gardner, M.; Goble, A.; Griffiths-Jones, S.; Harris, D. E.; Katzer, F.; Larke, N.; Lord, A.; Maser, P.; McKellar, S.; Mooney, P.; Morton, F.; Nene, V.; O'Neil, S.; Price, C.; Quail, M. A.; Rabbinowitsch, E.; Rawlings, N. D.; Rutter, S.;

Saunders, D.; Seeger, K.; Shah, T.; Squares, R.; Squares, S.; Tivey, A.; Walker, A. R.; Woodard, J.; Dobbelaere, D. A. E.; Langsley, G.; Rajandream, M. A.; McKeever, D.; Shiels, B.; Tait, A.; Barrell, B.; Hall, N. *Science* **2005**, *309*, 131-133.

(48) Warscheid, B.; Fenselau, C. *Anal. Chem.* **2003**, *75*, 5618-5627.

(49) Warscheid, B.; Jackson, K.; Sutton, C.; Fenselau, C. *Anal. Chem.* **2003**, *75*, 5608-5617.

(50) Mesnage, S.; TosiCouture, E.; Mock, M.; Gounon, P.; Fouet, A. *Mol. Microbiol.* **1997**, *23*, 1147-1155.

(51) Williams, D. D.; Turnbough, C. L. *J. Bacteriol.* **2004**, *186*, 566-569.

(52) Mignot, T.; Mock, M.; Fouet, A. *Mol. Microbiol.* **2003**, *47*, 917-927.

(53) Nishihar.T; Nozu, Y.; Okada, Y. *Journal of Biochemistry* **1970**, *67*, 403-&.

(54) Hua, L.; Low, T. Y.; Sze, S. K. *Proteomics* **2006**, *6*, 586-591.

(55) Weber, K.; Konigsbe.W *Journal of Biological Chemistry* **1967**, *242*, 3563-&.

(56) Thomas, J. J.; Falk, B.; Fenselau, C.; Jackman, J.; Ezzell, J. *Analytical Chemistry* **1998**, *70*, 3863-3867.

(57) Davis, J. E. *Journal of Molecular Biology* **1963**, *6*, 203-&.

(58) Li, L.; Golding, R. E.; Whittal, R. M. *Journal of the American Chemical Society* **1996**, *118*, 11662-11663.

(59) Swatkoski, S.; Russell, S. C.; Edwards, N.; Fenselau, C. *Analytical Chemistry* **2006**, *78*, 181-188.

(60) Nishihara, T.; Fujisaki, S.; Nishimura, Y.; Minami, Y.; Yubisui, T. *Microbiology and Immunology* **2006**, *50*, 61-66.

(61) Nesvizhskii, A. I.; Aebersold, R. *Molecular & Cellular Proteomics* **2005**, *4*, 1419-1440.

(62) www.RMIDB.org.

(63) Shenk, T. *Fundamental Virology* **1996**, 979-1016.

(64) Yao, Z. P.; Afonso, C.; Fenselau, C. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1953-1956.

(65) www.expasy.org.

(66) Webster, A.; Leith, I. R.; Hay, R. T. *Journal of Virology* **1994**, *68*, 7292-7300.

(67) Diouri, M.; KeyvaniAmineh, H.; Geoghegan, K. F.; Weber, J. M. *Journal of Biological Chemistry* **1996**, *271*, 32511-32514.

(68) Lee, S. W.; Berger, S. J.; Martinovic, S.; Pasa-Tolic, L.; Anderson, G. A.; Shen, Y. F.; Zhao, R.; Smith, R. D. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 5942-5947.

(69) Planta, R. J.; Mager, W. H. *Yeast* **1998**, *14*, 471-477.

(70) Swatkoski, S. G., J.; Laine, O.; Petrov, A.; Dinman, J.; Gutierrez, P.; Edwards, N.; Fenselau, C.; , Indianapolis, IN 2007.

(71) http://ribosome.miyazaki-med.ac.jp/.

(72) Hathout, Y.; Setlow, B.; Cabrera-Martinez, R. M.; Fenselau, C.; Setlow, P. *Applied and Environmental Microbiology* **2003**, *69*, 1100-1107.

(73) Wu, S. L.; Kim, J.; Hancock, W. S.; Karger, B. *Journal of Proteome Research* **2005**, *4*, 1155-1170.

(74) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 2193-2198.