STRATEGIES FOR SHOTGUN PROTEOMICS: MIMICKING AN ARGC DIGEST AND ISOLATING N-TERMINAL PEPTIDES

by

DHANASHRI BAGAL

(Under the Direction of I. JONATHAN AMSTER)

ABSTRACT

Shotgun proteomic assays utilizing peptide assignment by accurate mass analysis can increase sample throughput. This thesis describes two shotgun proteomics strategies to facilitate proteomics analysis by accurate mass measurement using Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). ArgC digestion of a protein produces peptides that are more specific for protein identification than those from trypsinolysis. However, ArgC lacks specificity, displaying some trypsin-like behavior. This thesis reports a method for mimicking an ArgC digestion using lysine acetylation and trypsin digestion. Furthermore, it describes a novel technique for improving protein identification by isolating N-terminal peptides using a mass defect labeling approach to remove stray internal peptides. This approach to N-terminal peptide restricted shotgun proteomics provides a substantial gain in assignment specificity.

INDEX WORDS: Mass Spectrometry, Proteomics, FTICR MS, Accurate mass measurement, ArgC digestion, N-terminal peptides, Mass defect labeling

STRATEGIES FOR SHOTGUN PROTEOMICS: MIMICKING AN ARGC DIGEST AND ISOLATING N-TERMINAL PEPTIDES

by

DHANASHRI BAGAL

B.Tech., Mumbai Institute of Chemical Technology, India, 2003

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2006

© 2006

Dhanashri Bagal

All Rights Reserved

STRATEGIES FOR SHOTGUN PROTEOMICS: MIMICKING AN ARGC DIGEST AND ISOLATING N-TERMINAL PEPTIDES

by

DHANASHRI BAGAL

Major Professor: I. Jonathan Amster

Committee: Ron Orlando
Robert Phillips

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2006

To Mom and Dad….

// ACKNOWLEDGEMENTS heading
ACKNOWLEDGEMENTS

I would like to thank my family and friends for being there with me throughout this period. A special thanks to Abhishek, for being so patient with me and making this difficult phase easy. I would also like to thank my group members, Richard Wong, Jeremy Wolff, Hilda Hernandez, Chunyan Li, Li Jing, Jennifer Cobb, Lisabeth Hoffman and Doug Frey for all their ideas, support and assistance in the lab and for making graduate school at University of Georgia so memorable.

I would like to thank Dr. Robert Phillips for synthesizing the amine specific mass defect label. My thanks is additionally extended to professors Jon Amster, Ron Orlando and Robert Phillips for taking time out of their busy schedules to serve on my committee. And above all, I would like to thank my advisor Dr. Jon Amster for all the support, understanding and ideas, without which this research would not be possible.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

CHAPTER

1 LITERATURE REVIEW 1

Proteomics 2

Mass Spectrometry in Proteomics 5

References 25

2 EXPERIMENTAL 31

Instrumentation 32

Sample Preparation 40

References 43

3 MIMICKING AN ARGC DIGESTION BY LYSINE ACETYLATION AND TRYPSINOLYSIS FOR IMPROVED SHOTGUN PROTEOMICS BY ACCURATE MASS MEASUREMENT 44

Introduction 45

Sample Preparation 48

Results and Discussion 56

Conclusions 77

References 80

4 A MASS DEFECT LABELING APPROACH TO N-TERMINLAL PEPTIDE RESTRICTED SHOTGUN PROTEOMICS 83

Introduction 84

Experimental 90

Results and Discussion 92

Conclusions 98

References 99

5 CONCLUSIONS 101

References 105

CHAPTER 1

LITERATURE REVIEW

PROTEOMICS

A single genome can give rise to qualitatively and quantitatively different proteomes under different biological conditions.[1] Although, DNA microarrays can profile the amount of messenger RNA (mRNA) expressed from each gene, this technology is insufficient to fully characterize the biological state of a system because the level of mRNA does exhibit a 1:1 correlation with the level of protein expression.[1,2] Also, protein maturation and degradation are dynamic processes which cause the amount of active protein to vary considerably from the mRNA level. The need to complement mRNA expression analysis has thus resulted in the emergence of the field of proteomics to directly analyze protein expression levels from an organism.[3,4]

While precise definitions may vary, proteomics refers to a set of techniques that allows one to comprehensively study protein expression in an organism under a specific set of conditions. A major goal of proteomics is the global and quantitative measurement of the proteins expressed in cells or tissues.[5,6] Unlike the static nature of the genome, which is essentially identical in every cell of an organism, the proteome is dynamic, constantly changing and responding to internal and external stimuli. Proteomic studies cope with problems such as limited and variable sample material, a protein abundance dynamic range of more than 10^6 fold, post translational modifications and a plethora of perturbations.[7] Proteomic studies integrate separation science for the separation of proteins and peptides, analytical science for identification and quantification and bioinformatics for data management and analysis.[8] The two dominant

separation methods used in proteomics are 2-dimensional gel electrophoresis (2DE) and, increasingly, liquid chromatography.[9]

Classical Proteomics

Until recently proteome analysis has combined 2DE for protein separation, visualization and quantification with protein identification by mass spectrometry.[8] High resolution 2DE combines two orthogonal separations to improve the resolution of complex protein mixtures. The first dimension separates the proteins according to their isoelectric point in an immobilized pH gradient gel and the second dimension separates the proteins according to their molecular weight, by sodium dodecyl sulfate polyacrylamide gel electrophoresis.[10] Subsequent steps include spot excision, proteolytic in-gel digestion, peptide mass measurements by mass spectrometry, and database interrogation for protein identification (Figure 1.1).[6,11] However, the 2DE based proteome technology has difficulty in detecting some classes of proteins, including those of high or low molecular weight, extreme pI, or poor solubility. It has also been shown that the two dimensional gel electrophoresis-mass spectrometry (2DE-MS) approach is incapable of measuring low abundance proteins without pre-gel enrichment. Moreover, 2DE is labor intensive and difficult to automate.[4,12,13] An alternative and efficient approach is the use of liquid chromatography as a separation method as it is faster, easier to automate, and couples more readily to MS than 2DE.[5,13,14] One more attractive feature of the chromatographic systems is that they allow many dimensions of analysis to be coupled by analyte transfer between dimensions through automated valve switching.[15]

Shotgun Proteomics

Shotgun proteomics is a gel-less approach that utilizes liquid chromatography as a separation technique for a complex mixture of peptides that is generated by a single tryptic

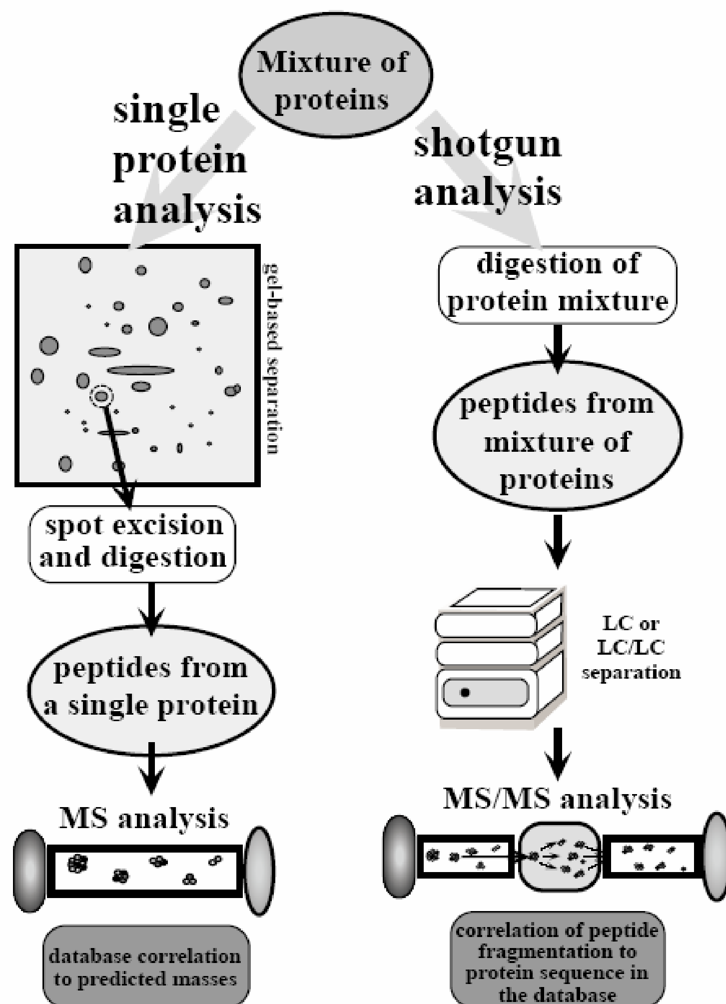


Figure 1.1 A comparison between classical proteomics (single protein analysis) and shotgun proteomics.[6]

digestion on the unseparated proteins (batch digestion). This high throughput technique results in a complex mixture of proteolytic peptides which is separated by liquid chromatography prior to the mass spectrometry analysis (Figure 1.1). Each peptide must be correlated with its precursor protein. This can be done by using tandem mass spectrometry to generate fragmentation data that

can be used to search a database or by using accurate mass measurement. The biggest advantage of shotgun proteomics over 2DE is the speed of the experiment (days versus weeks)[6] and the total amount of sample needed (50 µg versus 500 µg)[6,16] for analysis. Modern shotgun strategies for increasing protein identification and quantification in proteome analysis include isotope-coded affinity tags (ICAT),[17] mass-coded abundance tagging (MCAT),[18-20] isobaric tags for relative and absolute quantitation (iTRAQ),[21] stable isotope labeling of proteins/peptides,[22,23,24] global internal standard strategy (GIST),[9,25] along with accurate mass and time tag (AMT),[9] and shotgun proteomics that includes the use of 2 dimensional chromatography as in the multi-dimensional protein identification (MudPIT), a LC/LC-MS/MS strategy.[3,14]

MASS SPECTROMETRY IN PROTEOMICS

Mass spectrometry (MS) is a method that measures mass-to-charge ratio (m/z) of gas-phase ions with extremely high sensitivity. In the last 20 years MS has played an increasingly significant role in biological sciences and related studies. MS owes its success to the soft ionization techniques electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI). These ionization methods were developed in the late 1980's and have catalyzed changes in MS instrumentation and fundamentally changed the analysis of proteins. Both these ionization techniques made possible the generation of ions from large, non volatile, thermally labile analytes, such as proteins and peptides, without significant fragmentation.

Under the appropriate conditions, they are capable of maintaining non-covalent interactions during the ionization process.[26] Earlier soft ionization techniques like chemical ionization (CI) required a sample to have sufficient vapor pressure and thermal stability to enter the gas phase in the ion source. This requirement inhibited the application of mass spectrometry to biological macromolecules.[27] However, with ESI and MALDI, proteins can now be analyzed

to reveal elemental composition, complete or partial amino acid sequence,[28,29] post- translational modifications,[30,31] protein-protein interaction sites,[32,33] and even provide insight into conformational aspects.[34-36] The achievements in the MS field due to the introduction of soft ionization techniques were recognized by the awarding of the 2002 Nobel prize in chemistry to John Fenn and Koichi Tanaka for their pioneering work on ESI and MALDI respectively.[37,38]

Electrospray Ionization

Dole et al. first demonstrated the generation of macroions by electrospray,[39] but John Fenn was the first to couple ESI with MS.[37] This technique produces gaseous ionized molecules directly from the solution at atmospheric pressure. An advantage of sample ionization performed under atmospheric pressure conditions is that it yields an ionization efficiency which is 10^3-10^4 times greater than in a reduced pressure chemical ionization source.[40] Also, as ESI involves the continuous introduction of solution, it can easily be interfaced with liquid separation techniques such as high performance liquid chromatography (HPLC) or capillary electrophoresis (CE).[41]

Electrospray produces a fine spray of highly charged droplets by the application of a strong electric field to a liquid passing through a capillary tube. A potential difference of 2-4 kV is applied between the capillary and the counter electrode, separated by 0.3-2 cm, producing electric fields of the order of 10^6 V/m.[40] Under the influence of the electric field, the solution disperses into a fine spray of charged droplets. Either dry gas, heat or a combination of the two are used to cause the solvent to evaporate from each droplet. As the size of the droplet decreases, the charge density on its surface increases. The mutual coulombic repulsion between like charges on the surface becomes so great that it exceeds the forces of surface tension, thereby ejecting ions from the droplet, which assumes a conical shape with a “Taylor cone” (Figure 1.2).[7,40,42] Another possibility is that the droplet explodes releasing the ions.[7,41] The exact mechanism

involved in the production of isolated gaseous ions by ESI is yet not fully understood.[43] The ions enter a vacuum chamber through an orifice, and are directed by electrostatic lenses to the mass analyzer.

The ions generated by ESI of larger peptides and proteins usually carry multiple protons. This is an important phenomenon, as the mass spectrometer measures the mass-to-charge ratio. Multiple charging makes it possible to observe high mass molecules at low mass-to-charge using an instrument with a relatively small mass range. Multiple charging also offers advantages in tandem mass spectrometry. Upon fragmentation of a multiply charged precursor ion, more fragment ions can be observed as compared to fragmenting a singly charged precursor ion.[41]

One of the shortcomings of ESI is in the analysis of mixtures. For electrospray ionization of mixtures, the most polar analyte undergoes the greatest suppression.[44] In the high concentration range, solutions that contain multiple analytes suffer from competition and ion suppression during electrospray ionization.[42,45] However, this can be avoided by dilution and by using separation techniques such as high performance liquid chromatography (HPLC) or capillary electrophoresis (CE) prior to mass analysis.[42] Also, the presence of salts, buffers, detergents and other additives reduce the sensitivity drastically. Low flow electrospray, known as nanoelectrospray[46] improves sensitivity with its ability to preconcentrate the sample and desalt on-line prior to sample introduction into the mass spectrometer, and is thus more tolerant of salts and impurities.[41,43,47,48] The lower flow rates (nL/min) that are characteristic of nanoelectrospray MS not only improve ionization efficiency but also allow for longer analysis time.[47] Table 1.1 lists the advantages and disadvantages of ESI.

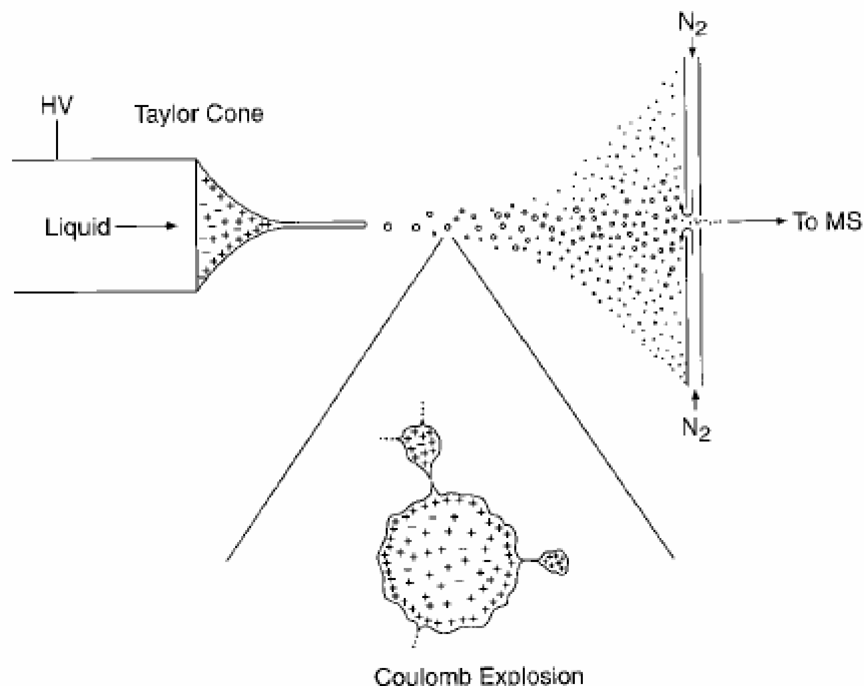


Figure 1.2 A schematic representation of a generic electrospray ionization source.[43]

Matrix Assisted Laser Desorption Ionization

Karas and Hillenkamp introduced the technique of matrix assisted laser desorption ionization (MALDI) in 1988, for the analysis of proteins with molecular masses exceeding 10 kDa.[49] MALDI desorbs and ionizes an analyte present in a crystalline matrix by using a pulsed laser to deposit energy on a nanosecond timescale. The analyte is co-crystallized with a large excess of a matrix material that strongly absorbs light from the laser. The matrix plays an important role in isolating the sample molecules from each other and absorbing the photons from the laser, usually a 337 nm N_2 laser.[7] The matrix molecules absorb the laser energy and rapidly evaporate into gas phase carrying the intact sample molecules. Little energy is transferred to the analyte molecules, allowing ionization without fragmentation. Unlike ESI, MALDI tends to form

singly-charged ions. The exact origin of ions produced by the MALDI process is not yet fully understood. The widely accepted mechanism involves gas phase proton transfer in the expanding matrix plume with photoionized matrix molecules (Figure 1.3). The matrix plays an important role in the ionization process and therefore different classes of analytes require different matrices. 2,5-dihydroxybenzoic acid (DHB) is a suitable matrix for analysis of peptides by Fourier transform mass spectrometry (FTMS).[7,43] For time of flight mass spectrometry (TOF-MS), α-cyano-4-hydroxy cinnamic acid is the best peptide matrix.[50] Proteins are best desorbed from sinapinic acid.[50,51] Other matrices have been reported for nucleic acids[52-54] and polymers.[50,55]

Conventional vacuum MALDI results in metastable losses (loss of H_2O or NH_3 from the molecular species) thus yielding extra peaks.[56,57] These are potentially pronounced with FTMS. These losses can be reduced considerably by means of collisional cooling at intermediate or high pressure.[58] Burlingame et al. reported MALDI at atmospheric pressure, (API-MALDI).[59] Later, intermediate pressure MALDI (IP-MALDI) was reported, wherein a collision gas (argon) is introduced to the ionization region as a short burst prior to firing of the laser, permitting the collisional thermalization of ions as they are formed and reducing metastable losses.[56,58]

The MALDI process is totally independent of the properties and size of the analyte and therefore easily allows desorption and ionization of the analytes with very high molecular weights. The pulsed laser that is required for MALDI produces packets of ions rather than a continuous beam. The resulting ion pulse must be coupled to a mass analyzer that can measure either a complete mass spectrum without field scanning(e.g. TOF-MS), or that can trap all the ions for subsequent mass analysis (e.g. FTMS).[7,60] Though MALDI can tolerate millimolar

Table 1.1 Advantages and disadvantages of ESI

Advantages	Disadvantages
Analysis of compounds with a molecular weight up to about 100,000 Da is possible	Presence of salts can reduce sensitivity drastically
ESI is very sensitive; typical sensitivity ranges from femtomole to low picomole	Analysis of mixtures is difficult as each compound gives rise to several signals corresponding to sample molecules carrying a range of protons
ESI is a softest ionization technique, making it possible to observe native biological complexes bound by non covalent interactions	Multiple charging can be confusing especially in complex mixture analysis
Easily adaptable to liquid chromatography techniques	On line coupling with LC or CE equipment is required for complex mixture analysis
Easily adaptable to tandem mass analyzers such as ion traps and triple quad instruments	Sample purity is important
Multiple charging allows for analysis of high mass ions with a relatively low m/z range instrument	
No matrix interference	

concentrations of salts, the major shortcoming of MALDI is matrix background at low molecular weight (below m/z 300) and the possibility of photodegradation during laser desorption/ ionization. Table 1.2 lists the advantages and disadvantages of MALDI.

Currently, genomics, proteomics and metabolomics increasingly require the analysis of highly complex mixtures and detection of various analytes over a wide range of concentrations, from hormones and growth regulators present at pM levels to proteins whose expression levels may vary of six or more orders of magnitude. These challenges demand instruments with better performance characteristics including sensitivity, mass resolution, accuracy, dynamic range and tandem mass spectrometry abilities.[61]

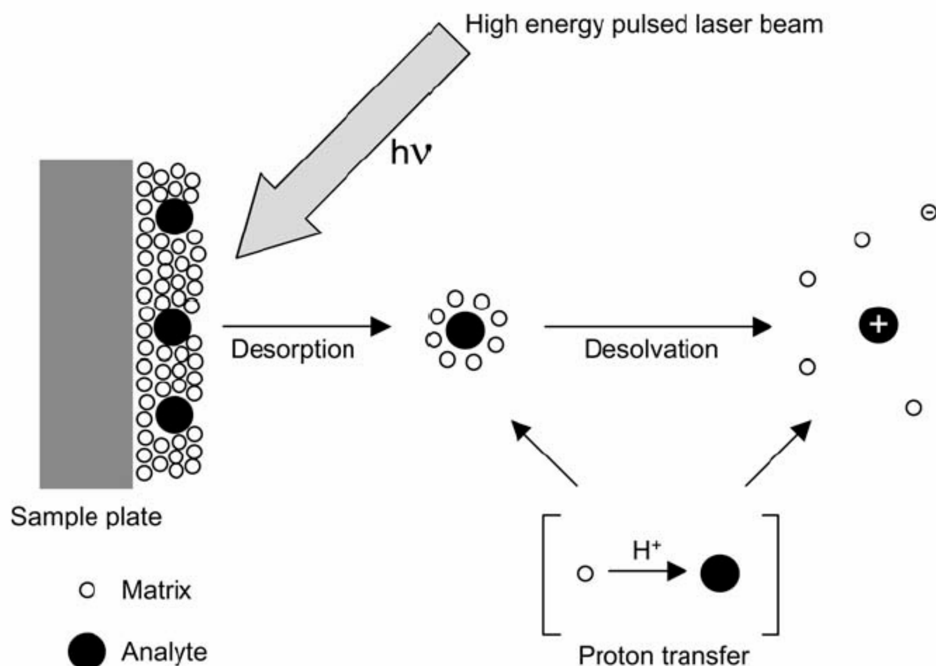


Figure 1.3 A schematic of the MALDI process.[40]

Table 1.2 Advantages and disadvantages of MALDI

Advantages	Disadvantages
Analysis of compounds with a molecular weight in excess of 300, 000 is possible	Analysis of compounds with a molecular weight below 700 Da is difficult due to presence of intense matrix signal
MALDI is very sensitive; typical sensitivity on the order of low fetmomole to low picomole. Attomole sensitivity is possible	Possibility of photodegraadtion by laser desorption
MALDI is a soft ionization technique with little or no fragmentation	Online line coupling with LC or CE equipment is difficult
Usually only singly charged ions and a small amount of doubly charged ions are generated making analysis of mixtures possible	Acidic matrix used in MALDI may cause degradation on some compounds
MALDI is reasonably tolerant towards presence of salts, buffers and other additives	

Fourier Transform Ion Cyclotron Mass Spectrometry

Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) is a technique known for its high resolution and mass accuracy as it offers 10-100 times mass resolving power and mass accuracy than any other mass analysis technique.[62] FTICR-MS derives from ion cyclotron resonance spectrometry, the theory of which was developed by Lawrence in 1930s.[63]

Comisarow and Marshall built the first FTICR instrument in 1978, applying the Fourier transform methods to ion cyclotron resonance spectrometry.[64] Of all mass spectrometric methods, FTICR-MS offers a unique combination of analytical qualities and is a complete ion laboratory by itself. Along with high resolution and high mass accuracy, it also combines non-destructive multichannel detection, long ion observation times, the capability to perform gas phase reactions on trapped ions and tandem mass analysis.[34]

The analyzer cell is the heart of the FTICR-MS instrument. Here the ions are stored, mass analyzed and detected. The analyzer cell operates within a strong, homogenous magnetic field of a superconducting magnet. The performance of the FTICR instrument improves as the magnetic strength increases and, commercial instruments are available with field strengths of 12 tesla.[65,66] Figure 1.4 shows the improvement in theoretical mass resolving power of FTICR instrument as a function of magnetic field strength.[67] The plot shows that mass resolving power decreases with increasing mass-to-charge. However, the resolution is improved by increasing the time domain acquisition period.

Several analyzer cell designs have been developed over time. A typical cubic analyzer cell consists of 6 plates arranged in the shape of a cube (Figure 1.5). The cell is oriented so that one opposing pair of plates is orthogonal to the direction of magnetic field and the 2 remaining pairs of plates lie parallel to the magnetic field. The plates orthogonal to the magnetic field are

called the trapping plates. A small symmetric voltage applied to the trapping plates helps store ions which then undergo a simple harmonic oscillation between the trapping plates along the magnetic axis. The remaining 2 pairs of plates which lie parallel to the magnetic field are used for excitation and detection of ions.

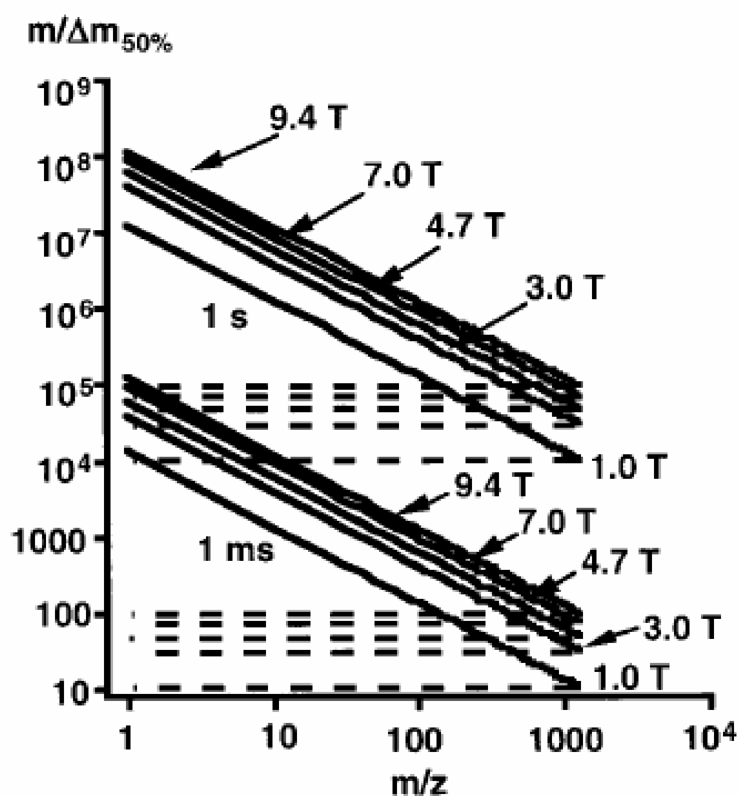


Figure 1.4 Theoretical mass resolving power for FTICR MS at 1.0, 3.0, 4.7, 7 and 9.4T as a function of mass to charge ratio. The dashed lines indicate mass resolving power of magnetic field swept instruments like magnetic sectors.[66]

Ion Motion

The basis of FTICR-MS is ion cyclotron motion. An ion experiences a Lorentz force that is perpendicular to both the direction of its velocity and magnetic field causing it to travel in a circular orbit that is perpendicular to the magnetic field (Figure 1.6). As a result of its interaction with the unidirectional magnetic field, each ion moves in a circular orbit. This is referred to as cyclotron motion. The cyclotron frequency (f_c) is determined by three physical parameters namely, the magnetic field (B), the charge present on the ion (q), and the mass of the ion (m) as shown by equation 1.1. In an FTICR mass spectrometer, the magnetic field is held constant and the mass-to-charge ratio of an ion can be determined by measuring the cyclotron frequency.[34, 65]

$$f_c = \frac{qB}{2\pi m}$$ Equation 1.1

An FTICR-MS experiment consists of a series of temporally separated events.[66] A simple experimental sequence consists of 4 events, namely quench, ion formation, ion excitation and detection. The quench event is used to empty the analyzer cell prior to ion formation which can take place either inside or outside the cell.[65] Most commonly in modern FTICR-MS instruments, the ionization and detection regions are physically separated with one or more stages of differential pumping.[68] An external ion source allows MALDI or ESI to be coupled to the ultrahigh vacuum analyzer regions. Ions are extracted from the external ion source and guided to the analyzer cell by means of ion optics, i.e. electrostatic ion optics or by a radiofrequency (RF) multipole ion guide. For example, a hexapole lens in an RF-only mode operates as an ion

guide.[69,70] The ions are then captured in the analyzer cell by means of gated trapping[71,72] or sidekick,[73] prior to mass analysis.

Ion excitation/detection

The excitation and detection events follow the ion formation event. Ions that are trapped in the analyzer cell have a relatively small cyclotron orbit (less than 1 mm). In order to detect ions, they are excited into coherent motion by applying a sinusoidal voltage to the excitation plates. An ion spirals outwards when its cyclotron frequency is in resonance with the applied RF electric field. Ions that are not in resonance with the sinusoidal frequency do not absorb energy and remain at the center of the cell. Ions of many masses can be detected simultaneously

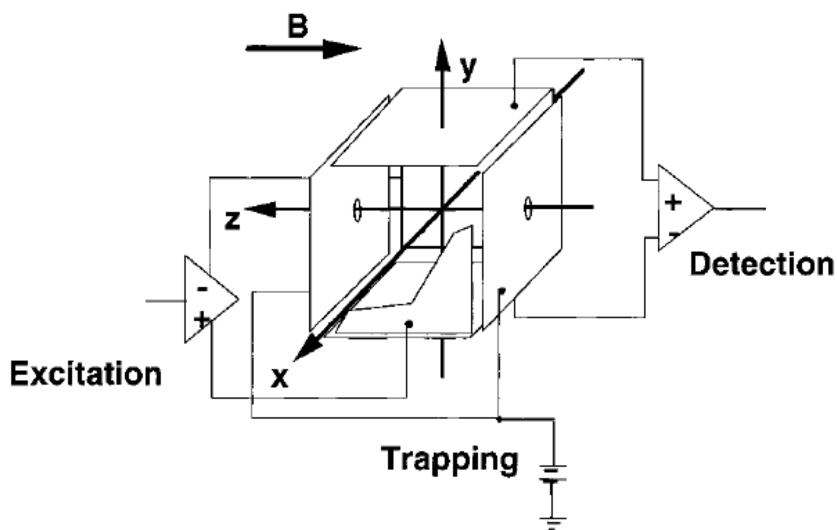


Figure 1.5 A cubic analyzer cell describing the electronic circuitry in detail. B is the direction of magnetic field.[66]

(broadband detection) if a range of frequencies is applied during the excitation event. A rapid frequency sweep called a “RF chirp” is commonly used for broadband detection. Application of a single RF frequency allows ions of a given m/z ratio to absorb energy and spiral outwards until they strike the excite or detect plates and are neutralized. This feature is used to remove mass selected ions from the analyzer cell. If the field is turned off before the ions strike the plates, they undergo cyclotron motion as shown in Figure 1.7. All ions of the same mass-to-charge ratio are excited coherently and undergo cyclotron motion as a packet. As they pass the cells electrodes, the coherently orbiting ion packet attracts electrons to the first one and then the other of the 2 detection plates through the external circuit that joins them. This alternating current is called the image current and its detection is a non-destructive process.[65]

The periodic cyclotron motion of the ions produces a sinusoidal image signal, which is amplified and digitized. Ions of many masses can be detected simultaneously (broadband detection) by applying many frequencies during the excitation event. The image current (transient) that results from ions of several mass-to-charge values is a composite of sinusoids of different frequencies and amplitudes, and constitutes the time domain signal. Applying a Fourier transform to the time domain transient yields the frequency components. The frequency spectrum is converted into the mass spectrum using a calibration equation that is derived from the cyclotron frequency expression. The trapping voltage significantly affects the ion population in the analyzer cell. Higher trapping voltages increase the ion population, which may cause frequency shifts due to space charge effects. Very low trapping voltages reduce ion densities and the accompanying space charge effects but adversely affect signal intensity. Therefore an optimum trapping voltage along with a calibration equation that corrects for the electric field

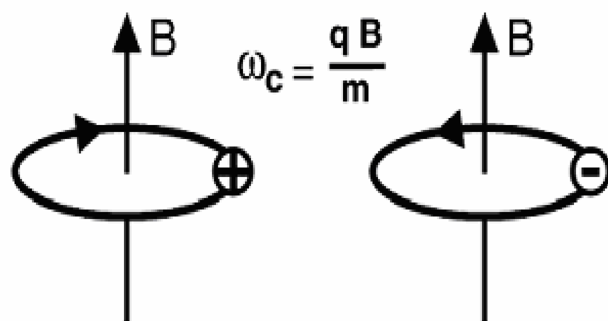


Figure 1.6 Ion cyclotron motion in the plane perpendicular to the magnetic field lines. Note that the positive and negative ions orbit in opposite senses.[66]

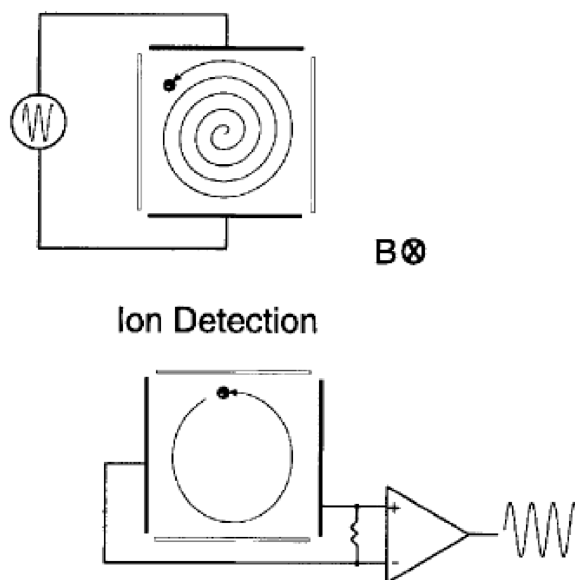


Figure 1.7 Ion excitation and detection in FTICR mass spectrometry. The magnetic field is directed into the plane of the paper. The image signal produced on the detection plates is amplified by using an amplifier.[65]

effects needs to be used for effective FTICR-MS analysis. Equation 1.2 shows the calibration equation that is used in our laboratory while performing FTICR-MS analysis, where f_{obs} is the observed frequency, m/z is the calculated mass-to-charge ratio and A and B are the calibration constants. The mass resolution improves with the increase in the length of the transient. The maximum resolution obtained is shown in equation 1.3 where R is the resolving power, f_c is the cyclotron frequency, and T is the duration of transient. Figure 1.8 compares the resolution obtained at different transient durations. The resolution obtained by performing a Fourier transform on the full 1.2 seconds transient is much better than that on the first 0.15 seconds of the same transient. However, the transient signal decays with time as the collisions between the ions and the neutrals destroy the coherence of the ion packet, thereby imposing a physical limitation to the duration of the transient. Thus an ultrahigh vacuum is required (10^{-10} Torr), especially during detection to ensure minimum collisions.[65]

$$\frac{m}{z} = \frac{A}{B + f_{obs}}$$ Equation 1.2

$$R = \frac{f_c T}{2}$$ Equation 1.3

Limitations of FTICR MS

A limiting factor to accurate mass measurement by FTICR-MS is space charge, which arises from the influence of the electric field on the ions trapped in the ion cell upon each other. In other words, mass accuracy is affected by the total number of ions trapped in the ICR cell. The electric field from the trapped ions produces space charge which can produce mass shifts that can be as large as a hundred ppm in extreme cases.[74,75] More commonly this effect limits mass accuracy to ~ 10 ppm.

In order to achieve reliable mass accuracy, a hybrid mass spectrometer that combines linear ion trap (LTQ) and FTICR technologies in a single instrument has been introduced (Figure 1.9). The LTQ-FTICR takes advantage of the strengths of both the linear ion trap and FTICR mass spectrometers and provides extraordinary mass accuracy (~5 ppm with external calibration).[76,77,78] The ions are generated in the source and then guided through ion optics to the linear ion trap. With a fast pre-scan, the LTQ determines the total ion current and uses this information to calculate the optimal injection time for collecting a specified number of ions. Ions are then injected for the calculated period of time, and then ejected from the ion trap into the FTICR analyzer cell, where they are excited and detected with high mass accuracy. This technology yields a reproducible ion population from scan-to-scan.[47] However, this technology can only be beneficial to continuous ion sources such as ESI, and cannot be applied to pulsed ion sources such as MALDI, where there is fluctuation in ion population between successive ionization events.

Ultrahigh mass accuracy offers an alternative to tandem mass spectrometry for proteomics analyses.[48,79] High mass accuracy can be used as an additional constraint in database searching.[34,62] In peptide mass fingerprinting, the more accurately masses are measured, the more

protein identifications can be made, with greater confidence in the final identification. Figure 1.10 shows the number of proteins from *Escherichia coli* that can be correlated to specific tryptic peptides measured at a mass accuracy of 1, 10 and 100 ppm. The data was generated by calculating the masses of all peptide generated by *in silco* digestion of the entire proteome from *E. coli* using trypsin as a protease. Overall, the plot suggests that as the mass accuracy increases, the search specificity increases significantly. However, even at 1 ppm mass accuracy, additional constraints for database search are required (especially for peptides below 1000 Da).

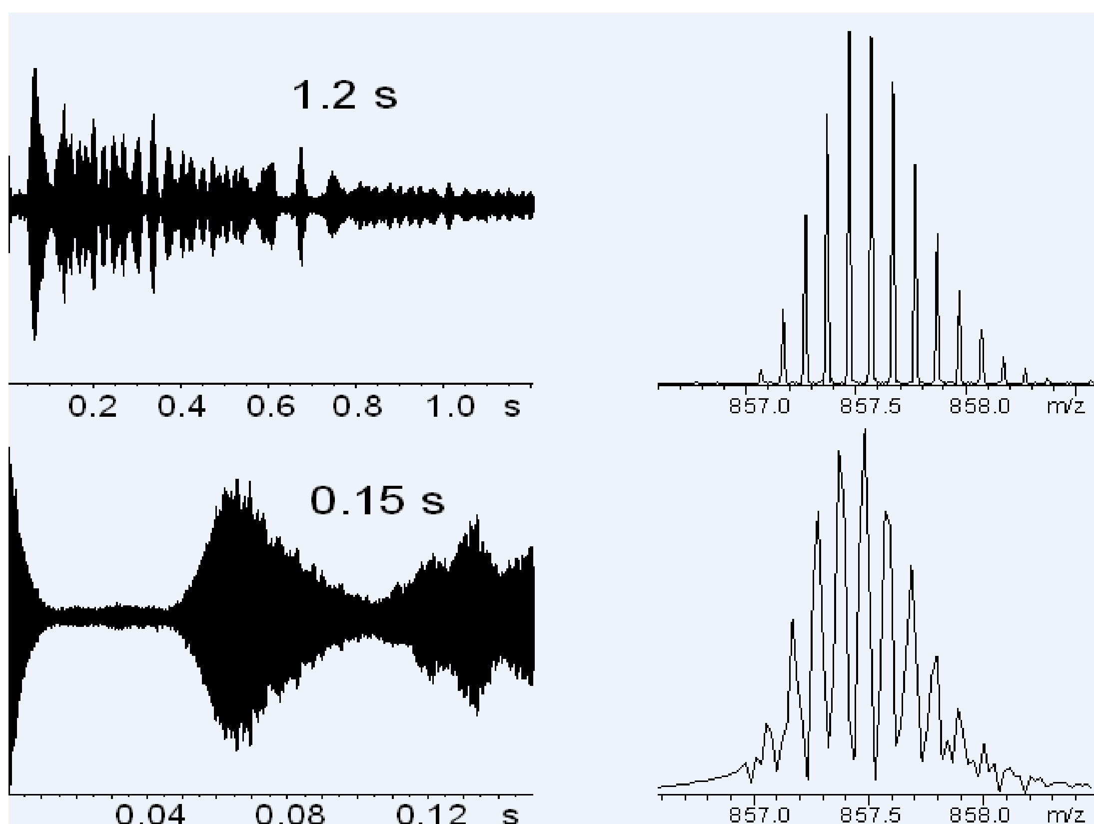


Figure 1.8 Comparing the duration of transient with mass resolution.[65]

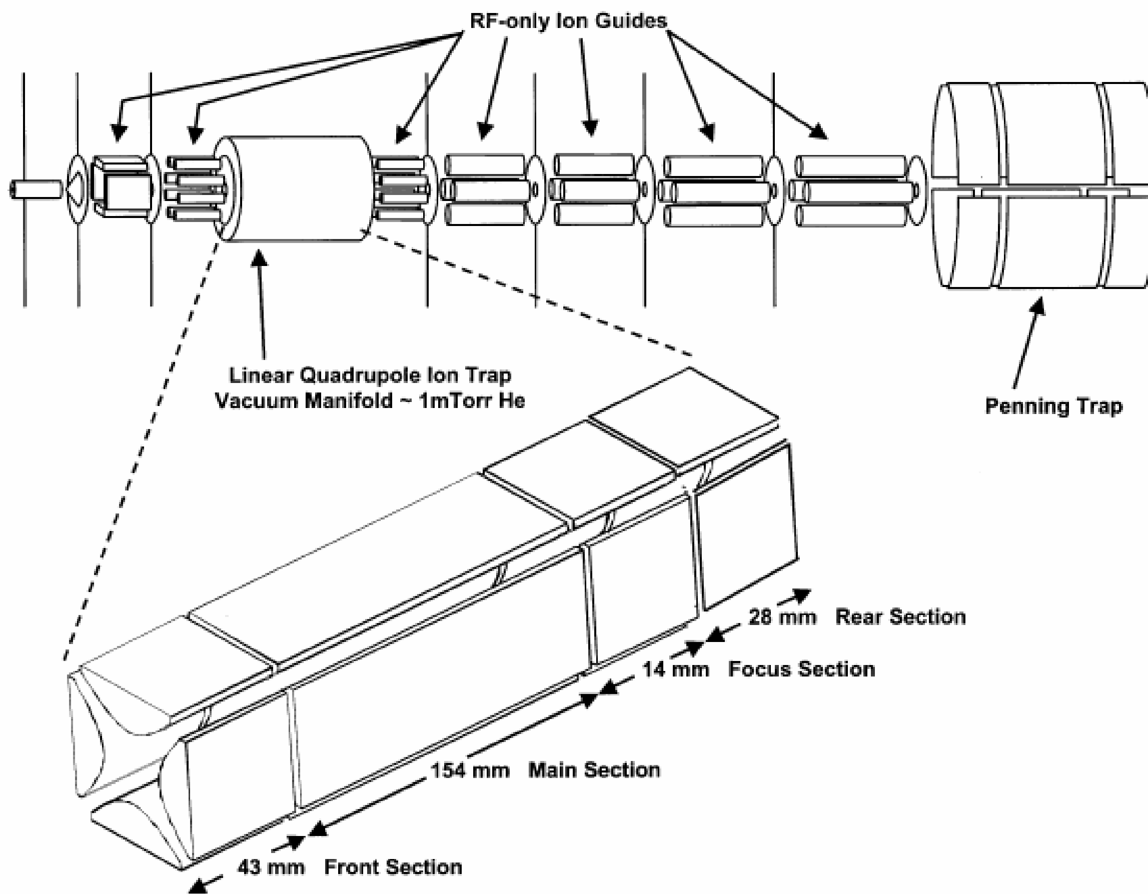


Figure 1.9 A schematic of a hybrid ESI- LTQ-FTICR mass spectrometer.[76]

Research in our laboratory is directed towards developing techniques of shotgun proteomics that rely upon accurate mass measurement, ^{15}N labeling, and by chemical derivatization methods such as mass defect labeling.[80,81] For comparative and quantitative proteomic analysis of cell cultures, proteins can be metabolically labeled by growing cells in ^{15}N

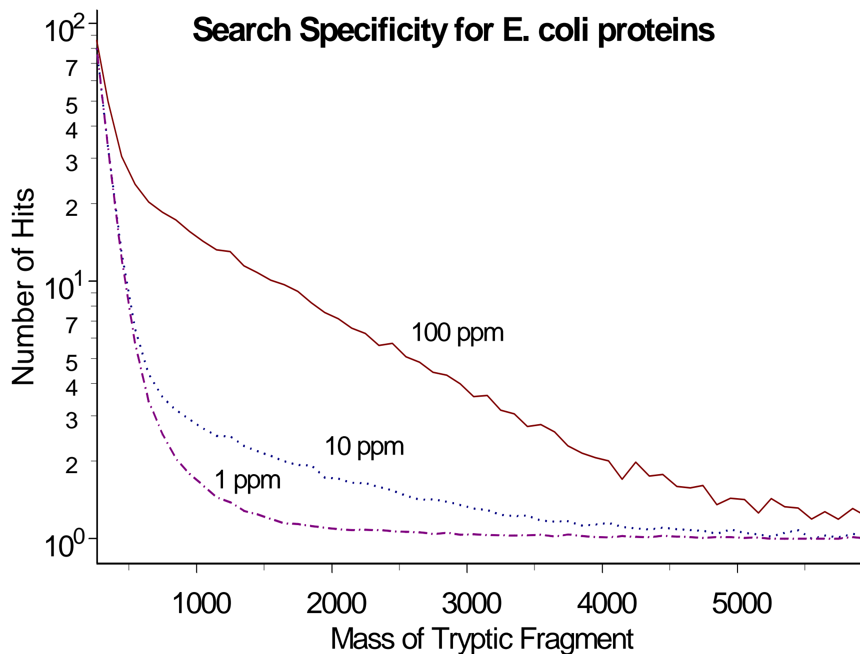


Figure 1.10 A plot of search specificity for *Escherichia coli* proteins identified from its tryptic peptides measured at 1, 10 and 100 ppm mass accuracy.[83]

enriched media. For this approach, two pools of the organism under study are grown on a medium that contains natural abundance of nitrogen and also on a medium that is enriched with ^{15}N. Both the proteome pools are combined and then compared. The proteome pools are digested once with a protease and then the resulting peptides are separated using liquid chromatography. Pairs of peptides elute and the difference in the spacing between the pairs depends on the number of nitrogen atoms in the respective peptide.[22,80,82] In theory, all peptides from the combined samples exist as analyte pairs of identical sequence but different masses. The peptide pairs have

the same physico-chemical properties and behave similarly under any conceivable isolation or separation step. Thus, the ratios between the intensities of the lower and upper mass components of these pairs of peaks provides an accurate measure of relative abundance of peptides (and hence protein) in the original protein mixture. High resolution and mass accuracy provided by FTICR allows the mass difference between the peaks to be accurately determined and the peptide pairs are therefore identified with high confidence.[34,62]

Scope of this Thesis

This thesis describes two different shotgun proteomics approaches to improve protein identification. Chapter 3 describes a strategy to yield large, unique peptides which are more useful for protein identification as compared to the regular tryptic peptides. Chapter 4 describes a novel technique to exclusively isolate N-terminal peptides using mass defect labeling for protein identification. Chapter 2 details the different experimentation protocols and instruments used for analysis.

References

(1) Dongre, A. R.; Opiteck, G.; Cosand, W. L.; Hefta, S. A. *Peptide Science* **2001**, *60*, 206-211.

(2) Williams, K. L. *Electrophoresis* **1999**, *20*, 678-688.

(3) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nature Biotechnology* **2001**, *19*, 242 - 247.

(4) Quadroni, M.; James, P. *Electrophoresis* **1999**, *20*, 664-677.

(5) Gygi, S. P.; Rist, B.; Aebersold, R. *Curr. Opin. Biotechnol.* **2000**, *11*, 396-401.

(6) McDonald, W. H.; Yates, J. R. *Disease Markers* **2002**, *18*, 99-105.

(7) Lane, C. S. *Cell. Mol. Life Sci.* **2005**, *62*, 848-869.

(8) Gygi, S. P.; Aebersold, R. *Curr. Opin. Chem. Biol.* **2000**, *4*, 489-494.

(9) Strittmatter, E. F.; Ferguson, P. L.; Tang, K.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 980-991.

(10) Patton, W. F.; Schulenberg, B.; Steinberg, T. H. *Curr. Opin. Biotechnol.* **2002**, *13*, 321-328.

(11) Lion, N.; Rohner, T. C.; Dayon, L.; Arnaud, I. L.; Damoc, E.; Youhnovski, N.; Wu, Z.-Y.; Roussel, C.; Josserand, J.; Jensen, H.; Rossier, J. S.; Przybylski, M.; Girault, H. H. *Electrophoresis* **2003**, *24*, 3533-3562.

(12) Sechi, S.; Oda, Y. *Curr. Opin. Chem. Biol.* **2003**, *7*, 70-77.

(13) Peng, J.; Gygi, S. P. *J. Mass Spectrom.* **2001**, *36*, 1083-1091.

(14) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Anal. Chem.* **2001**, *73*, 5683 -5690.

(15) Geng, M.; Ji, J.; Regnier, F. E. *J. Chromatogr. A* **2000**, *870*, 295-313.

(16) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc Natl Acad Sci USA* **2000**, *97*, 9390-9395.

(17) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

(18) Krause, E.; Wenschuh, H.; Jungblut, P. R. *Anal. Chem.* **1999**, *71*, 4160-4165.

(19) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163-170.

(20) Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. *Anal. Biochem.* **2000**, *287*, 110-117.

(21) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Mol Cell Proteomics* **2004**, *3*, 1154-1169.

(22) Goshe, M. B.; Smith, R. D. *Curr. Opin. Biotechnol.* **2003**, *14*, 101-109.

(23) Moritz, B.; E.Meyer, H. *Proetomics* **2003**, *3*, 2208-2220.

(24) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol Cell Proteomics* **2002**, *1*, 376-386.

(25) Hamdan, M.; Righetti, P. G. *Mass Spectrom. Rev.* **2002**, *21*, 287-302.

(26) Aebersold, R.; Goodlett, D. R. *Chemical Reviews* **2001**, *101*, 269-295.

(27) McLafferty, F. W.; Turecek, F. *Interpretation of Mass Spectra*, Fourth ed.; University Science Books, 1993.

(28) Yamaguchi, M.; Nakazawa, T.; Kuyama, H.; Obama, T.; Ando, E.; Okamura; Ueyama, N.; Norioka, S. *Anal. Chem.* **2005**, *77*, 645-651.

(29) Bruni, R.; Gianfranceschi, G.; Koch, G. *J. Peptide Sci.* **2005**, *11*, 225-234.

(30) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255-261.

(31) Wilkins, M. R.; Gasteiger, E.; Gooley, A. A.; Herbert, B. R.; Molloy, M. P.; Binz, P.-A.; Ou, K.; Sanchez, J.-C.; Bairoch, A.; Williams, K. L.; Hochstrasser, D. F. *J. Mol. Biol.* **1999**, *289*, 645-657.

(32) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.

(33) Bader, G. D.; Hogue, C. W. V. *Nat. Biotechnol.* **2002**, *20*, 991-997.

(34) Heeren, R. M. A.; Kleinnijenhuis, A. J.; McDonnell, L. A.; Mize, T. H. *Anal Bioanal Chem* **2004**, *378*, 1048-1058.

(35) Wyttenbach, T.; Bowers, M. T. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 9-14.

(36) Kaltashov, I. A.; Eyles, S. J. *Mass Spectrom. Rev.* **2002**, *21*, 37-71.

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

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

(39) Dole, M.; Hines, R. L.; Mack, L. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *Macromolecules* **1968**, *1*, 96-97.

(40) Hoffman, E. d.; Stroobant, V. *Mass Spectrometry*, Second ed.; John Wiley & Sons, 2002.

(41) Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*; MCC Press, San Diego, 2003.

(42) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362-387.

(43) Cornelis, E. C.; Hop, A.; Bakhtiar, R. *Biospectroscopy* **1997**, *3*, 259-280.

(44) Bonfiglio, R.; King, R. C.; Olah, T. V.; Merkle, K. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1175-1185.

(45) Mallet, C. R.; Lu, Z.; Mazzeo, J. R. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 49-58.

(46) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1-8.

(47) Gershon, D. *Nature Methods* **2005**, *2*, 465.

(48) Quenzer, T. L.; Emmett, M. R.; Hendrickson, C. L.; Kelly, P. H.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 1721-1725.

(49) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299 - 2301.

(50) Burton, R. D.; Watson, C. H.; Eyler, J. R.; Lang, G. L.; Powell, D. H.; Avery, M. Y. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 443-446.

(51) Fitzgerald, M. C.; Parr, G. R.; Smith, L. M. *Analytical Chemistry* **1993**, *65*, 3204-3211.

(52) Nordhoff, E. A. I.; Cramer, R. A. O.; Stahl, B.; Karas, M.; Hillenkamp, F.; Crain, P. F.; Chait, B. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 771-776.

(53) Tang, K.; Taranenko, N. I.; Allman, S. L.; Chen, C. H.; Chág, L. Y.; Jacobson, K. B. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 673-677.

(54) Berkenkamp, S.; Kirpekar, F.; Hillenkamp, F. *Science* **1998**, *281*, 260-262.

(55) Nielen, M. W. F. *Mass Spectrom. Rev.* **1999**, *18*, 309-344.

(56) O'Connor, P. B.; Costello, C. E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1862-1868.

(57) Zaia, J. *Mass Spectrom. Rev.* **2004**, *23*, 161-227.

(58) Baykut, G.; Jertz, R.; Witt, M. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1238-1247.

(59) Laiko, V. V.; Baldwin, M. A.; Burlingame, A. L. *Anal. Chem.* **2000**, *72*, 652-657.

(60) Guerrera, I.; Kleiner, O. *Bioscience Reports* **2005**, *25*, 71-93.

(61) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. *J. Mass Spectrom.* **2005**, *40*, 430-443.

(62) Marshall, A. G. *Int. J. Mass spectrom.* **2000**, *200*, 331-356.

(63) Lawrence, E. O.; Edlefsen, N. E. *Science* **1930**, *72*, 376.

(64) Comisarow, M. B.; Marshall, A. G. *J. Chem. Phys.* **1979**, *71*, 4434.

(65) Amster, I. J. *J. Mass Spectrom.* **1996**, *31*, 1325-1227.

(66) Marshall, A. G.; Hendrickson, C. L. *Int. J. Mass spectrom.* **2002**, *215*, 59-75.

(67) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.

(68) Watson, C. H.; Wronka, J.; Laukien, F. H.; Barshick, C. M.; Eyler, J. R. *Anal. Chem.* **1993**, *65*, 2801 - 2804.

(69) Lowery, K. S.; Griffey, R. H.; Kruppa, G. H.; Speir, J. P.; Hofstadler, S. A. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1957-1961.

(70) Kofel, P.; Allemann, M.; Kellerhals, H.; Wanczek, K. P. *Int. J. Mass Spectrom. Ion Processes* **1985**, *65*, 97-103.

(71) Knobeler, M.; Wanczek, K. P. *Int. J. Mass Spectrom. Ion Processes* **1997**, *163*, 47-68.

(72) Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 296-303.

(73) Hofstadler, S. A.; Sannes-Lowery, K. A.; Griffey, R. H. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 945-951.

(74) Taylor, P. K.; Amster, I. J. *Int. J. Mass spectrom.* **2003**, *222*, 351-361.

(75) Easterling, M. L.; Mize, T. H.; Amster, I. J. *Anal. Chem.* **1999**, *71*, 624-632.

(76) Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. *J. Proteome Res.* **2004**, *3*, 621-626.

(77) Page, J. S.; Bogdanov, B.; Vilkov, A. N.; Prior, D. C.; Buschbach, M. A.; Tang, K.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 244-253.

(78) Belov, M. E.; Zhang, R.; Strittmatter, E. F.; Prior, D. C.; Tang, K.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 4195-4205.

(79) Koomen, J.; Hawke, D.; Kobayashi, R. *Cancer Invest.* **2005**, *23*, 47-59.

(80) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *PNAS* **1999**, *96*, 6591-6596.

(81) Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. *Anal. Chem.* **In press**.

(82) Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Paa-Toli, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 2132-2139.

(83) Mize, T. H. Thesis (Ph. D.), University of Georgia, Athens, 2002.

CHAPTER 2

EXPERIMENTAL

This chapter provides insights into the instrumentation and sample preparation techniques described for the shotgun proteomics experiments performed in Chapters 3 and 4. The research presented in the following chapters share common techniques and instruments including high performance liquid chromatography (HPLC) for separation and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) for mass analysis. This chapter offers a detailed description of the instruments and sample preparation techniques.

INSTRUMENTATION

High Performance Liquid Chromatography

Complex peptide mixtures are separated using reversed phase liquid chromatography by an Ultimate Plus system (Dionex LC Packings, Sunnyvale, CA) (Figure 2.1). The mobile phase provided to the system is filtered (0.45 μm pore size filter) prior to use and is degassed using helium gas. The desired mobile phase composition is achieved by low pressure mixing of solvents followed by the use of high pressure micropump which delivers the mobile phase to a flow splitting unit that provides flow rates of 50 nL/min to 200 μL/min. The gradient is delivered to a reverse phase HPLC column (75 μm-300 μm I.D.), and the eluent is then passed to a UV detector that can simultaneously monitor up to 4 wavelengths in the region of 190-740 nm.

For the experiments reported here, offline separation of the complex proteome digest for MALDI-FTICR analysis is achieved with a reverse phase nano HPLC column (75 μm I.D. x 15 cm) packed with C18 stationary phase. The proteome digest (0.5 μg/mL) is injected (full loop injection) at a flow rate of 30 μL/min using a 20 μL injection loop by the FAMOS well plate

Figure 2.1 A picture of the Unlimate Plus Nano LC system.

microautosampler (Figure 2.2) onto a precolumn. A flow rate of 0.3 μL/min is used to elute the components from the analytical column. A shallow solvent gradient changing from mobile phase A (95/5/0.1% water/acetonitrile/trifluoroacetic acid by volume) to mobile phase B (100% acetonitrile) is used for the 140 minute separation.

The eluted peptides are detected by the UV detector and are then spotted on the MALDI target by use of a Probot micro fraction collector. The short capillary connection between the outlet and the needle of the Probot provides extremely low dwell time, avoiding mixing of separated components. The Probot is a robotic fraction collector designed for off-line collection of fractions eluting from the HPLC system onto a MALDI mass spectrometer target. For the experiments reported here, the elution volume collected per spot was 0.45 μL.

The robotic table is an important component of the Probot unit and can be positioned in the X, Y and Z directions by means of 3 stepper motors. The distance from point to point, the time spent at each point, and the speeds of movement are easily programmed. A fully automated system where the matrix can be added co-axially at the needle tip is also possible. However, for all the work reported in this thesis, the matrix was manually added to the collected fractions after they had dried, prior to MALDI-FTICR-MS analysis. This ensures that the matrix droplet has the same solvent composition for each spot, providing uniform crystal morphology at each spot. The entire separation process is automated with minimal sample handling, thereby decreasing contamination and increasing reproducibility.

For MALDI experiments, the choice of matrix plays an important role. 2,5-dihydroxy benzoic acid (DHB) is a relatively "cool" matrix and is suitable for analysis of peptides and proteins.[1] Typically an excess of the matrix (1:1000 of analyte to matrix) is used for MALDI analysis.[2,3] For a typical MALDI-FTICR analysis, 0.5 μL of the analyte is spotted on the MALDI target (7 X 7 target) and is allowed to dry. 0.5 μL of 2,5- dihydroxy benzoic acid (1 M) solution made in 50/50/ 0.1 % acetonitrile/ water/ trifluoro acetic acid (by volume) is applied to the sample spot and allowed to dry. This step is repeated thus forming fine crystals of the analyte embedded in the matrix.

Figure 2.2 A picture of FAMOS Well Plate Microautosampler.

Intermediate Pressure MALDI-FTMS (9.4 T)

Shotgun proteomics studies utilizing matrix assisted laser desorption ionization (MALDI) as the ionization technique were performed on the 9.4 T BioApex FTICR mass spectrometer with an intermediate pressure Scout100 (IP-MALDI) source (Bruker Daltonics, Billerica, MA).

The IP-MALDI source overcomes a major limitation of conventional vacuum MALDI, namely metastable fragmentation which is result of high internal energies imparted to the ions.[4,5] Such metastable losses yield substantial extra peaks, principally from loss of H_20 or NH_3 from the molecular species. IP-MALDI reduces the abundance of these losses to a negligible level. These metastable losses are mitigated by the use of collisional cooling at intermediate pressure.[5,6] Figure 2.3 shows a schematic of the intermediate pressure MALDI source that utilizes argon at 10^{-3} Torr for collisional cooling and a 337 nm UV Laser for ionization. A typical MALDI experimental sequence starts with the opening of the pulsed valve, allowing the collision gas (argon) to be pulsed in the vicinity of the target (for collisional cooling), thereby increasing the pressure of the source to 10^{-3} Torr. An X-Y manipulator positions the target aligning the analyte spot with the laser beam. After the pulsed addition of the collision gas, the UV laser is triggered, and it emits a 100 μJ pulse of ~10 ns duration. The laser output is directed onto the analyte spot, and focused to a size of ~0.5 mm in diameter which translates to approximately 10 MW/cm^2 irradiance. The pulsed valve closes after the firing of the laser and after a short delay (for pumping down); the ions are sent to the hexapole for accumulation. Ions undergo collisional thermalization during their desorption. They are trapped in a hexapole ion guide, operating in rf-only mode. After accumulation of ions produced by up to 10 laser shots, the voltage is reduced on the extraction plate of the hexapole unit, thus releasing the ions. They are guided to the analyzer cell by means of electrostatic ion transfer optics (Figure 2.4).[7] The ions are then trapped in the analyzer cell using Sidekick™, a process where a small voltage is applied to a pair of electrodes at the entrance of the analyzer cell. This voltage pushes the ions off axis, slightly expanding their magnetron radius.[8] This process increases trapping efficiency and thus enhances detection.

ESI-Q-FTMS (7.0 T)

Experiments involving use of electrospray ionization for proteomics analysis as described in Chapter 3 were performed on a 7.0 T Apex-Qe FTMS (Bruker Daltonics, Billerica, MA) with an Apollo-II electrospray ionixation source. This hybrid Q-FTMS combines the quadrupole mass analyzer collision cell, and FTICR mass analyzer into a single instrument. Apart from in-cell accumulation and dissociation, it also allows for precursor selection and collisional activated dissociation (CAD) in a hexapole collision cell that lies outside the magnet.[9] Figure 2.5 shows the quadrupole interface with the FTICR mass spectrometer, which consists of a quadrupole mass filter (which be used as an ion guide when operated under rf only mode), a hexapole collision cell, and associated lenses between the ion source and the standard ion transfer optics. The quadrupole mass filter consists of four precision rods that are aligned parallel with one another. It operates with rf and dc potentials applied to rods that permit selection of ions of a narrow range of mass-to-charge (1-10 m/z units). The selected ions can then be dissociated by CAD in the hexapole which is operated in the rf-only mode. The hexapole collision cell operating in an rf-only mode acts as an ion guide and not a mass filter. This helps to keep ions on axis while they undergo collisions. For a tandem mass spectrometry experiment, the ions are allowed to undergo energetic collisions with the collision gas (argon), thereby dissociating the precursor ion and forming product ions. The product ions are then cooled and trapped by further collisions with the gas. The collision cell is used to trap and store ions before transferring them to the analyzer cell. Alternatively, the quadrupole can be used in the rf-only mode to pass all ions without mass selection, and the potentials of the hexapole can be set to prevent colliosional dissociation.

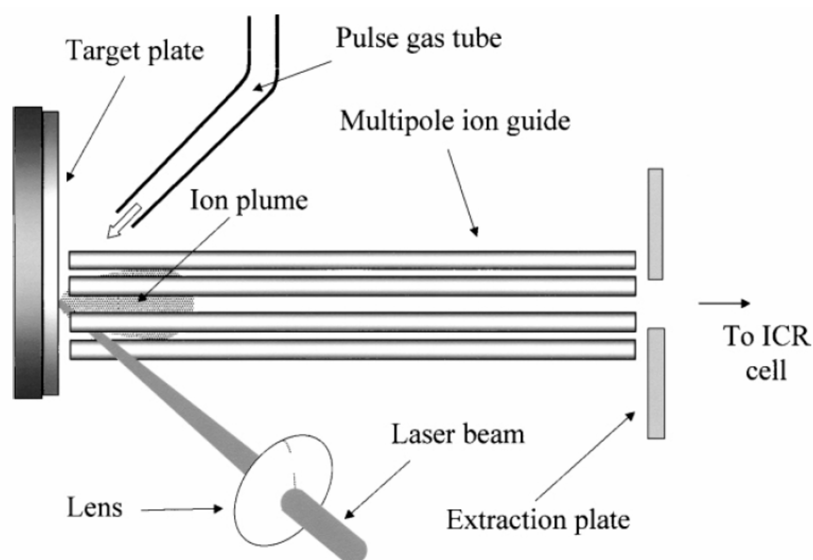


Figure 2.3 A schematic of Intermediate pressure MALDI source.[6] Pressure in the hexapole region during desorption is approximately 10^{-3} Torr.

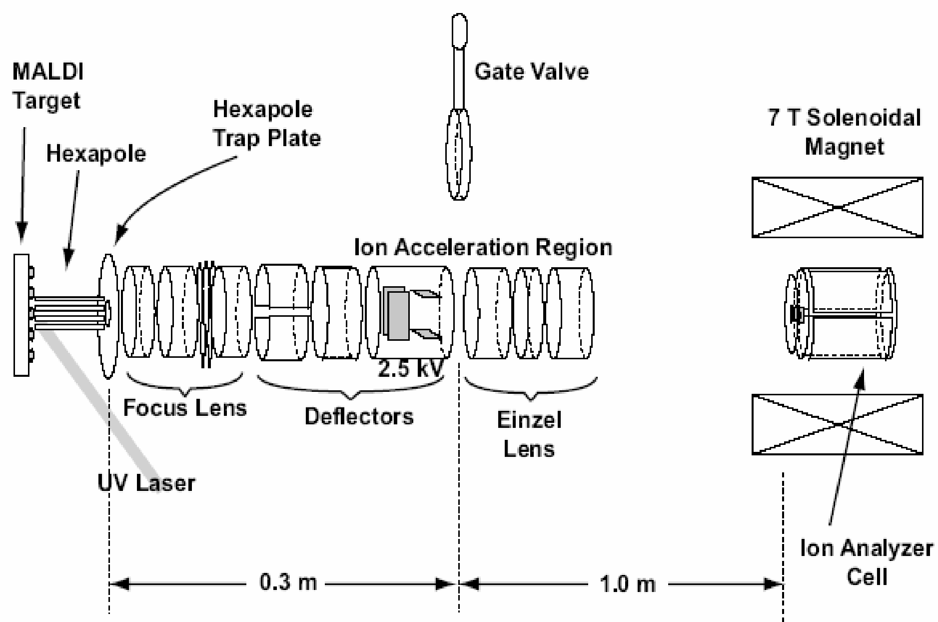


Figure 2.4 A schematic showing the electrostatic ion transfer optics including Focus lens, deflectors and Einzel lens.[7]

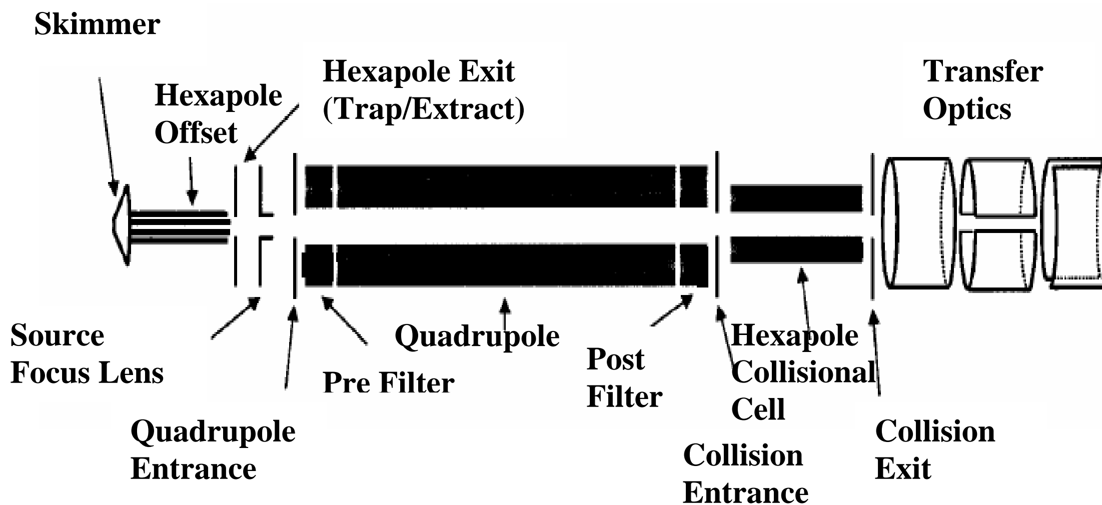


Figure 2.5 A schematic of the Quadrupole interface along with the ion optics.

An Agilent 1100 series HPLC system used in conjunction with Q-FTMS allows online capillary LC/MS and LC/MS/MS studies. The separation is performed on a reverse phase capillary HPLC column (300 μm I.D. x 15 cm) with C18 stationary (Dionex LC Packings, Sunnyvale, CA) phase with a flow rate of 0.4 mL/min. A manual injection of the analyte is performed using a 1 μL injection loop. The 60 minute gradient used for the online separation is shallow and runs from 5% mobile phase A (100% water, 0.1% formic acid) to 50% mobile phase B (100% acetonitrile, 0.1% formic acid). The gradient is then held at a 50% organic for 10 minutes to ensure the elution of all the components and then is changed accordingly for cleaning and conditioning the column.

SAMPLE PREPARATION

Shotgun proteomics requires batch digestion of all the proteins resulting in a highly complex mixture of peptides. Any developmental strategy such as chemical labeling of amino acids etc. needs to be first tested on peptide and protein standards before experimenting on a real biological sample. The choice of peptide and protein standards for initial experimentation is thus important. For example, for labeling the lysine residue and N-terminus (as described in this Chapter 3) substance P, Arg-Pro-Lys-Pro-Glu-Phe-Gly-Leu-Met-NH_2, is used as a peptide standard as it has 1 lysine residue and 1 N-terminal amino group available for reaction. After a reaction protocol is developed for the peptide, the reaction is then tested on protein standards such as bovine serum albumin (BSA), ovalbumin, and ubiquitin and then finally on a real biological sample (whole cell lysate). All the protein and peptide solutions are made at 1 mg/mL in 50 mM ammonium bicarbonate or phosphate buffer (pH 8). It is important to consider the stability of the protein in the buffer and also its compatibility with mass spectrometry analysis.

Labeling and Digestion Procedures

Unlike protein standards, proteins extracted from a biological source often contain contaminants from the biological matrix and therefore need to be purified prior to analysis. Gel permeation chromatography (GPC) is used for sample purification. In GPC, a protein solution is passed through a column bed of an inert and microporous stationary phase. For these experiments, Sephadex G-25 (Aldrich, St.Louis, MO) is used in a 3 mL spin column. Small molecules penetrate into the pores of the stationary phase and are retained on the column. Large molecules such as proteins, pass directly through the column. Once the proteome is free of low mass contaminants, denaturation and reduction of disulfide bonds is performed. Proteins are tightly folded and need to be denatured for effective digestion. Also, the disulfide linkages

between cysteine residues result in a more tightly bound protein. Denaturation is commonly achieved by heating to 95 ^{0}C. For some experiments, guanidine hydrochloride was used as a chaotropic agent to assist denaturation.[10,11] Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) is used to reduce disulfide linkages. Guanidine hydrochloride is removed prior to digestion to avoid interference during mass analysis.

To mimic an ArgC digest as described in Chapter 3, the *M. maripaludis* proteome was denatured using 6 M guanidine hydrochloride, 14 µL of 50 mM tris (2-carboxyethyl) phosphine hydrochloride, and heat (at 95 ^{0}C) for 15 minutes. Upon cooling, 250 µL of the proteome solution was reacted with iodoacetamide to alkylate the cysteine residues. 30 µL of a 50 mM solution of iodoacetamide dissolved in methanol was added to the proteome sample and was incubated for 90 minutes at room temperature in the dark. Acetylation of the free amines was then carried out using N-acetoxysuccinimide at pH 9 for 5 hours. N-acetoxysuccinimde was dissolved in water at a concentration of 18 mg/mL. The reagent was added in two steps. 40 µL of reagent solution was added at the beginning of the reaction and an additional 40 µL of the reagent solution was added after 30 minutes of the reaction. Treatment with excess N-hydroxylamine (10 µL) for 10 minutes was followed. The residual reagent was removed by gel permeation chromatography prior to enzymatic digestion. Trypsin is the most commonly used protease for enzymatic digetsion, as it is highly specific and less expensive.[12] The modified grade of trypsin, in which the lysines are dimethylated (Promega, Madison, WI) was used in order to reduce autolysis. The digestion protocol involved a 50:1 (w/w) ratio of substrate to enzyme. 50 µL of trypsin solution prepared at 40 µg/mL, using 20 mM phosphate buffer was added to the proteome sample and incubated for 18 hours at 37 ^{0}C. After digestion, the sample was refrigerated at 0 ^{0}C to prevent further digestion or degradation.

The following procedure was used for initial studies in isolating the N-terminal peptides as described in Chapter 3. 1 mg/mL solutions of the peptide standards, substance P, Arg-Pro-Lys-Pro-Glu-Phe-Gly-Leu-Met-NH_2, and angiotensin II, β-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe, were prepared using 0.1 M phosphate buffer and 0.1 M mixture of ammonium bicarbonate and dimethylformamide (50/50 by volume) respectively. 100 μL of substance P peptide was acetylated with a 50 fold excess of N-acetoxysuccinimide at pH 9 for 5 hours. The reagent was dissolved in water at a concentration of 18 mg/mL and was added in parts in the first hour. The first part comprising of 13 μL of reagent solution was added at the beginning of the reaction and the second part comprising of 13 μL of the same was added after 30 minutes of the reaction. Treatment with excess N-hydroxylamine (5 μL) for 10 minutes was then followed. A mixture of pure substance P and acetylated substance P (50/50 by volume) was prepared at 0.1 mg/mL at pH 8. Internal peptides were removed from the solution using the Reacti-bind maleic anhydride microplate wells (Pierce, Rockford, IL.). Substance P mixture, at 100 μg/mL was used for reaction in 10 different wells at 37 ^{0}C with a one hour reaction time in each well. At one hour intervals the sample was transferred by a pipette to a fresh well. After 10 hours of reaction (10 wells, one hour per well), the sample mixture was analyzed by mass spectrometry. For mass defect labeling of internal peptides, the reagent N-2,5-dibromobenzoyloxysuccinimide that was synthesized, and tested on a peptide standard. 100 μL angiotensin II peptide was reacted with 30 μL of the mass defect label which was dissolved in dimethylformamide to make a saturated solution. After 24 hours of reaction the peptide was analyzed by mass spectrometry.

References

(1) Cornelis, E. C.; Hop, A.; Bakhtiar, R. *Biospectroscopy* **1997**, *3*, 259-280.

(2) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299 - 2301.

(3) Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*; MCC Press, San Diego, 2003.

(4) Zaia, J. *Mass Spectrom. Rev.* **2004**, *23*, 161-227.

(5) O'Connor, P. B.; Costello, C. E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1862-1868.

(6) Baykut, G.; Jertz, R.; Witt, M. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1238-1247.

(7) Wong, R. L.; Amster, I. J. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 205-212.

(8) Hofstadler, S. A.; Sannes-Lowery, K. A.; Griffey, R. H. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 945-951.

(9) Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. *J. Proteome Res.* **2004**, *3*, 621-626.

(10) Creighton, T. E. *2nd edition, W.H.Freeman and Company* **1993**.

(11) Görg, A.; Weiss, W.; Dunn, M. J. *Proetomics* **2005**, *5*, 826-827.

(12) Olsen, J. V.; Ong, S.-E.; Mann, M. *Mol Cell Proteomics* **2004**, *3*, 608-614.

CHAPTER 3

MIMICKING AN ARGC DIGESTION BY LYSINE ACETYLATION AND TRYPSINOLYSIS FOR IMPROVED SHOTGUN PROTEOMICS BY ACCURATE MASS MEASUREMENT

INTRODUCTION

Protein identification is usually the rate determining step in any proteomic strategy and many of the advances in the proteomics field are focused on improving the speed and sensitivity with which proteins are identified.[1] This chapter describes a new shotgun proteomics strategy to improve protein identification specificity and the overall speed of analysis when using accurate mass measurement for identification. The method utilizes lysine acetylation and trypsin digestion to produce peptides that are similar to those produced by ArgC digestion. The developed technique is tested on a proteome from *Methanococcus maripaludis*, which is an archaeal methanogen. Methanogens are ubiquitous microorganisms that catalyze the terminal step in the anaerobic food chain by reducing simple compounds to methane which is a useful energy source and a powerful greenhouse gas.[2] *M. maripaludis* is a methane producing archaeon, strictly anerobic and like all methanogens, belongs to the *Euryarchaeota* kingdom in the *Archaea* domain. It is a mesophilic hydrogenotropic methanogen that utilizes H_2 as an electron donor to reduce CO_2 to methane.[2,3] *M. maripaludis* is an ideal model organism to study archeal biology and is distinguished amongst all the methanogens by its rapid growth.[4] Successful implementation of the shotgun proteomic strategy described in this chapter on *M. maripaludis* would allow high throughput proteomics studies of real biological samples.

In general, any proteomics assay that utilizes peptide mass fingerprinting most commonly uses trypsin as a protease to digest the protein. Trypsin has high cleavage specificity, and is stable under a wide variety of conditions.[5] Trypsin cleaves C-terminal to the lysine and arginine residues of the protein and thereby produces peptides. However, proteolytic cleavage of all

proteins with trypsin presents a highly complex mixture.[5] Lysine and arginine represent 10% of all amino acids in a protein, and so trypsin cleavage produces peptides that are on an average approximately 10 amino acids in length.

In silico tryptic digestion of the 1722 proteins from the *M. maripaludis* genome, will generate peptides with a mass distribution as shown in Figure 3.1. The X-axis coordinate represents the peptide molecular weight and the Y-axis coordinate represents the number of peptides that are formed (bin size-25 amu). The plot shows that majority of the peptides that result from tryptic digestion have molecular weight less than 1000 Da. Figure 3.2 shows the number of proteins from *M. maripaludis* that can be correlated to specific tryptic peptides measured at a mass accuracy of 1, 10 and, 100 ppm with and without nitrogen stoichiometry as an identification constraint. The plot shows that as the mass accuracy increases, the identification specificity increases significantly. However, even at 1 ppm mass tolerance, especially at 1000 Da, additional constraints for database search are required. Research in our laboratory has shown that at 10 ppm mass tolerance, and using nitrogen stoichiometry as a search constraint, increased peptide specificity is observed.[6,7] Metabolic labeling using ^{15}N labeling not only yields nitrogen stoichiometry, but also allows quantitative studies.[8,9] Two pools of the organism under study are grown on a medium that contains natural abundance of nitrogen and also on a medium that is enriched with ^{15}N. Both the pools are combined and digested once using trypsin and the resulting peptides are separated using liquid chromatography. Pairs of peptides elute and the spacing between them depends on the number of nitrogen atoms present in the respective peptide. The peptide pairs have identical sequence but differ in mass, as one of them is labeled with heavy nitrogen. 43% of peptides from *M. maripaludis* can be identified using 10 ppm mass accuracy and nitrogen constraint versus about 15% using only accurate mass measurement. According

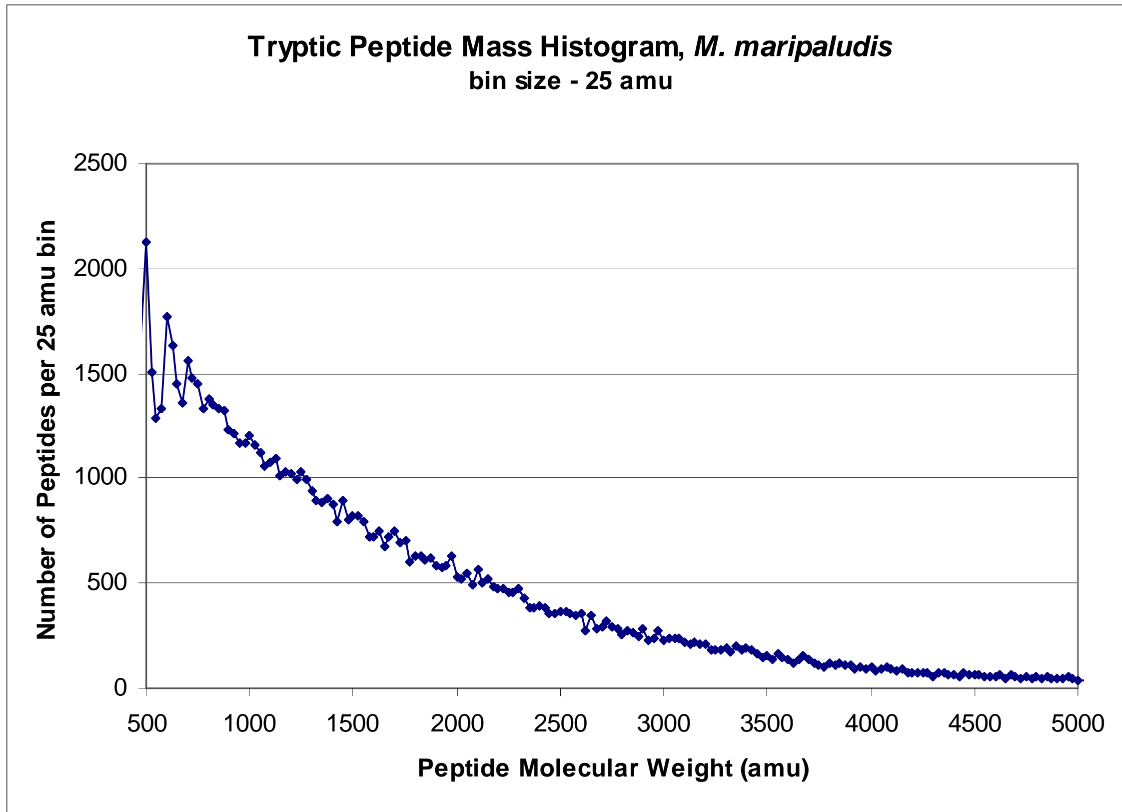


Figure 3.1 A histogram showing the number of peptides, along with their molecular weight that can be formed by tryptic digestion of all the proteins from *M. maripaludis*, with a bin size of 25 amu.

to the graph, as the molecular weight increases, the peptide specificity increases significantly, showing that larger peptides contribute more as unique peptides. The goal of the shotgun proteomics approach described in this chapter is to increase this fraction of unique peptides in order to improve protein identification.

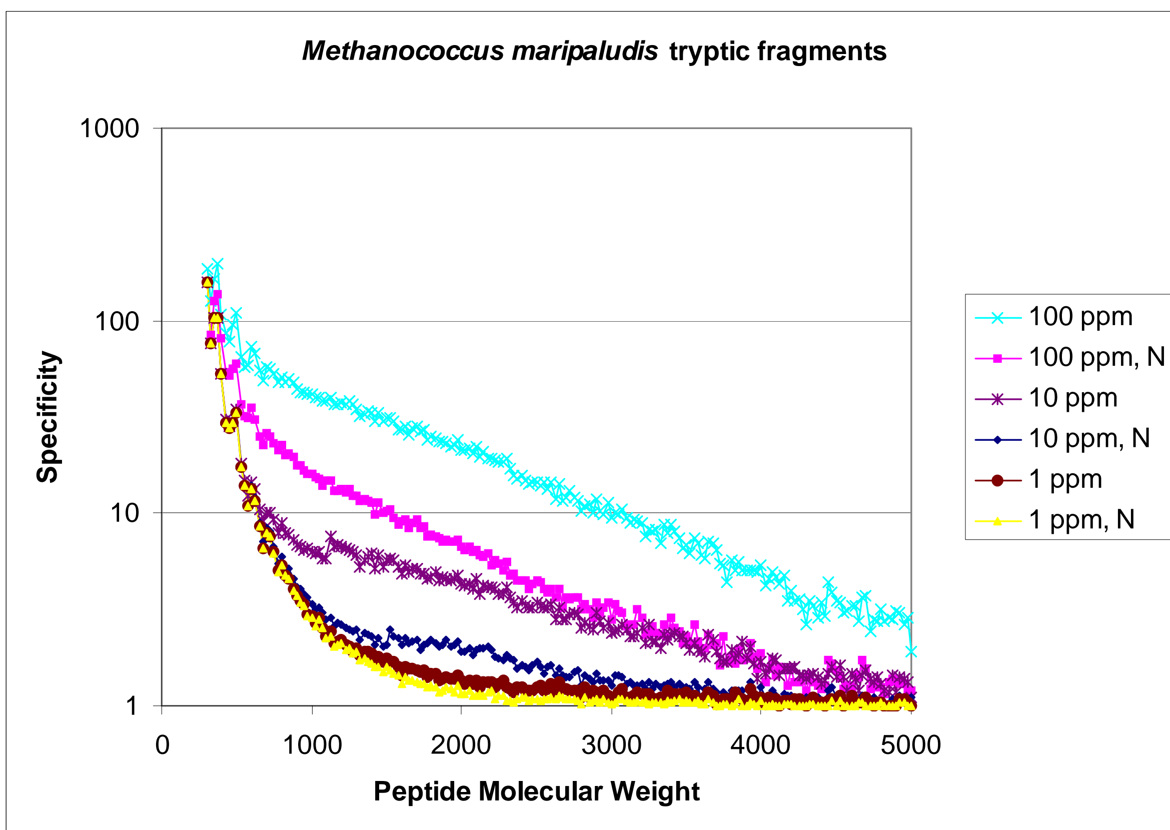


Figure 3.2 A plot of search specificity for *M. maripaludis* proteins identified from its tryptic peptides measured at 1, 10 and 100 ppm mass accuracy with and without nitrogen stoichiometry.

The proteolytic enzyme ArgC cleaves amide bonds on the C-terminal side of the arginine residues. *In silico* digestion performed on the *M. maripaludis* proteome, considering ArgC as a protease, will generate peptides with a mass distribution that is very different compared to trypsin. Figure 3.3 shows the mass distribution that results from *in silico* digestion using trypsin and ArgC as a protease. The graph shows how peptide identification specificity by accurate mass

measurement (10 ppm mass tolerance) varies as a function of molecular weight for the methanogen. The dark bars represent a histogram of all the peptides that are predicted to be formed with up to 1 missed cleavage. The number of peptides that can be assigned to a single protein using accurate mass measurement, constrained by the number of nitrogen atoms within the peptides are shown by the light bars, and constitute 89% (75% for peptides below m/z 4000) of all peptides, when using ArgC as a protease and 43% of all peptides, when using trypsin as a protease. The high number of unique peptides generated by ArgC digestion yields higher sequence coverage and improves protein identification compared to trypsin digestion. Another advantage of using ArgC digestion is that fewer peptides are generated compared to tryptic digestion, thus resulting in less complex batch digestion mixtures, enabling faster and easier analysis. However, ArgC has two drawbacks in its application for this purpose; it is expensive and it lacks the specificity of trypsin. ArgC preferentially cleaves after arginine residues, but it also exhibits a small degree of trypsin-like behavior (cleaving after lysine). Thus an economical and practical way to carry out an ArgC-like digest would be to modify the lysine residues so that tryptic digestion cleaves only after arginine. In order to mimic an ArgC digest by the technique discussed above, a derivatizing agent that is highly specific for lysine is required.

Lysine Derivatization

The side chain of the lysine residue is a linear chain of four methylene groups terminated by an amino group with an intrinsic pKa value of 11.1, and therefore it is ionized under physiological conditions. The ε-amino group of lysine readily undergoes a variety of reactions that include acylation, alkylation, arylation and amidination reactions.[10]

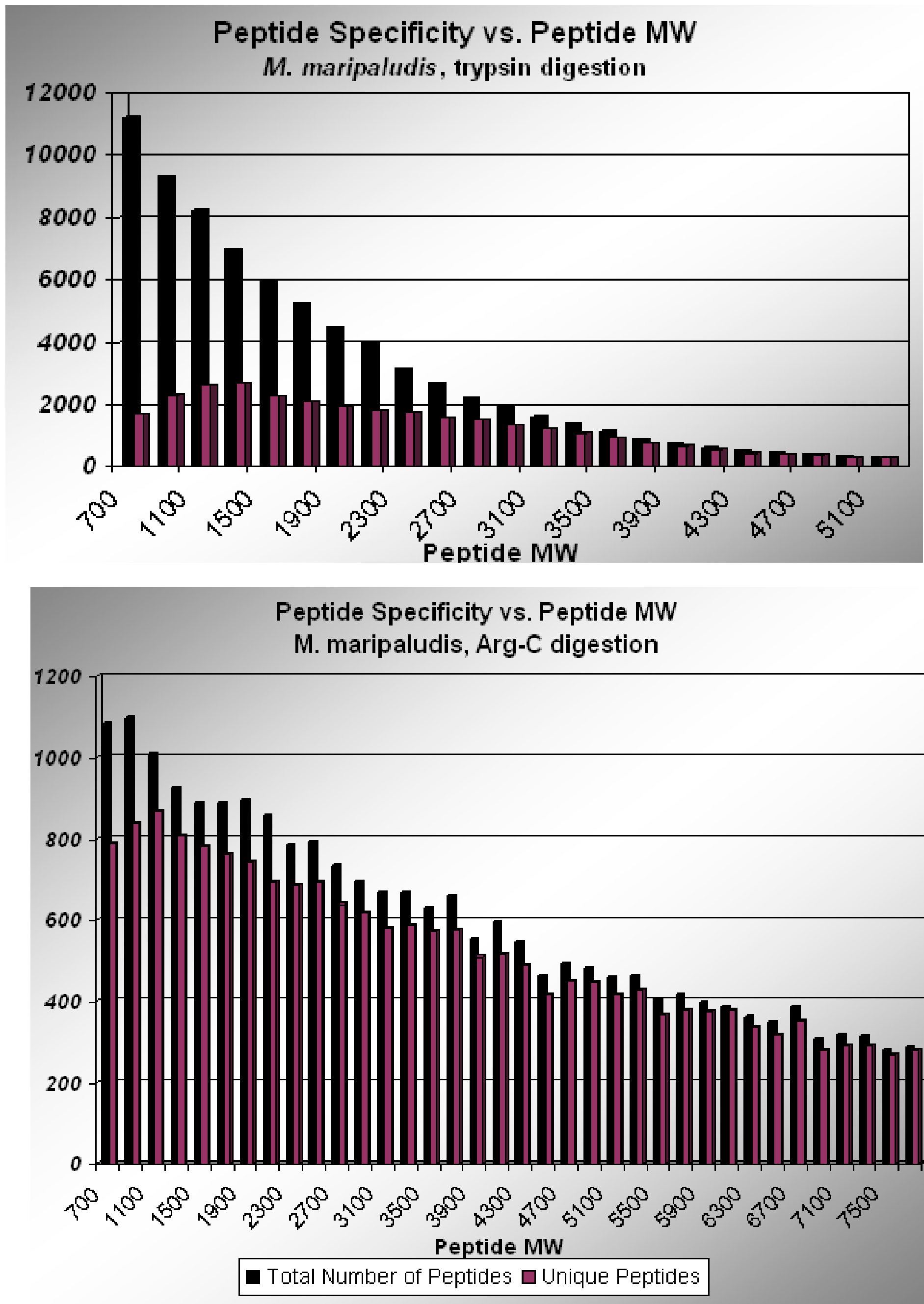


Figure 3.3 Results of *in silico* tryptic and ArgC digest of the *M. maripaludis* proteome. Histograms with 200 amu bins, where the dark bars represent a histogram of all the peptides that are predicted to be formed with up to 1 missed cleavage. The lighter bars are the histogram of the number of peptides that can be identified uniquely by accurate mass measurement, constrained by the number of nitrogen atoms within the peptides.

Recent studies indicate that MALDI is more efficient at ionizing peptides that contain arginine residues versus those that contain lysine, with the arginine-containing peptides being detected with 4-18 fold more intensity than that of the lysine containing peptides.[11] This increase in intensity of the arginine-containing peptides was attributed to the greater condensed phase basicity and/or gas-phase basicity of the arginine-containing peptides. One result of this bias is the relatively low percentage of sequence coverage that is achieved through database searching since only half of peptides generated by tryptic digestion are intense enough to be used with confidence. Researchers have developed a derivatization procedure to facilitate *de novo* sequencing of the lysine terminated peptides through the guanidination of the ε-amino group of the lysine by using O-methylisourea.[12-14] Thus this reaction converts the lysine residues into more basic homoarginine, thereby dramatically increasing the sensitivity of detection of these peptides by MALDI-TOF mass spectrometry. However, this approach of converting lysine residues to homoarginine residues cannot be used in this research, as trypsin will cleave after the arginine and homoarginine residues and will eventually yield a tryptic digest instead of the desired ArgC-like digest. Therefore a specific derivatizing agent that blocks trypsinolysis at lysine residues needs to be used. Also, any reagent that reacts readily with the lysine residue will inevitably react with the N-terminus of the protein. Thus, the database search also needs to account for the modification of the N-terminal amino groups of all the proteins secreted from *M. maripaludis*.

Different reagents are claimed to be lysine specific including S-ethyltrifluoro thioacetate,[15-20] citraconic anhydride,[21-24] 2,4,6-trinitrobenzene 1-sulfonic acid,[25] formaldehyde,[26] and N-acetoxysuccinimde.[27-32] All the above reagents have been tested in the work reported here, and N-acetoxysuccinimide was selected for lysine derivatization as it is water soluble

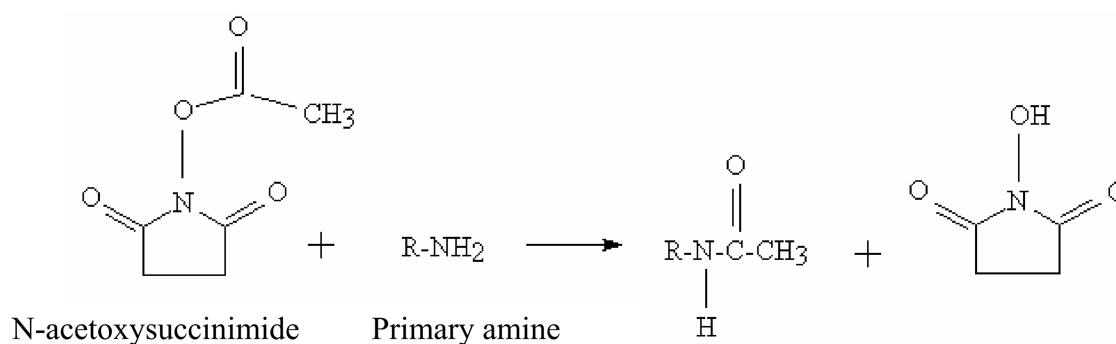


Figure 3.4 Reaction of N-acetoxysuccinimide with a primary amine.

(hence protein friendly), reacts readily and is highly specific to lysine and N-terminal amino groups.[33] This reagent acetylates primary amines as shown in Figure 3.4. The reaction adds 42.0107 Da mass for every primary amine present in a protein or peptide. Reaction conditions reported in the literature suggest basic conditions (~ pH of 9) and a 50 fold excess of the reagent with 5 hours of reaction time.[30] There is a possibility of acetate esters being formed, but treatment with excess hydroxylamine causes aminolysis of acetate esters within minutes.[30,33,34] In this study, use of N-acetoxysuccinimde as an acetylating agent for the lysine residues and the N-terminus of the proteins is first tested on the peptide standards substance P and angiotensin II, and the intact protein, ubiquitin, with 7 lysine residues. Additionally, a large protein, bovine serum albumin, is tested for the generation of ArgC-like peptides. The results obtained from the above experiments allow the application of this technique to a proteome of *M. maripaludis,* to test this method as a shotgun proteomics approach where endogenous ^{15}N labeling, ArgC-like peptide generation, and accurate mass measurement are used to improve peptide identification specficity.

SAMPLE PREPARATION

The peptide standards, substance P and angiotensin II are dissolved in 50 mM phosphate buffer at a concentration of 1 mg/mL. 100 µL of each peptide solution is allowed to react with a 50 fold excess of N-acetoxysuccinimide at pH 9 for 5 hours. The reagent is dissolved in water at a concentration of 18 mg/mL and is added in two parts in the first hour. The first aliquot of 13 µL of reagent solution is added at the beginning of the reaction and the second aliquot comprising of 13 µL of the same solution, is added after 30 minutes after initiation of the reaction. Treatment with excess hydroxylamine (5 µL) for 10 minutes then follows.

1 mg/mL of the protein standards, ubiquitin and bovine serum albumin, are prepared in 50 mM phosphate buffer. The general protocol involves heat denaturation with 6 M guanidine hydrochloride and 12 µL of 50 mM tris(2-carboxyethyl)phosphine hydrochloride (to reduce the disulfide bonds) for 15 minutes at 95 ^{0}C. Upon cooling, 250 µL of the protein solution is allowed to react with iodoacetamide to alkylate the cysteine residues to prevent formation of disulfide bonds. A 50 fold excess of a 50 mM solution of iodoacetamide dissolved in methanol is added to the 250 µL of 1 mg/mL protein sample and is incubated for 90 minutes at room temperature in the dark. Acetylation of the free amines is carried out as described above. Treatment with excess hydroxylamine (5 µL), is followed by removal of residual reagents by gel permeation chromatography in a 3 mL spin column. For enzymatic digestion, modified trypsin is used in order to prevent autolysis. 50 µL of trypsin prepared at 40 µg/mL, using 20 mM phosphate buffer is added to about 250 µL of protein solution. The solution is thoroughly mixed and then incubated for 18 hours at 37 ^{0}C.

M. maripaludis proteins were harvested by the research group of Dr.William B. Whitman (Department of Microbiology, University of Georgia). Wild type *M. maripaludis* was grown to

mid-logarithmic and stationary stages using ammonium sulfate as the sole source of nitrogen with naturally occurring isotopic composition and with 98% ^{15}N-enriched composition respectively. Equal amounts of protein extracts from the two growth stages were mixed prior to denaturing the proteome. For denaturation, 6 M guanidine hydrochloride and 14 µL of 50 mM tris(2-carboxyethyl)phosphine hydrochloride is added to the protein solution, followed by heating for 15 minutes at 95 ^{0}C. Upon cooling, 250 µL of the proteome solution is allowed to react with iodoacetamide. 30 µL of a 50 mM solution of iodoacetamide, dissolved in methanol is added to the proteome sample. The solution is incubated for 90 minutes at room temperature in the dark. Acetylation of the free amines is then carried out using N-acetoxysuccinimide which is dissolved in water at a concentration of 18 mg/mL. The reagent is added in two steps; 40 µL of reagent solution is added at the beginning of the reaction and an additional 40 µL of the reagent solution is added 30 minutes after the onset of the reaction. After the acetylation reaction is completed, excess hydroxylamine (10 µL) is added and allowed to react for 10 minutes. The residual reagent is removed by gel permeation chromatography in a spin column prior to enzymatic digestion. 50 µL of trypsin prepared at 40 µg/mL, using 20 mM phosphate buffer is added to proteome sample and incubated for 18 hours at 37 ^{0}C.

Instrumentation

The labeled peptides and protein digests were analyzed on the 9.4 T BioApex FTICR mass spectrometer with an intermediate pressure Scout100 MALDI source (Bruker Daltonics, Billerica, MA). The whole protein analysis for ubiquitin was performed in the reflectron mode on the Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics, Billerica, MA) of the Chemical and Biological Sciences Mass Spectrometry Facility (CBSMSF, Department of Chemistry, University of Georgia).

For offline LC/MS studies, chromatography was performed with an Ultimate Plus HPLC system (Dionex LC Packings, Sunnyvale, CA). 20 µL of 0.5 mg/mL of the *M. maripaludis* proteome digest was injected using the autosampler on the reverse phase nano C18 column (75 µm I.D. x 15 cm). Over the course of the 140 minute gradient, 80 fractions were collected onto two stainless steel MALDI targets using the Probot micro-fraction collector. The matrix was added manually to the dried analyte. The fractions were then analyzed using a 9.4 T FTICR mass spectrometer with an intermediate pressure MALDI source. 12 scans were acquired, with 2 laser shots per scan. The instrument was calibrated using ~ 400 nL of a (1 mg/mL) tryptic digest of bovine serum albumin prior to mass analysis. The calibrant peaks were fitted with a calibration equation (Equation 3.1), where A and B are the calibration constants and f_{obs} is the observed frequency.

$$\frac{m}{z} = \frac{A}{B + f_{obs}} \qquad \text{Equation 3.1}$$

For online LC/MS/MS studies, 1 µL of 1 mg/mL of the *M. maripaludis* proteome digest was manually injected onto the reverse phase capillary C18 column (300 µm I.D. x 15 cm) of an Agilent 1100 series HPLC system hyphenated to the 7.0 T Apex-Qe FTMS (Bruker Daltonics, Billerica, MA). The peptide fractions were analyzed online over a 55 minute duration. A spectrum was acquired every 1 second and three of the most abundant peptides from each scan were selected for tandem mass analysis using collisional activated dissociation by argon gas. The

instrument was calibrated using 1 μg/mL peptide standard solution mixture of angiotensin I and B-chain.

RESULTS AND DISCUSSION

Labeling Peptides and Proteins using N-acetoxy succinimide

To test the chemistry for mimicking an ArgC-like digest, the reaction scheme (acetylation using N-acetoxysuccinimide) was first tested on peptide and protein standards. Substance P, Arg-Pro-Lys-Pro-Glu-Phe-Gly-Leu-Met-NH_2, was the choice for testing the acetylation step as it has two potential sites for reactivity namely, the lysine residue and the N-terminus. Figure 3.5 shows the MALDI-FTICR mass spectrum of the peptide before and after acetylation using N-acetoxysuccinimide. The mass difference of 84 Da between the control and the reacted peptide shows that the reagent quantitatively reacted with both the lysine residue and the N-terminus. Angiotensin II, β-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe, has only one site for reactivity, namely the N-terminus. Figure 3.6 shows the MALDI-FTICR mass spectrum for the acetylated angiotensin II before and after addition of the N-hydroxylamine. The mass difference of 84 Da between the control and the labeled peptide shows that two residues were acetylated. Treatment with excess hydroxylamine for 10 minutes resulted in a mass spectrum with mass difference of 42 Da between the control and the labeled peptide suggesting that only the N-terminus underwent complete acetylation. The tyrosine residue probably underwent acetylation, and treatment with hydroxylamine aminolyzed the ester. This suggests that treatment with hydroxylamine is an important step in the reaction protocol.

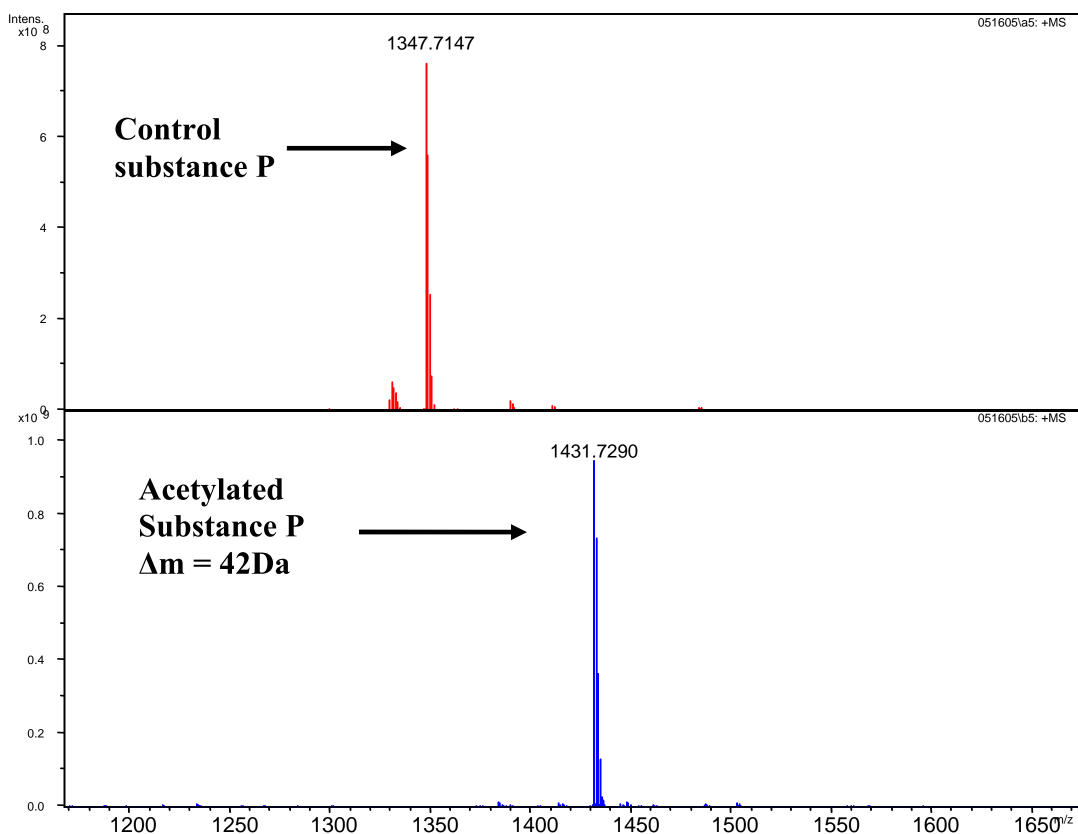


Figure 3.5 Substance P before and after treatment with N-acetoxysuccinimide. The mass shift of 42 Da shows that the reagent reacted quantitatively with the lysine residue and N-terminus.

Ubiquitin, a relatively small protein (8 kDa), was chosen for validating the acetylation reaction on a whole protein. It has 7 lysine residues, no cysteine residues, and 1 N-terminal amino group and thus has 8 possible sites for acetylation. Figure 3.7 shows the mass spectrum of the underivatized and the derivatized protein. The whole protein was analyzed using MALDI-TOF mass spectrometry. The mass difference of 336 Da between the control and the labeled peaks suggests that all possible 8 sites were acetylated.

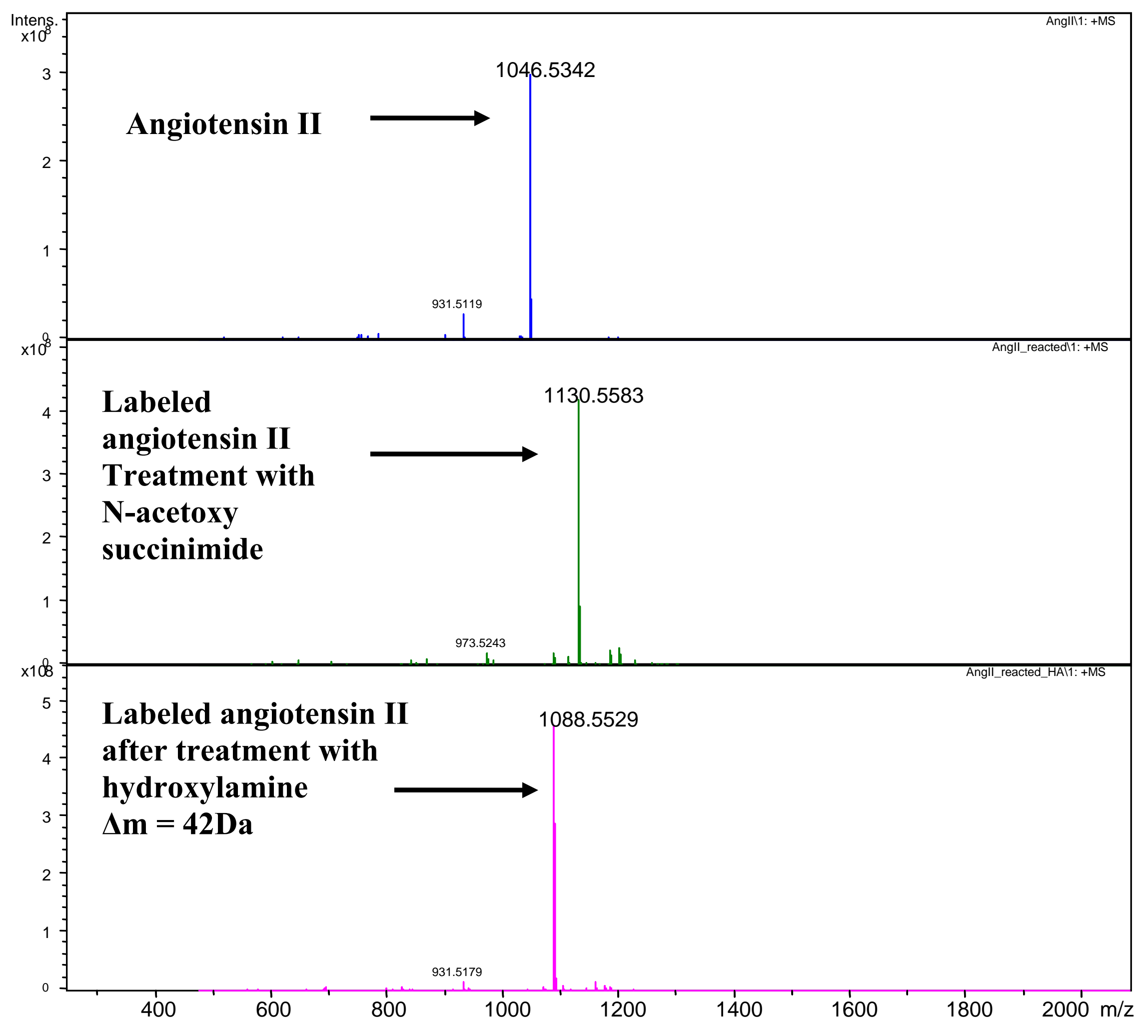


Figure 3.6 Acetylation of angiotensin II using N-acetoxysuccinimide followed by treatment with hydroxylamine for hydrolysis of acetate esters. The mass difference of 42 Da in the resultant mass spectrum shows that only the N-terminus of the peptide underwent labeling.

A larger protein, bovine serum albumin (66 kDa) was then used to test the procedure for generating ArgC-like peptides. Figure 3.8a shows the MALDI-FTICR mass spectrum of the tryptic and ArgC-like digests of bovine serum albumin (BSA). All of the intense peaks

corresponding to the abundant peptides observed in the ArgC-like digest of BSA corresponded to those of the expected ArgC-like peptides. Figure 3.8b shows the protein coverage obtained from tryptic and ArgC-like peptides. This result demonstrated successful ArgC-like digestion, implying that the experimental methods were suitable for implementation on a real biological sample.

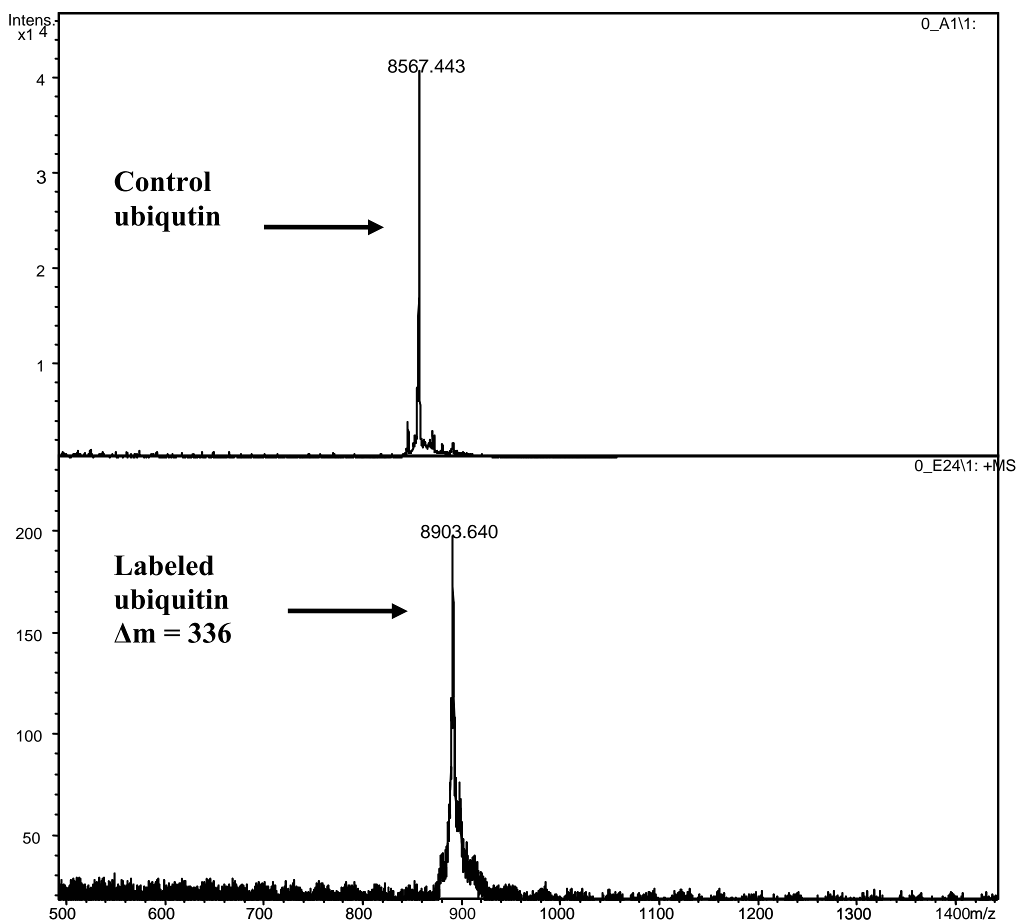


Figure 3.7 Acetylation of whole protein ubiquitin by N-acetoxysuccinimide. The mass difference of 336 Da shows that all the seven lysine residues and the N-terminus were acetylated quantitatively.

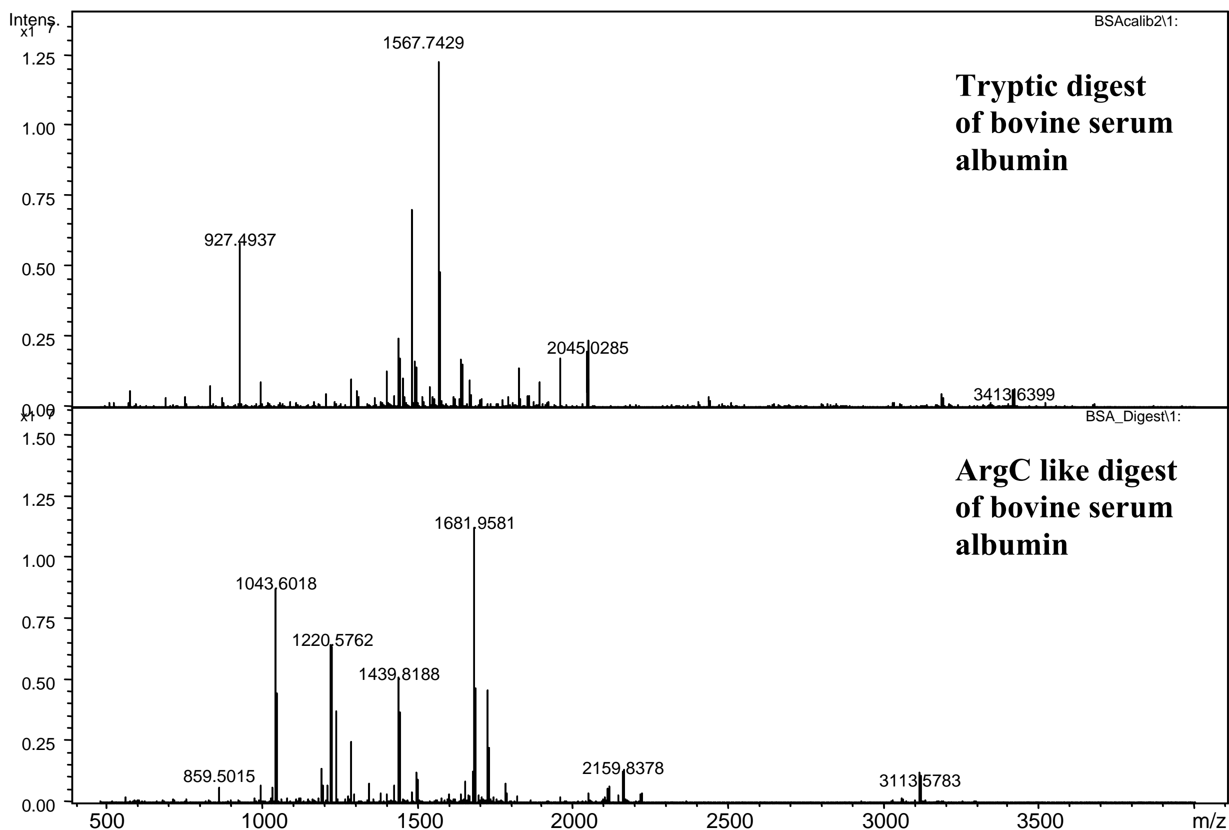


Figure 3.8a Mass spectra and protein coverage map for Tryptic and ArgC like digests of bovine serum albumin. Tryptic peptides are red in color and ArgC peptides are underlined in blue.

Alternative Amino Acid Derivatization Chemistry

Other reagents were also tested on peptide and protein standards for chemically labeling of lysine residues. S-ethyl trifluorothioacetate reacts with primary amines under basic conditions causing a mass shift of 96 Da (Figure 3.9). The reagent was tested on ubiqutin. The general

Sequence coverage of BSA using tryptic peptides

MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF
KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CQAEDKGAC
LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK YQEAKDAFL
GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
STQTALA

Sequence coverage of BSA using ArgC peptides

MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF
KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CQAEDKGAC
LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK YQEAKDAFL
GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
STQTALA

Figure 3.8b The sequence coverage for bovine serum albumin using tryptic and ArgC-like peptides.

protocol involved heat denaturation of the protein using 6 M guanidine hydrochloride for 15 minutes at 95 ^{0}C. Upon cooling, the chaotropic agent was removed by gel permeation chromatography in a 3 mL spin column. A 1000 fold excess of the reagent dissolved in dioxane was allowed to react with the protein for 2 hours at pH 8 using triethylamine as a base. After 2 hours, the residual reagent was removed by gel permeation chromatography in a spin column and

the protein was analyzed using MALDI-TOF mass spectrometry. Figure 3.10 shows the mass spectrum of the whole protein before and after reaction with S-ethyl trifluorothioacetate. Ubiquitin, as discussed previously, has 8 possible sites for modification. As observed in the mass spectrum, only 5 out of the 8 sites underwent the modification. The incomplete reaction with ubiquitin, the pungent odour of the reagent, and the requirement of an organic medium in the reaction mixture mitigated against further application of this reaction chemistry.

2,4,6-trinitrobenzenesulfonic acid reacts completely with the primary amines under basic conditions (pH 9) producing a mass shift of 210 Da (Figure 3.11). The resultant labeled peptides are hydrophobic in nature. The reaction was tested on substance P, a peptide standard which has one lysine residue and one N-terminus available for reaction. 125 μL of 1 mg/mL of the peptide solution was prepared using 0.1 M sodium bicarbonate solution. 30 μL of the reagent made at 0.01 % (weight /volume) in sodium bicarbonate buffer was added to the peptide solution and allowed to react for 2 hours at 37 ^{0}C. Figure 3.12 shows the mass spectrum of substance P

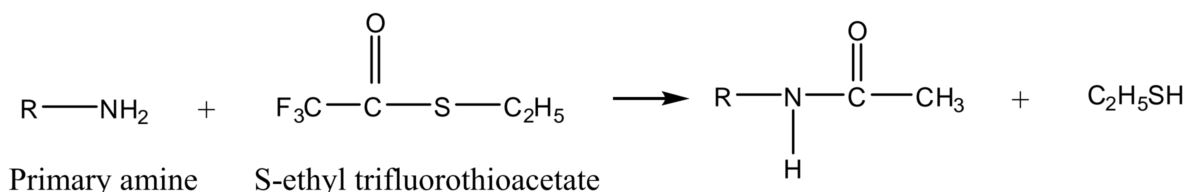


Figure 3.9 Reaction of S-ethyl trifluorothioacetate with a primary amine.

before and after reaction with the reagent. The mass spectrum shows that although substance P was completely modified, the label underwent photo-decomposition when analyzed by MALDI. This reagent is therefore not useful for analysis by MALDI mass spectrometry.

Reductive methylation of primary amines using formaldehyde as an alkylating agent in presence of a reducing agent (dimethylamine borane complex) is used to alkylate lysine residues (Figure 3.13). The reaction adds 2 methyl groups to primary amines, producing a mass shift of 28 Da. The reaction is performed on the shaker in the dark at 4 ^{0}C. 10 µL of a 1 M solution of the reducing agent was added to 150 µL of peptide solution, followed by addition of 20 µL of (1 M) formaldehyde. The mixture was then allowed to react for 2 hours at 4 ^{0}C. The same step was repeated and then 5 µL of ammonium bicarbonate was added and the reaction was carried out for 24 hours under the same reaction conditions. After 24 hours, 50 µL of (1 M) ammonium sulfate was added and the reaction mixture was incubated for 1 hour. Figure 3.14 shows the mass spectrum of substance P before and after reaction with formaldehyde. The mass spectrum shows that the reagent quantitatively reacted with substance P. However, the total time of reaction was close to 30 hours.

After having examined all these reagents for reactivity with primary amines, N-acetoxysuccinimide appeared to be the best choice. It is protein friendly (water soluble) and reacts rapidly and completely with primary amines. N-acetoxysuccinimide was therefore used in chemical labeling of the primary amines for all successive experiments.

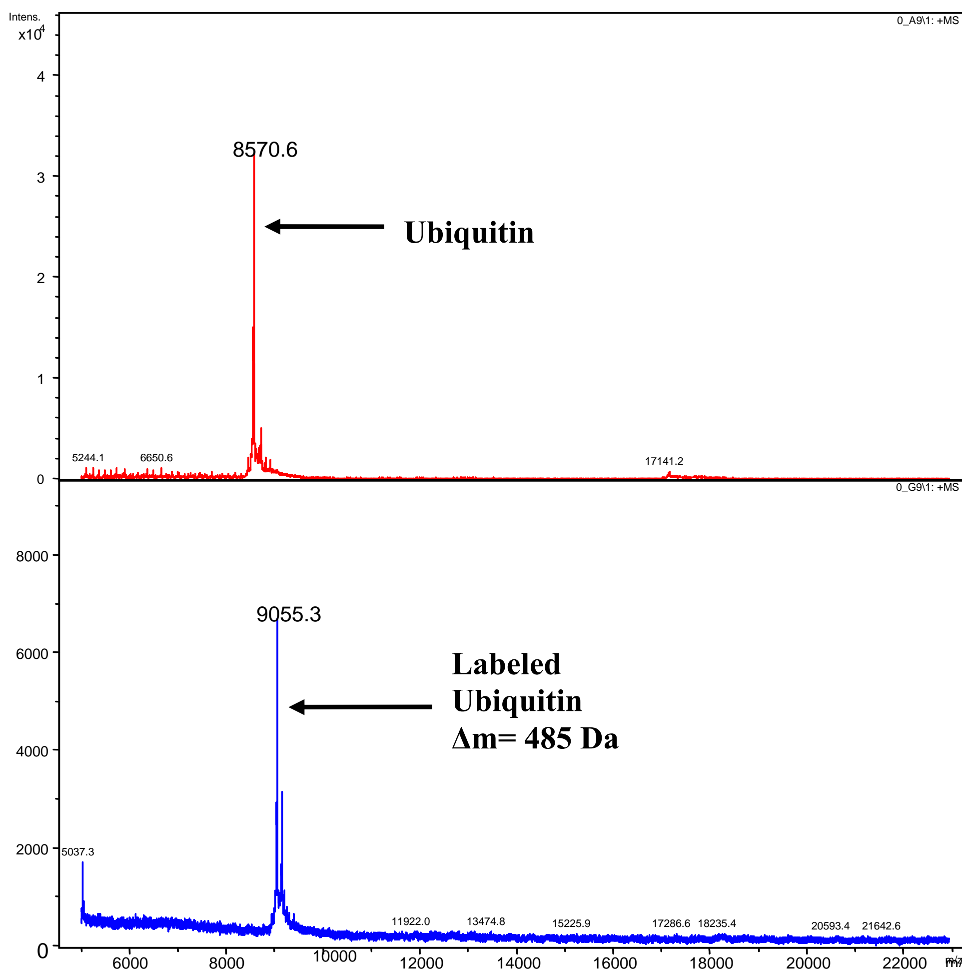
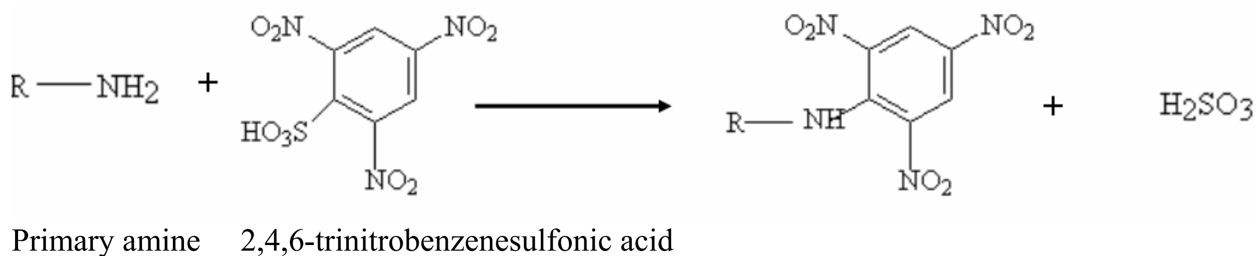


Figure 3.10 Reaction of S-ethyl trifluorothioacetate with ubiquitin analyzed using MALDI-TOF mass spectrometry. The mass difference of 485 suggests that almost 5 sites out of the possible 8 sites underwent modification implying that the reaction was incomplete.



Primary amine 2,4,6-trinitrobenzenesulfonic acid

Figure 3.11 Reaction of primary amine with 2,4,6-trinitrobenzenesulfonic acid.

Proteome analysis using LC-MALDI-FTICR MS

Batch digestion of the entire proteome from *M. maripaludis* generates a large number of peptides which are separated offline by liquid chromatography and then analyzed by MALDI-FTICR. Since the proteome is metabolically labeled, peptide pairs of identical sequence but different masses co-elute and the difference in the spacing between the pairs yields the number of nitrogen atoms present in the respective peptide. Figure 3.15 shows the two dimensional plot of the entire liquid chromatography (LC) run during which the analyte fractions were collected and analyzed. The horizontal axis corresponds to the mass to charge ratio of the peptides. The vertical axis corresponds to the elution time. Each spot indicates a peptide at a particular m/z value and relative elution time. As the elution time increases, the organic component in the mobile phase increases, thereby eluting hydrophobic peptides. The plot indicates that higher mass peptides are observed later in the LC run.

Manual peak picking was required to identify labeled/unlabeled peptide pairs. An in-house search program called Shotpep was used for database searching. 1053 pairs of peptides were submitted for the database search and 46 % of the peptides were uniquely identified (at a mass tolerance of 10 ppm) that contributed to 163 proteins. 13% of the peptides were non-unique and contributed to the identification of more than one protein. Figure 3.16 shows the resultant protein coverage map for *M. maripaludis*. Each of the 1722 squares represents a single open reading frame presented in the order in which they occur in the genome. The red and purple squares represent the 163 proteins for which at least one unique peptide was observed. The purple squares correspond to proteins with a low percentage of sequence coverage and the red squares correspond to proteins with a high percentage of sequence coverage. The green and yellow squares correspond to the proteins for which only non-unique peptides were observed.

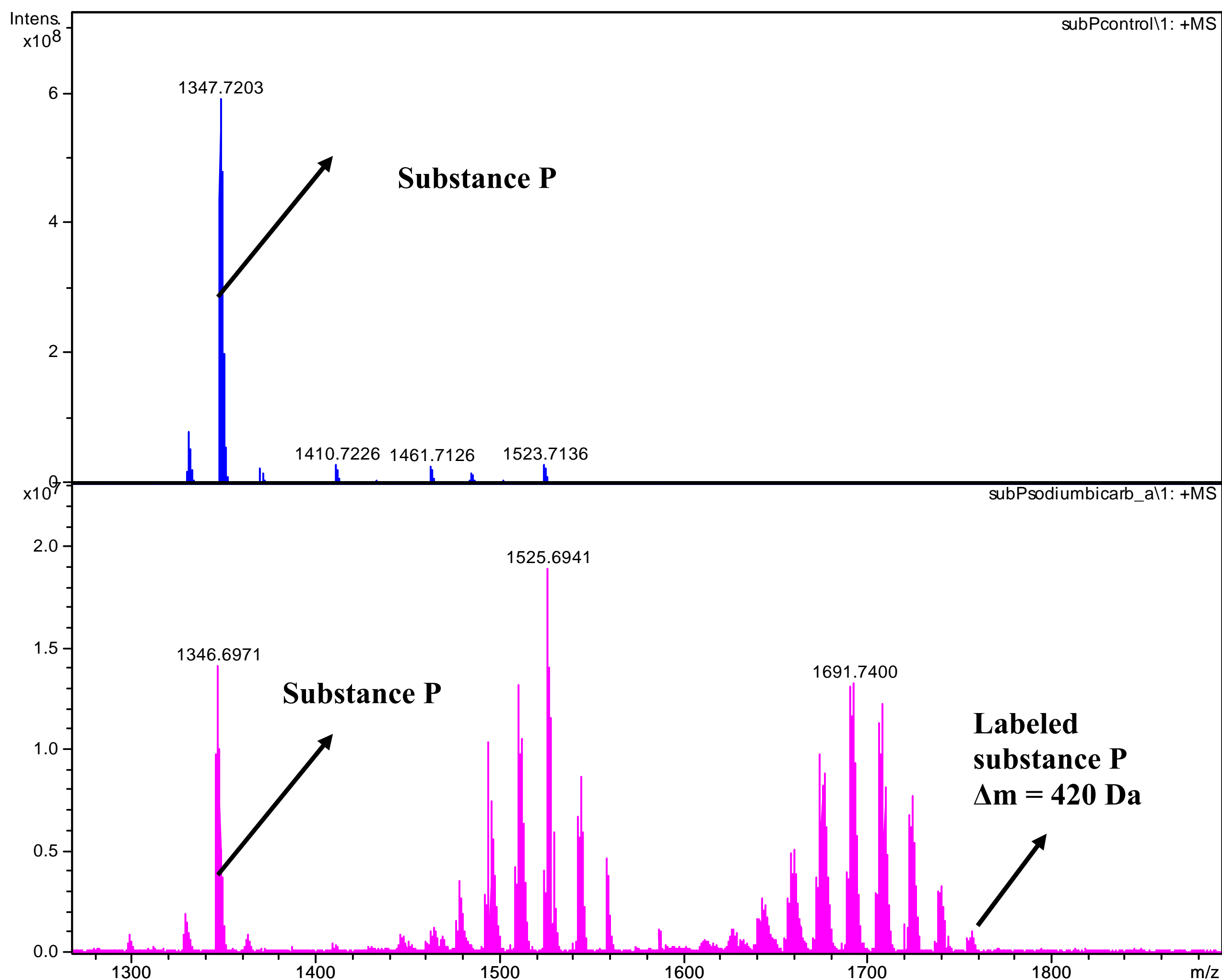
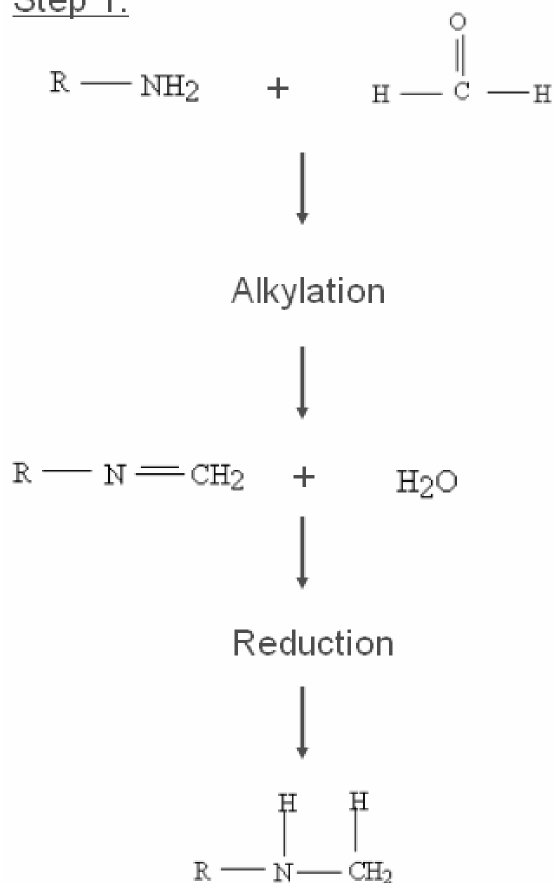


Figure 3.12 Reaction of substance P peptide with 2,4,6-trinitrobenzenesulfonic acid. The mass difference of 420 Da shows that both the lysine residue and the N-terminus of the peptide were labeled. The spectrum shows that even though the reaction was complete, the labeled peptide underwent photoactive decomposition during mass analysis under MALDI conditions.



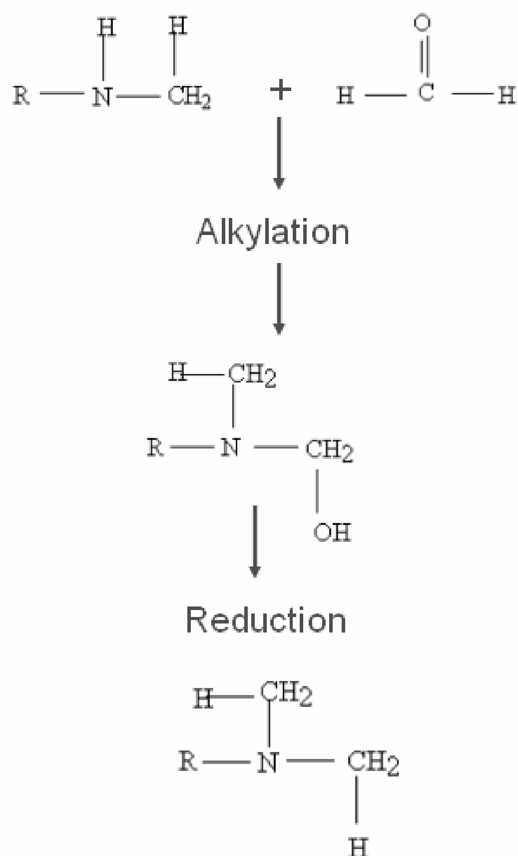


Figure 3.13 Reductive methylation of primary amines where formaldehyde is the alkylating agent.

The black squares represent proteins for which no peptides were observed. According to the results, 9.5% of all potential proteins from *M. maripaludis* were uniquely identified using nitrogen stoichiometry, accurate mass measurement (at 10 ppm) and ArgC digestion to improve identification specificity.

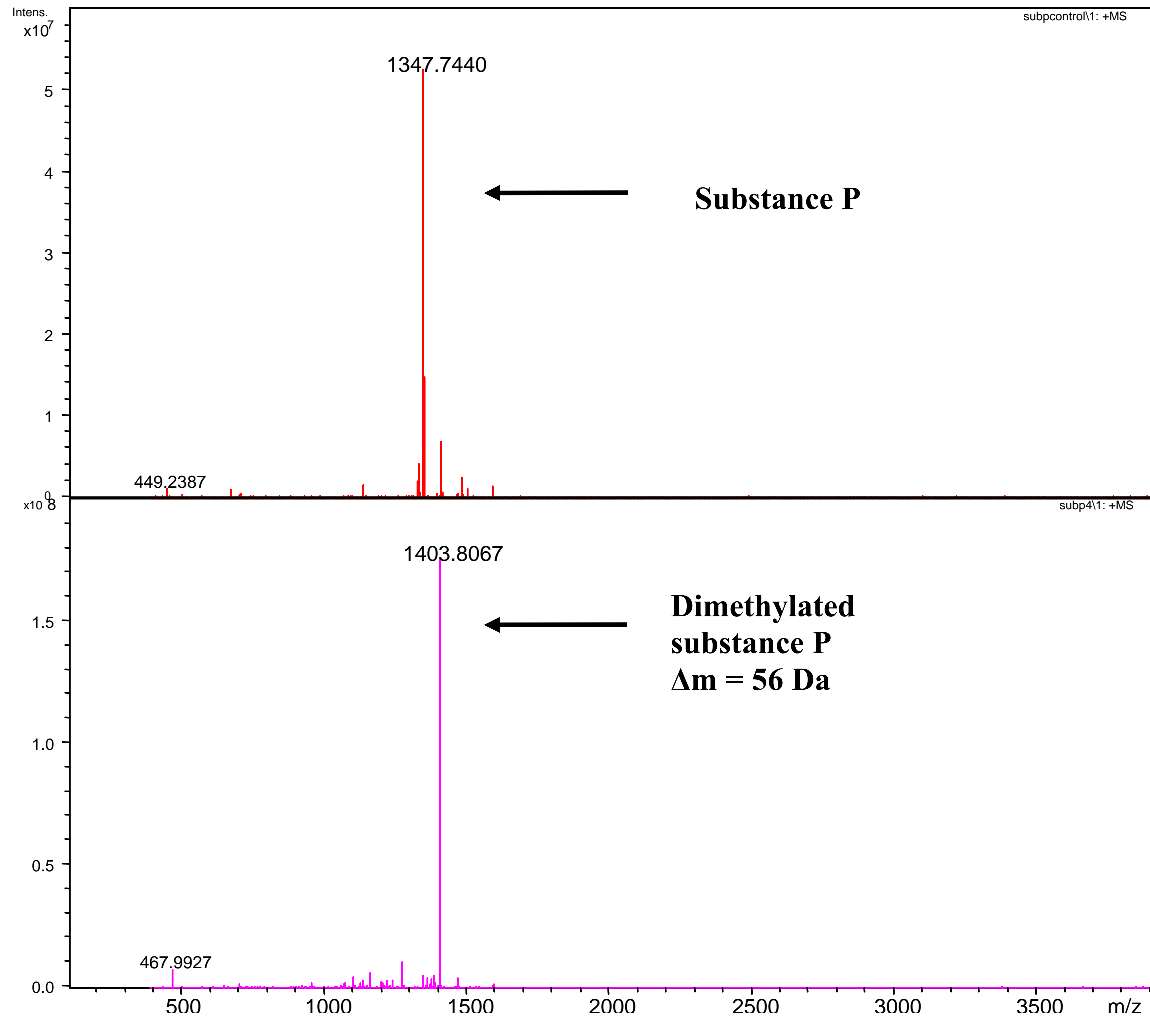


Figure 3.14 Reductive methylation of substance P using formaldehyde as an alkylating agent. The mass spectrum shows that substance P was entirely dimethylated. The mass difference of 56 Da shows that both the lysine residue and the N-terminus of the protein underwent complete dimethylation.

Proteome analysis using LC-ESI-FTICR MS

Online separation by liquid chromatography followed by mass spectrometry and by tandem mass spectrometry was performed on the complex mixture of peptides generated by

batch digestion of *M. maripaludis* proteome using ESI-FTICR-MS for mass analysis. Figure 3.17 shows the total ion chromatogram for the LC/MS run and the LC/MS/MS run. 1118 peptide pairs from the LC/MS run were submitted for database searching and 48% of these were uniquely identified (at a mass tolerance of 10 ppm) leading to the identification of 166 proteins. 11% of the peptides were non-unique and contributed to the identification of more than one protein.

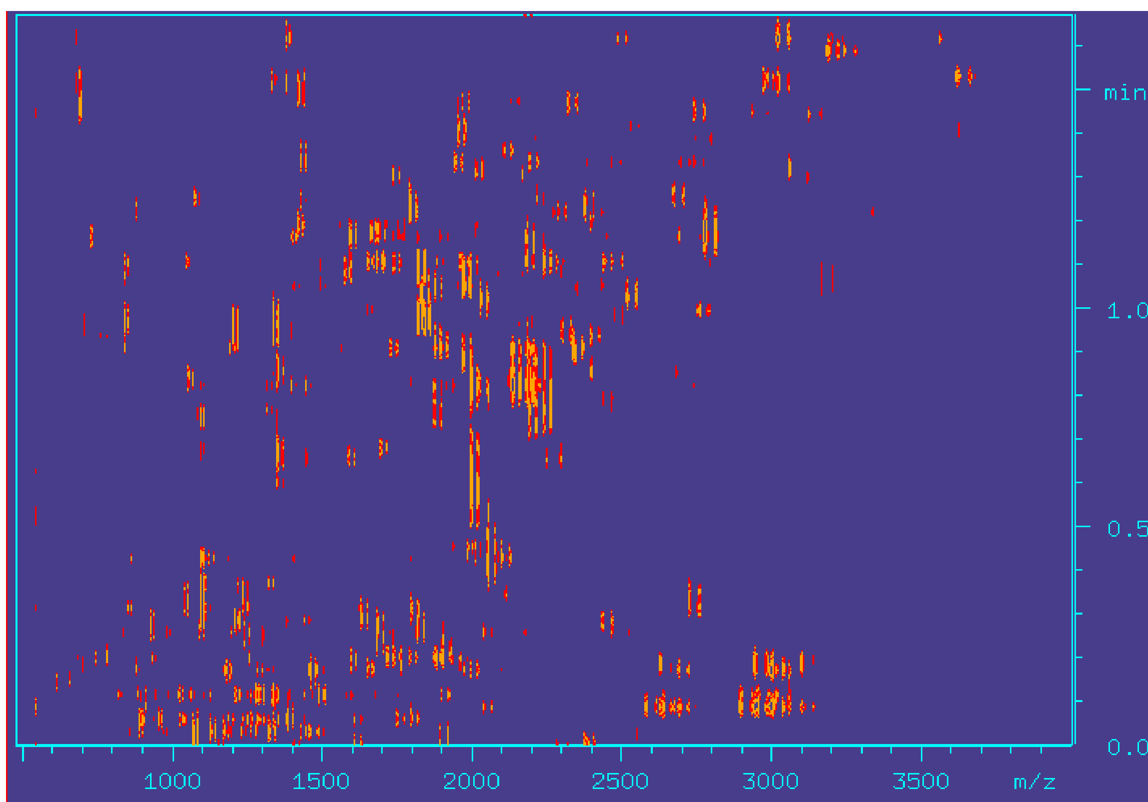


Figure 3.15 A two dimensional plot of the entire LC run of the *M. mariplaudis* proteome digest during which analyte fractions were collected and analyzed. The contrast was adjusted to show only the most abundant peptides.

Figure 3.18 shows the resultant protein coverage map for *M. maripaludis* by ESI studies. According to the results, about 10% of proteins from *M. maripaludis* were uniquely identified using number of nitrogens, accurate mass measurement (at 10 ppm) and unique peptide generation technique as constraints.

In both the MALDI and ESI studies, some of the proteins that were identified had high sequence coverage (up to 98%). However, the 10% of proteins identified by the individual techniques present a very low number. This low number of identified proteins observed in both MALDI and ESI studies is due to the presence of unidentified peptides (40%) that do not correspond to any entries in the database. These peptides are of very low abundance (less than 2% relative intensity) and fall below the threshold for tandem mass analysis in the LC/MS/MS experiment. Tandem mass analysis was performed on the proteome sample using LC-ESI-FTICR-MS and about 700 peptides were selected for tandem mass analysis. Only two of the 700 peptides corresponded to the "unidentifiable" peptides. Since the intensity of both precursor ions was low, the resultant tandem mass spectra were very weak and thus not informative. Ignoring these unidentifiable peptides, about 78% of peptides were unique indicating that higher peptide specificity was observed by using peptides generated by ArgC-like digestion.

An interesting observation was identification of 4 histone proteins from *M. mariplaudis* with very high sequence coverage. The ε-amino groups of lysine residues in histones frequently undergo post-translational modification, primarily acetylation,[35] and were previously not accounted for in trypsin digestion experiments, where the search assumes no post-translational modification. However, in this experiment, the lysine residues and the N-terminal amino groups are acetylated prior to enzymatic digestion, and peptides that are already acetylated *in vivo* will appear in the list of possible peptide masses, allowing identification of acetylated histones.

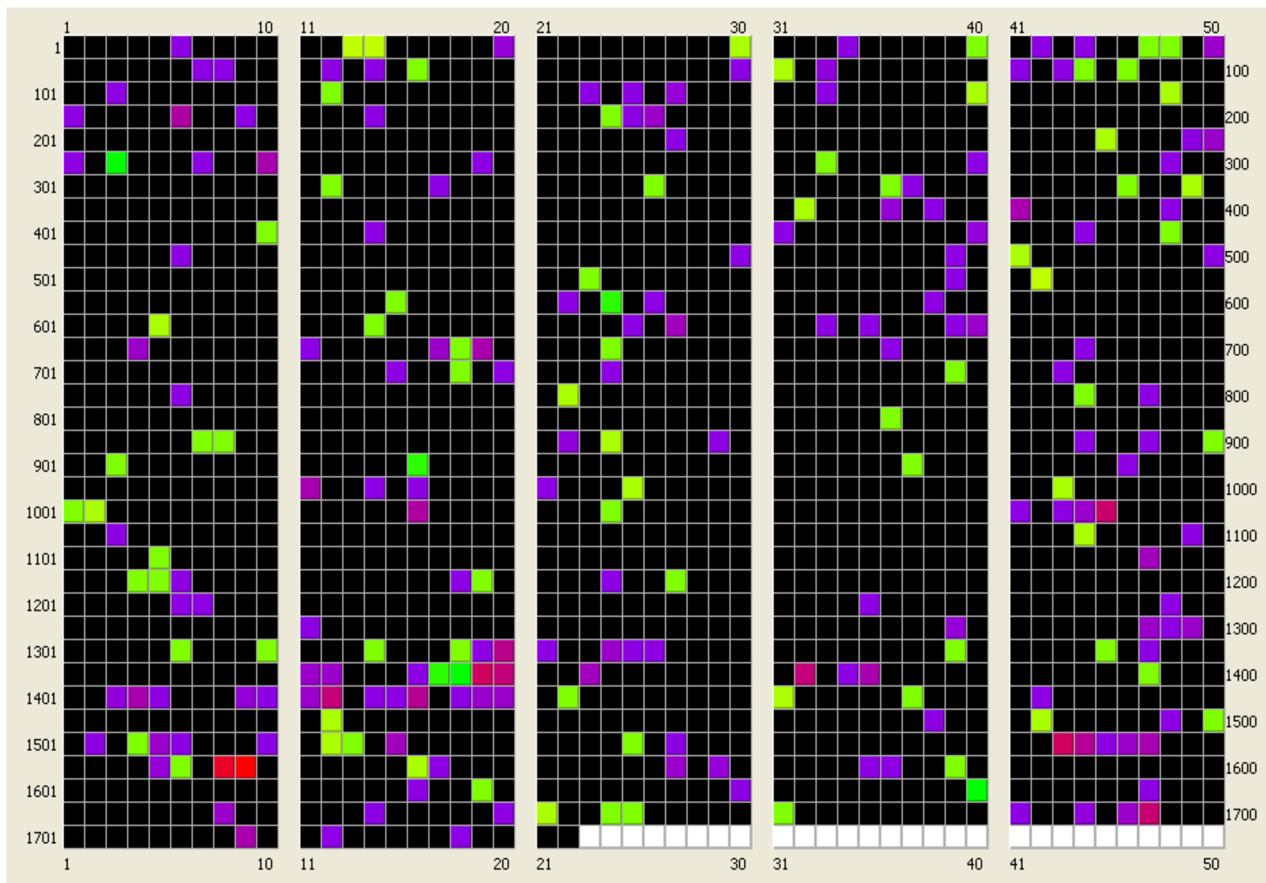


Figure 3.16 Proteome coverage map for *M. maripaludis* resulting for MALDI-MS results where all squares represent 1722 ORFs from *M. maripaludis*. Red and purple squares contribute to themost abundant proteins. Green and yellow squares represent proteins for which non- unique peptides were observed. 163 proteins were identified at 10 ppm mass tolerance.

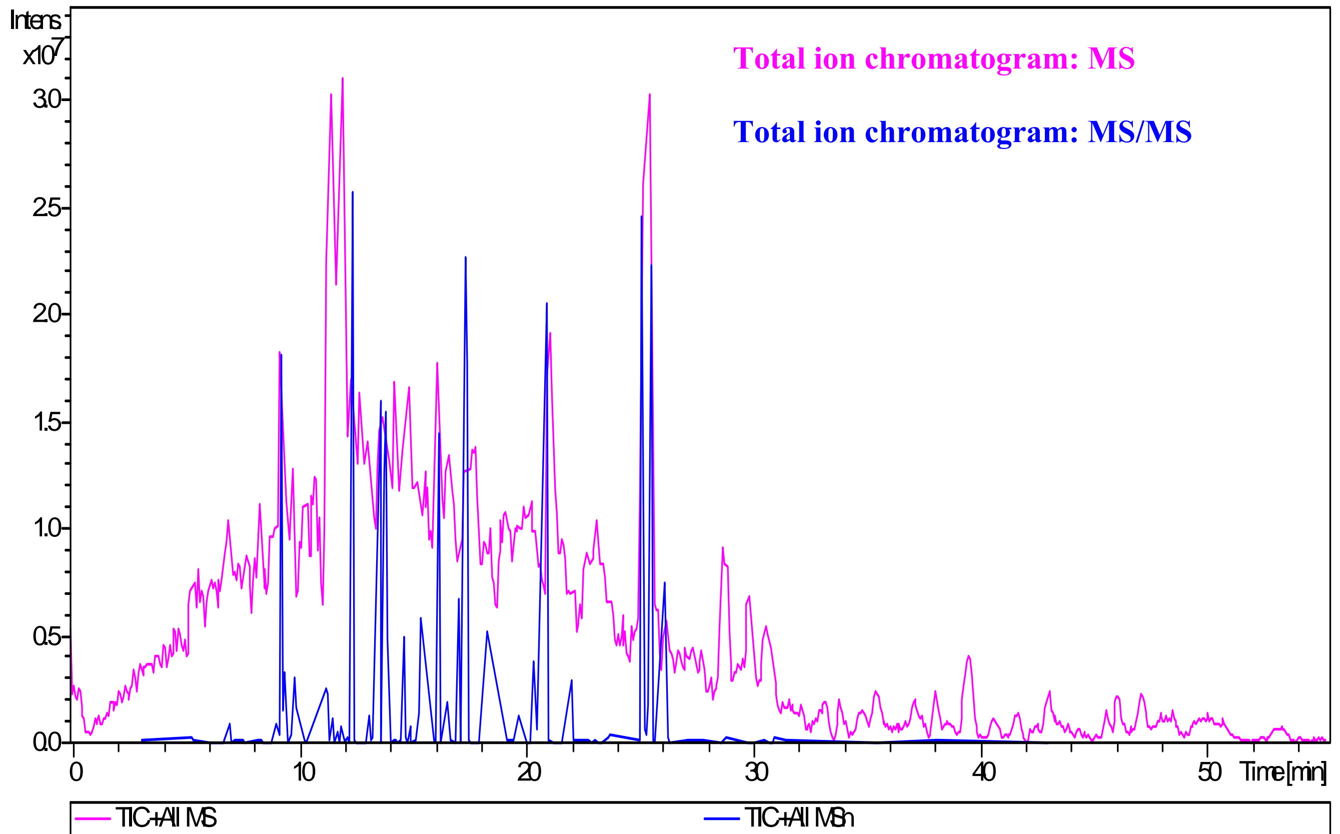


Figure 3.17 The total ion chromatogram for the entire MS and MS/MS run.

Comparing MALDI -MS and ESI- MS results

Figure 3.19 shows the peptide mass distribution of the peptides detected by MALDI and ESI measurements. The plot shows that higher mass peptides were observed using ESI. As expected, for MALDI measurements, the peptides were clustered in the mass range of 1000-2500 Da. Ignoring the unidentifiable peptides, for MALDI measurements 77% peptides contributed to the identification of 163 proteins and for ESI measurements 80% peptides contributed to the identification of 166 proteins.

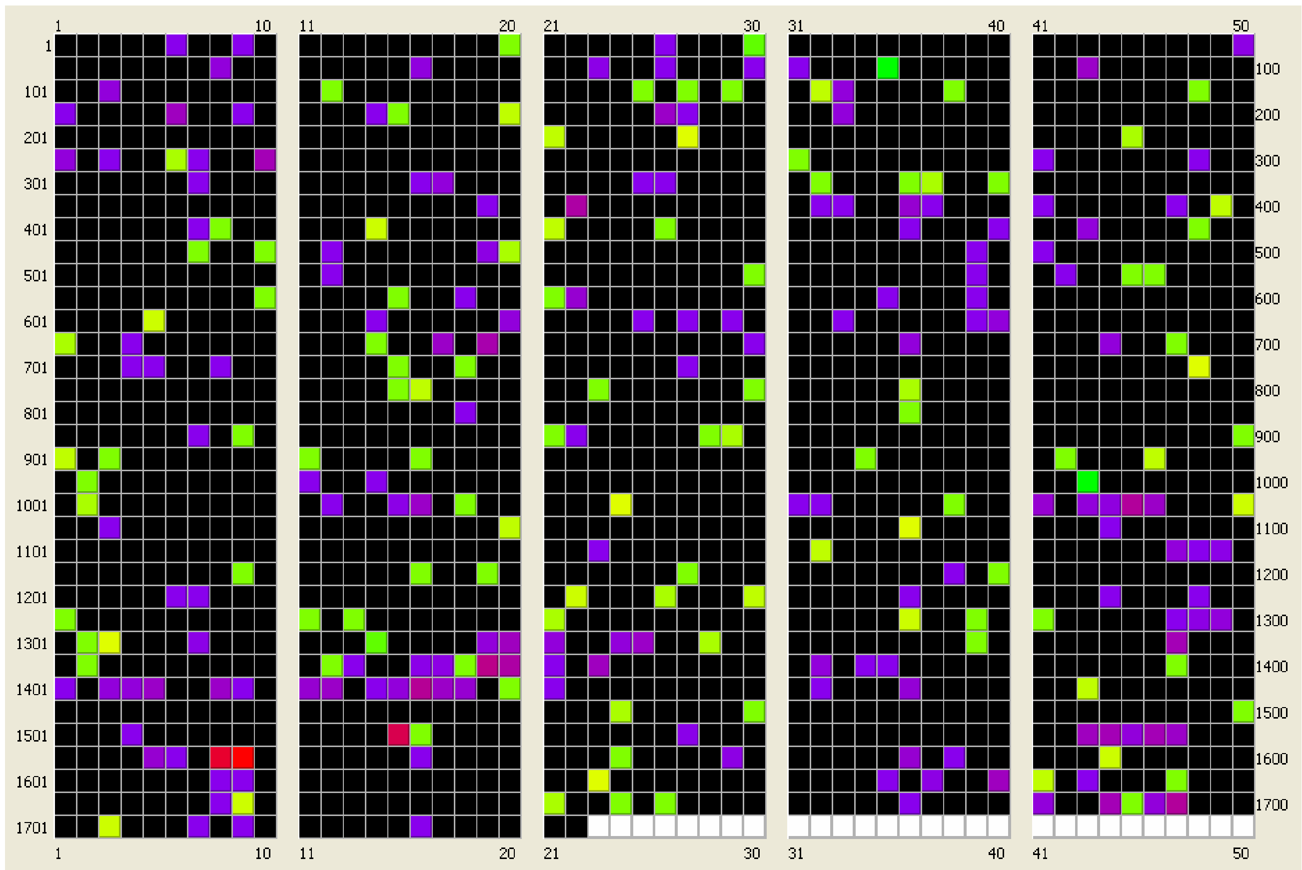


Figure 3.18 Proteome coverage map resulting for ESI-MS results where all squares represent 1722 ORFs from *M. maripaludis*. Red and purple squares contribute to the most abundant proteins. Green and yellow squares represent proteins for which non- unique peptides were observed. 166 proteins were identified at 10 ppm mass tolerance.

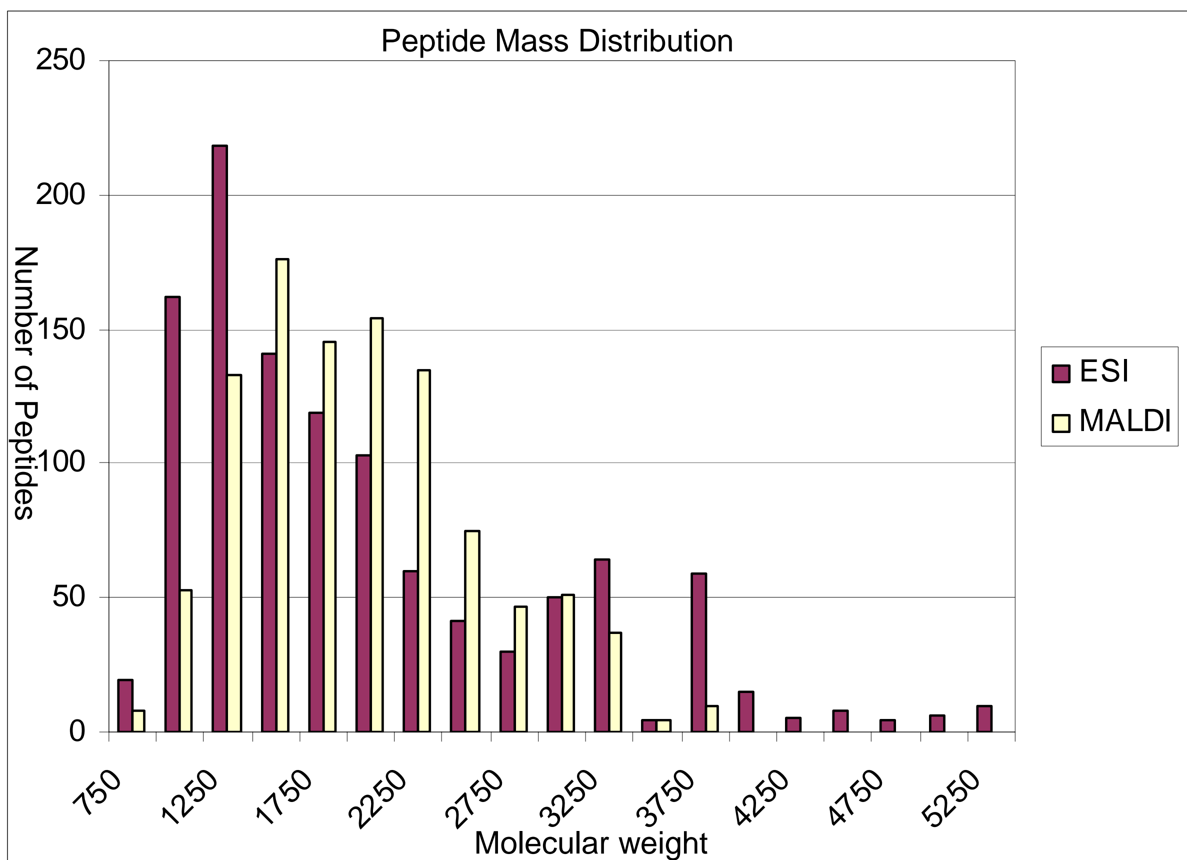


Figure 3.19 A plot of peptide mass distribution resulting from the *M. maripaludis* proteome analysis using MALDI and ESI-FTMS, 2 different ionization techniques. The plot shows the wide distribution of masses observed by ESI as compared to MALDI.

MALDI and ESI ionization techniques are known to complement each other.[36-38] Figure 3.20 shows the proteome coverage map that resulted from combining the MALDI and ESI data. Overall, 239 proteins were identified using MALDI and ESI studies of which only 90 were common to both (Figure 3.21). An example of this feature is the detection of the surface layer protein (MMP0383) which is the most abundant protein secreted by *M. maripaludis*. This protein was not detected by MALDI-FTMS analysis, but was observed by ESI-FTMS. Interestingly, this most abundant protein is fairly large (59 kDa) and contains 31 lysine residues and only two arginine residues. Due to its abundance and size, a tryptic digestion of the surface layer protein produces numerous peptides of high abundance, suppressing ion signal of other peptides. This obscures peptides of low abundance proteins and therefore lowers the detection limit of these proteins. However, an ArgC-like digestion of the surface layer protein results in two large peptides (MH^+ = m/z 25099.2849 and 32403.7116) and one small peptide (MH^+ = m/z 1447.6990) (Figure 3.22). The two large peptides are difficult to observe under MALDI-FTMS conditions due to their size and the small peptide has a low ionization yield under MALDI conditions as it is not very basic in nature. Thus, this effectively excludes the surface layer protein from analysis and allows detection of the lower abundance proteins. This feature of exclusion of the surface layer protein can be used for future proteomic studies of *M. maripaludis* where low abundance proteins are of importance. The small peptide resulting from the ArgC-like digestion of the surface layer protein is slightly hydrophobic in nature, and was detected by ESI-MS studies (Figure 3.21), but not by MALDI-MS. Therefore only one unique peptide for the surface layer protein was observed by ESI studies yielding a protein sequence coverage of only 2% and thus resulting in a purple square in the protein coverage map (Figure 3.18).

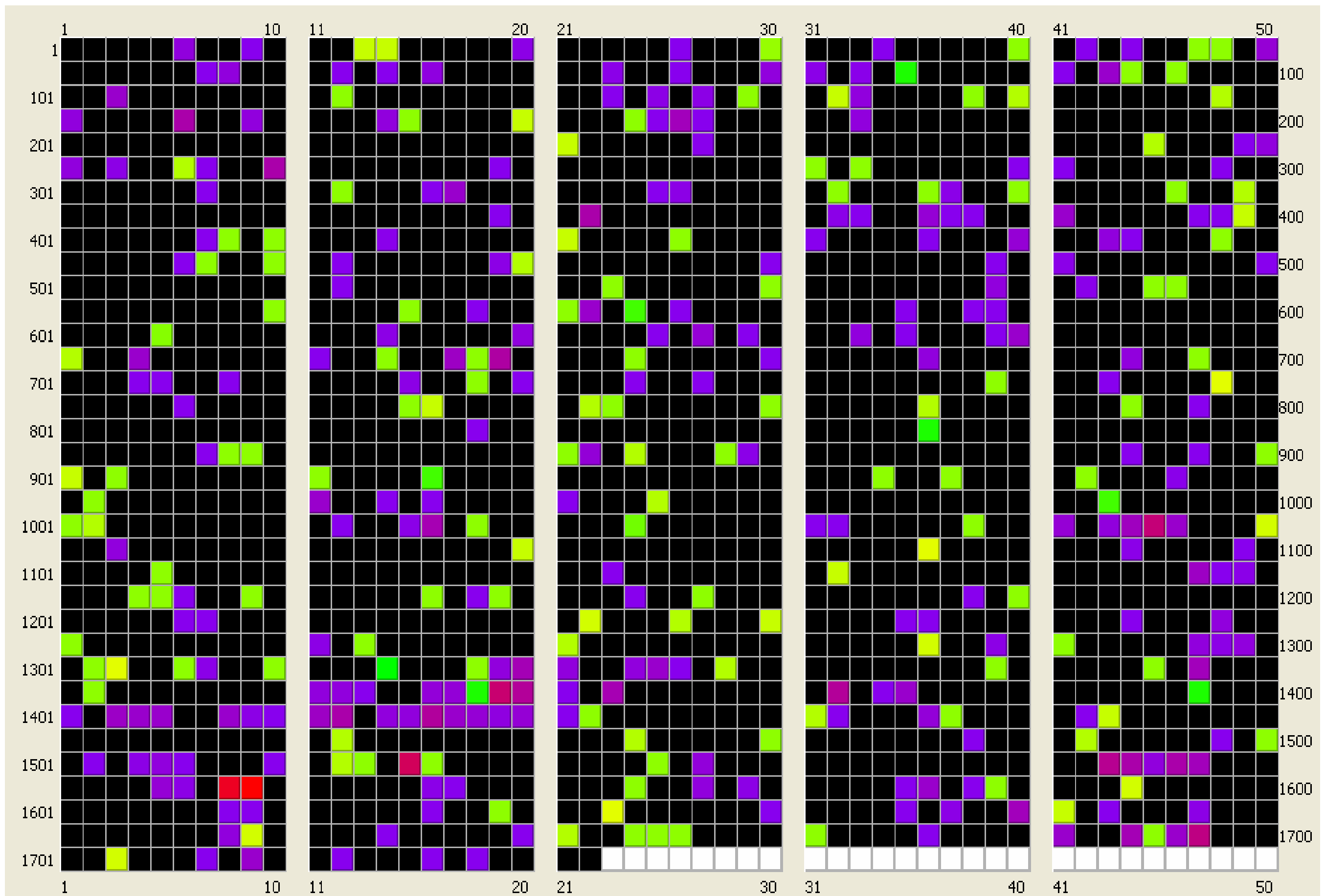


Figure 3.20 Proteome coverage map resulting for ESI and MALDI-MS analysis where all squares represent 1722 ORFs from *M.maripaludis*. Red and purple squares contribute to the most abundant proteins. Green and yellow squares represent proteins for which non- unique peptides were observed. 239 proteins were identified at 10 ppm mass tolerance.

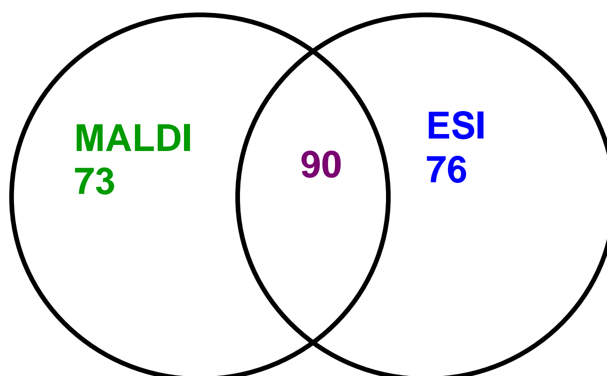


Figure 3.21 Venn diagram of the 163 proteins were identified by MALDI MS and 166 proteins were identified by ESI MS at a mass tolerance of 10 ppm. Overall 239 proteins were identified, of which 90 were common to both measurements.

CONCLUSIONS

This chapter introduces a novel method of generating larger peptides which have higher protein identification specificity for accurate mass analysis. The method implements simple quantitative chemical labeling of the free amines in the proteins followed by enzymatic digestion yielding ArgC-like peptides. The peptide patterns formed from enzymatic digestion using trypsin as a protease are very different from the ones formed by using ArgC as a protease, thus resulting in different types of proteins being identified. Combined results from a tryptic and ArgC-like digestion result in more proteins being identified. MALDI and ESI analysis of the proteome from *M. maripaludis* resulted in the identification of 14% of all potential proteins at a mass tolerance of 10 ppm. Only 37% of the identified proteins proteins were common to both MALDI and ESI

MMP0383 S-layer protein, MW:58948

MAMSMKKIGAIAVGGAMVASALATGAFAAEKVGDVDAFAADVVADGNANVDIVVGSNAAA
LDVVSAANVAAKIGSLMFKEGTVEDGSAQITVSASAESDEIKDLWATAPAAIAGQDKALV
ITAPDDDYTAVATGILGFADGVPAGTGLNLFGITTADAAVDLEDLQTLMIVEDADPSGWD
FYGTGNDFEAAEVLVAVVNTETTVPASTFGVGTPDLEVLEDEVVYASLAFVEDQYGNAAG
DYAALFPGMRIPLLGQEMVVVKLDTDDDLAIVGTEAYEGVIKQGDSYDVGNGYTVKIANV
LKQSGTAYKVDVQLLKDGKVVASKFDTVDKPAAPAAATQLKLVYKDMGIVVNAAWTDVGD
NYGYAELVIADNVVEMELGEEFIPDYEAYTVDLTAAGLTLSDKATAAGTIVGIALKYVGD
DVDKISKDKTFTIADYAKLDFDEDSGKIALKFLMDESKDVNINIGEKVSVLNAEIKLADI
EAEAKEAVVMTAPIAVLDSEASLDAADKGLILVGGPVVNALTAELADAGLVAIDNESPAT
LAVAAGAANGNDVLVVAGGDRAATEEAAEALIDML

Peptide 1:
MAMSMKKIGAIAVGGAMVASALATGAFAAEKVGDVDAFAADVVADGNANVDIVVGSNAAA
LDVVSAANVAAKIGSLMFKEGTVEDGSAQITVSASAESDEIKDLWATAPAAIAGQDKALV
ITAPDDDYTAVATGILGFADGVPAGTGLNLFGITTADAAVDLEDLQTLMIVEDADPSGWD
FYGTGNDFEAAEVLVAVVNTETTVPASTFGVGTPDLEVLEDEVVYASLAFVEDQYGNAAG
DYAALFPGMR

Peptide 2:
IPLLGQEMVVVKLDTDDDLAIVGTEAYEGVIKQGDSYDVGNGYTVKIANV
LKQSGTAYKVDVQLLKDGKVVASKFDTVDKPAAPAAATQLKLVYKDMGIVVNAAWTDVGD
NYGYAELVIADNVVEMELGEEFIPDYEAYTVDLTAAGLTLSDKATAAGTIVGIALKYVGD
DVDKISKDKTFTIADYAKLDFDEDSGKIALKFLMDESKDVNINIGEKVSVLNAEIKLADI
EAEAKEAVVMTAPIAVLDSEASLDAADKGLILVGGPVVNALTAELADAGLVAIDNESPAT
LAVAAGAANGNDVLVVAGGDR

Peptide 3:
AATEEAAEALIDML

Observed by ESI - MS

Figure 3.22 The sequence of the surface layer protein from *M. maripaludis* and the possible peptides that can be generated by ArgC-like digestion.

measurements, thus contributing to a total of 239 proteins being identified. The identified proteins had increased sequence coverage, especially for smaller proteins (up to 98%).

For both MALDI and ESI studies, 40% of the peptides submitted for the database search did not correspond to the identification of any protein. Even with a simple tryptic digestion of the proteome, about 32% of the peptides are unidentified. This high number of unidentified peptides may be due to the presence of contamination or errors in the genome database for *M. maripaludis*. In this study, however, this high number may be due to incomplete labeling of the lysine residue, which may result in some tryptic fragments. These unidentified peptides are of low abundance (< 2% relative intensity) and performing tandem mass analysis on them is a challenge. Ignoring the number of unidentified peptides, in both MALDI and ESI experiments, almost 78% of the peptides contributed to being unique peptides which is fairly close to the theoretical expected number of 89%. This shotgun proteomics approach of generating unique peptides looks promising and can be applied to other biological samples to perform high throughput studies.

References

(1) McDonald, W. H.; Yates, J. R. *Disease Markers* **2002**, *18*, 99-105.

(2) Hendrickson, E. L.; Kaul, R.; Zhou, Y.; Bovee, D.; Chapman, P.; Chung, J.; Conway de Macario, E.; Dodsworth, J. A.; Gillett, W.; Graham, D. E.; Hackett, M.; Haydock, A. K.; Kang, A.; Land, M. L.; Levy, R.; Lie, T. J.; Major, T. A.; Moore, B. C.; Porat, I.; Palmeiri, A.; Rouse, G.; Saenphimmachak, C.; Soll, D.; Van Dien, S.; Wang, T.; Whitman, W. B.; Xia, Q.; Zhang, Y.; Larimer, F. W.; Olson, M. V.; Leigh, J. A. *J. Bacteriol.* **2004**, *186*, 6956-6969.

(3) Tumbula, D. L.; Whitman, W. B. *Mol. Microbiol.* **1999**, *33*, 1-7.

(4) Haydock, A. K.; Porat, I.; Whitman, W. B.; Leigh, J. A. *FEMS Microbiol. Lett.* **2004**, *238*, 85-91.

(5) Olsen, J. V.; Ong, S.-E.; Mann, M. *Mol Cell Proteomics* **2004**, *3*, 608-614.

(6) Wong, R. L.; Amster, I. J. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 205-212.

(7) Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. *Anal. Chem.* **In press**.

(8) Goshe, M. B.; Smith, R. D. *Curr. Opin. Biotechnol.* **2003**, *14*, 101-109.

(9) Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Paa-Toli, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 2132-2139.

(10) Creighton, T. E. *2nd edition, W.H.Freeman and Company* **1993**.

(11) Hamdan, M.; Righetti, P. G. *Mass Spectrom. Rev.* **2002**, *21*, 287-302.

(12) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348–2356.

(13) Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. *Anal. Biochem.* **2000**, *287*, 110-117.

(14) Sechi, S.; Oda, Y. *Curr. Opin. Chem. Biol.* **2003**, *7*, 70-77.

(15) Goldberger, R. F.; Anfinsen, C. B. *Biochemistry* **1962**, *1*, 401-405.

(16) Perham, R. N.; Jones, G. M. T. *European J. Biochemistry* **1967**, *2*, 84-89.

(17) Dopheide, T. A. A.; Moore, S.; Stein, W. H. *J. Biol. Chem.* **1967**, *242*, 1833-1837.

(18) Adriaensens, P.; Box, M. E.; Martens, H. J.; Onkelinx, E.; Put, J.; Gelan, J. *Eur. J. Biochem.* **1988**, *177*, 383-394.

(19) Smith, M. B. *Biochim. Biophys. Acta* **1980**, *592*, 303-313.

(20) Levy, D.; Paselk, R. A. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1973**, *310*, 398-405.

(21) Nakagawa, Y.; Capetillo, S.; Jirgensons, B. *J. Biol. Chem.* **1972**, *247*, 5703-5708.

(22) Stenberg, Y.; Drakenberg, T. r.; Dahlback, B. r.; Stenflo, J. *Eur. J. Biochem.* **1998**, *251*, 558-564.

(23) Becker, L.; McLeod, R. S.; Marcovina, S. M.; Yao, Z.; Koschinsky, M. L. *J. Biol. Chem.* **2001**, *276*, 36155-36162.

(24) Kadlcik, V.; Strohalm, M.; Kodicek, M. *Biochem. Biophys. Res. Commun.* **2003**, *305*, 1091-1093.

(25) Sashidhar, R. B.; Capoor, A. K.; Ramana, D. *J. Immunol. Methods* **1994**, *167*, 121-127.

(26) Means, G. E.; Feeney, R. E. *Biochemistry* **1968**, *7*, 2193-2201.

(27) Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. *J. Mass Spectrom.* **2004**, *39*, 322-328.

(28) Geng, M.; Ji, J.; Regnier, F. E. *J. Chromatogr. A* **2000**, *870*, 295-313.

(29) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. *Nat. Biotechnol.* **2003**, *21*, 566-569.

(30) Chakraborty, A.; Regnier, F. E. *J. Chromatogr. A* **2002**, *949*, 173-184.

(31) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. *Anal. Chem.* **2001**, *73*, 5142-5149.

(32) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. *J. Chromatogr. B* **2000**, *745*, 197-210.

(33) Regnier, F. E.; Riggs, L.; Zhang, R.; Xiong, L.; Liu, P.; Chakraborty, A.; Seeley, E.; Sioma, C.; Thompson, R. A. *J. Mass Spectrom.* **2002**, *37*, 133-145.

(34) Yu, Y.; Cui, J.; Wang, X.; Liu, Y.; Yang, P. *Proetomics* **2004**, *4*, 3112-3120.

(35) Grunstein, M. *Nature* **1997**, *389*, 349-352.

(36) Bodnar, W. M.; Blackburn, R. K.; Krise, J. M.; Moseley, M. A. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 971-979.

(37) Stapels, M. D.; Barofsky, D. F. *Anal. Chem.* **2004**, *76*, 5423-5430.

(38) Medzihradszky, K. F.; Leffler, H.; Baldwin, M. A.; Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 215-221.

CHAPTER 4

A MASS DEFECT LABELING APPROACH TO N-TERMINAL PEPTIDE RESTRICTED SHOTGUN PROTEOMICS

INTRODUCTION

Current non-gel techniques for proteome analysis known as shotgun proteomics rely on the mass spectrometric analysis of highly complex peptide mixtures generated by the digestion of unseparated protein mixtures, referred to as batch digestion.[1,2] Liquid chromatography followed by mass spectrometric analysis of this complex mixture of peptides can be accomplished either by tandem mass spectrometry or by accurate mass measurements.[3] High resolution mass analysis and accurate mass measurement restricts the enormous number of molecular formulas that might be represented by a particular peptide mass. A narrow mass tolerance constrains the database search, and increases the specificity of the peptide molecular weight for protein identification.[3,4] Research in our laboratory is concentrated on developing shotgun proteomics strategies based on accurate mass measurement by using Fourier transform ion cyclotron mass spectrometry for mass analysis.

One of the challenges of shotgun proteomic analysis is the complexity of the peptide mixture that results from a batch digestion, which may contain up to 100,000 components. Each protein in the proteome will yield 50 peptides on average. The complexity of a batch digest can be reduced by isolation of the subset of peptides, as one does not need complete sequence coverage for each protein.[1,5] This approach reduces the complexity of the mass spectrum and can increase sample throughput. For example, isolation of either the N-terminal or the C-terminal tryptic peptides after proteolytic digestion will not only considerably reduce the sample complexity but will also provide a substantial gain in assignment specificity.[5,6]

This chapter explores the feasibility of analyzing the *Methanococcus maripaludis* proteome by mass spectrometric identification of isolated N-terminal peptides and using them for protein identification. *M. maripaludis* is a methanogenic archea with 1722 ORFs. The small lookup table of potential peptides along with accurate mass measurement offers increased confidence in protein identification. For almost all the proteins, translation begins at a start codon, which is almost always AUG (encodes methionine), thus incorporating a methionine residue at the N-terminus of the protein.[7] However, the N-terminal methionine is frequently absent in mature proteins, the result of post-translational processing.[7-9] The database needs to account for this possibility. Therefore, considering all the proteolytic possibilities with up to 1 missed cleavage, and plus or minus methionine at residue 1, the database will contain 6800 peptides. In contrast, the database of all tryptic peptides from *M. maripaludis* consists of ~95,000 peptides. Thus one would expect much higher identification specificity for the isolated N-terminal fragments than for the full set of tryptic peptides. Figure 4.1 shows how the peptide identification specificity of the N-terminal fragments from *M. maripaludis* varies as a function of peptide molecular weight. For masses above 700 Da, at 10 ppm mass tolerance 89% of the N-terminal peptides are unique in nature. At 10 ppm mass tolerance, using nitrogen stoichiometry as a constraint, 98% of the peptides are unique. These statistics demonstrate the increased specificity in protein identification using the small set of isolated N-terminal peptide fragments.

Several techniques for isolating N-terminal peptides have been reported in the literature. One approach relies upon alkylation of cysteine residues followed by acetylation (using N-acetoxysuccinimide) of all free amino groups, including the ε-amino group of the lysine residue and the primary amine of the protein N-terminus.[10-12] A tryptic digestion of the treated proteins results in fully acetylated peptides from the N-terminus of the proteins and internal peptides with

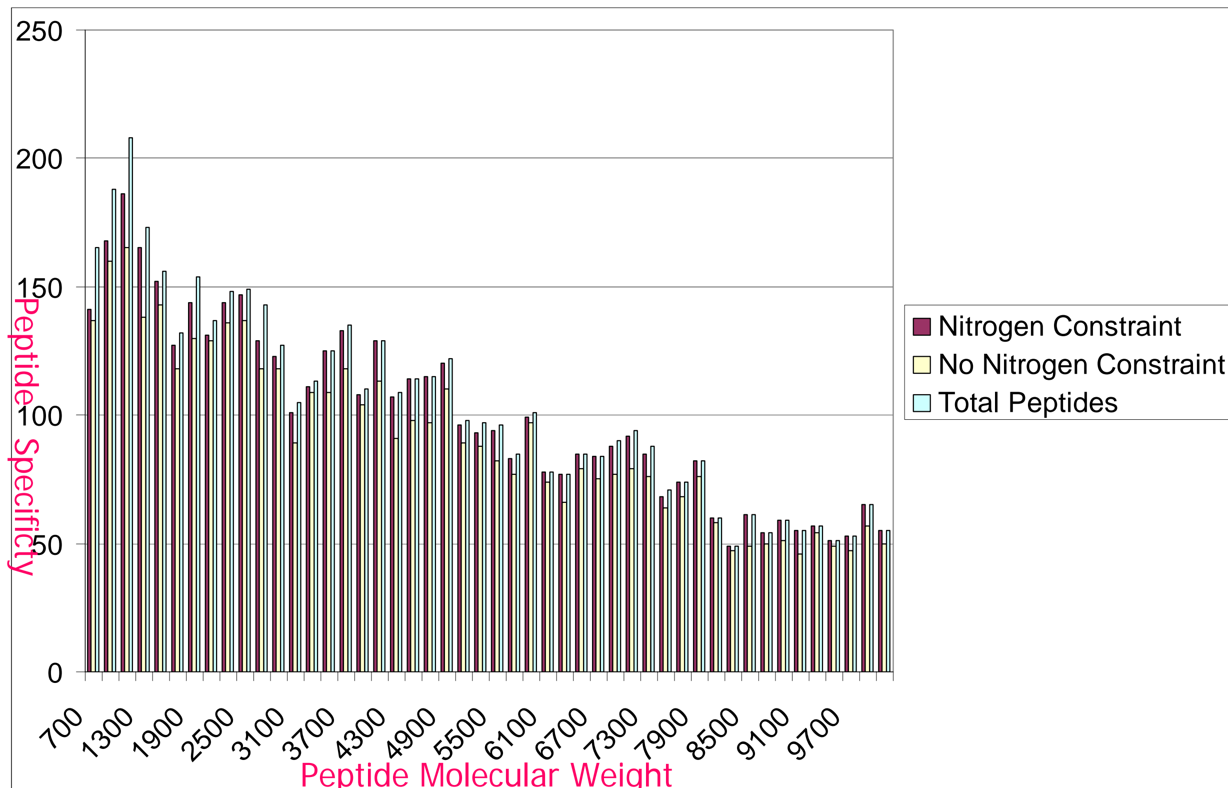


Figure 4.1 The graph shows the variation of peptide specificity with peptide molecular weight (200 amu bins) for all the N-terminal peptide fragments from M.maripaludis measured at 10 ppm mass tolerance with and without nitrogen stoichiometry as a search constraint.

a free amino group at their N-terminus. The free amine of the internal peptides is exploited by different reaction schemes to selectively isolate N-terminal peptides. Kris Gevaert and coworkers have described modification of the internal peptides with 2,4,6-trinitrobenzenesulfonic acid (TNBS) to form highly hydrophobic trinitrophenyl peptides.[2] These hydrophobic internal peptides appear at late retention times during chromatography with a reverse phase column. The unaltered N-terminal peptides appear at normal elution times and can be easily collected for

further mass spectrometry analysis (Figure 4.2). Other research groups have described modification of the internal peptides by biotinylation and removal of the biotinylated internal peptides with immobilized streptavidin, leaving behind exclusively the set of N-terminal peptides which can be further analyzed by mass spectrometric techniques (Figure 4.3).[5,9]

This chapter describes an approach based on pull-down of the internal peptides with commercially available reacti-bind maleic anhydride activated polystyrene plates. The 96 well plate utilizes a direct and simple approach to covalently attach amine containing peptides to the microplate wells. The reaction of the maleic anhydride groups with primary amines results in the formation of an amide bond which is stable at and above neutral pH conditions (Figure 4.4). If a mixture of blocked N-terminal peptides and internal peptides is subjected to the microplate well, the internal peptides that have free amines will covalently bond to the microplate well and the N-terminal peptides can be removed for mass spectrometric analysis. However, such purifications never provide 100% purity, and it is expected that a fraction of internal peptides will not bind to the wells. A solution to this problem lies in tagging the primary amine of the remaining internal peptides with a mass defect label. Mass defect labeling has been used to provide higher identification specificity for cysteine-containing peptides.[13] A mass defect label which reacts specifically with primary amine groups could distinguish the small portion of left over internal peptides from the N-terminal peptides. Figure 4.5 shows the proposed scheme for isolating N-terminal peptides using mass defect labeling. Initial experimentation is carried out on simple peptide standards to validate this technique.

Figure 4.2 Scheme summarizing the chemistry used for isolating N-terminal peptides exclusively by using 2, 4, 6- trinitrobenzenesulfonic acid [2].

H_2N — K — K — COOH
NH_2 NH_2

Acetylation

HN — K — K — COOH
NH NH

Tryptic proteolysis

HN — K — R
NH

H_2N — K — R
NH

H_2N — COOH

H_2N — R

Internal peptides

N-biotinylation

HN — K — R
NH

b HN — K — R
NH

b HN — COOH

b HN — R

Passage over
immobilized streptavidin

Biotinylated peptides
retained
and discarded

HN — K — R
NH

Acetyl

b Biotin

N-terminal peptides
analyzed by MS or MS/MS

Figure 4.3 Scheme summarizing the chemistry used for isolating N-terminal peptides exclusively using biotinylation and streptavidin.[5]

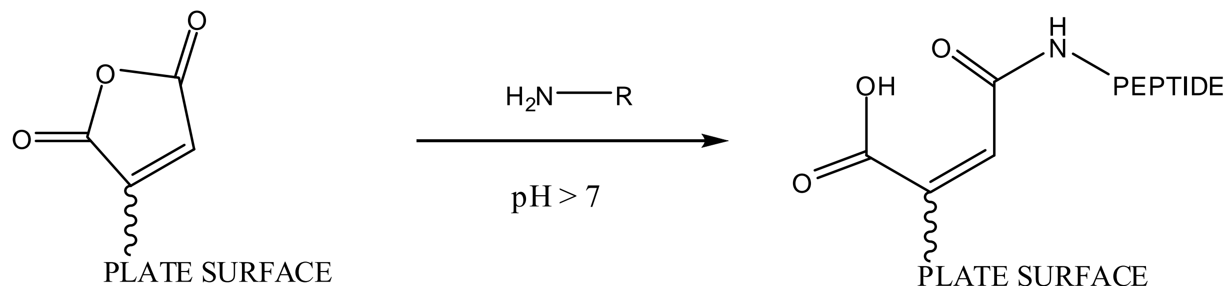


Figure 4.4 Reaction scheme for the commercially available reacti-bind maleic anhydride activated polystyrene plates for capturing the internal peptides, where R is a peptide.

EXPERIMENTAL

1 mg/mL solutions of the peptide standards, substance P and angiotensin II are prepared using 0.1 M phosphate buffer and 0.1 M mixture of ammonium bicarbonate and dimethyl formamide (50/50 by volume) respectively. 100 µL of substance P peptide is acetylated with a 50 fold excess of N-acetoxysuccinimide at pH 9 for 5 hours. The reagent is dissolved in water at a concentration of 18 mg/mL and is added in two parts in the first hour. The first part comprising of 13 µL of reagent solution is added at the beginning of the reaction and the second part comprising of 13 µL of the same is added after 30 minutes of the reaction. Treatment with excess hydroxylamine (5 µL) for 10 minutes is followed. A mixture of pure substance P and acetylated substance P (50/50 by volume) is prepared at 0.1 mg/mL at pH 8. Internal peptides were removed from the solution using the reacti-bind maleic anhydride activated microplate (Pierce, Rockford, IL.).

The capacity of each 100 µL well is 1 µg of peptide per well. The reaction protocol includes incubation of the plate for 1 hour at 37 ^{0}C on a shaker. Instead of using a dilute sample

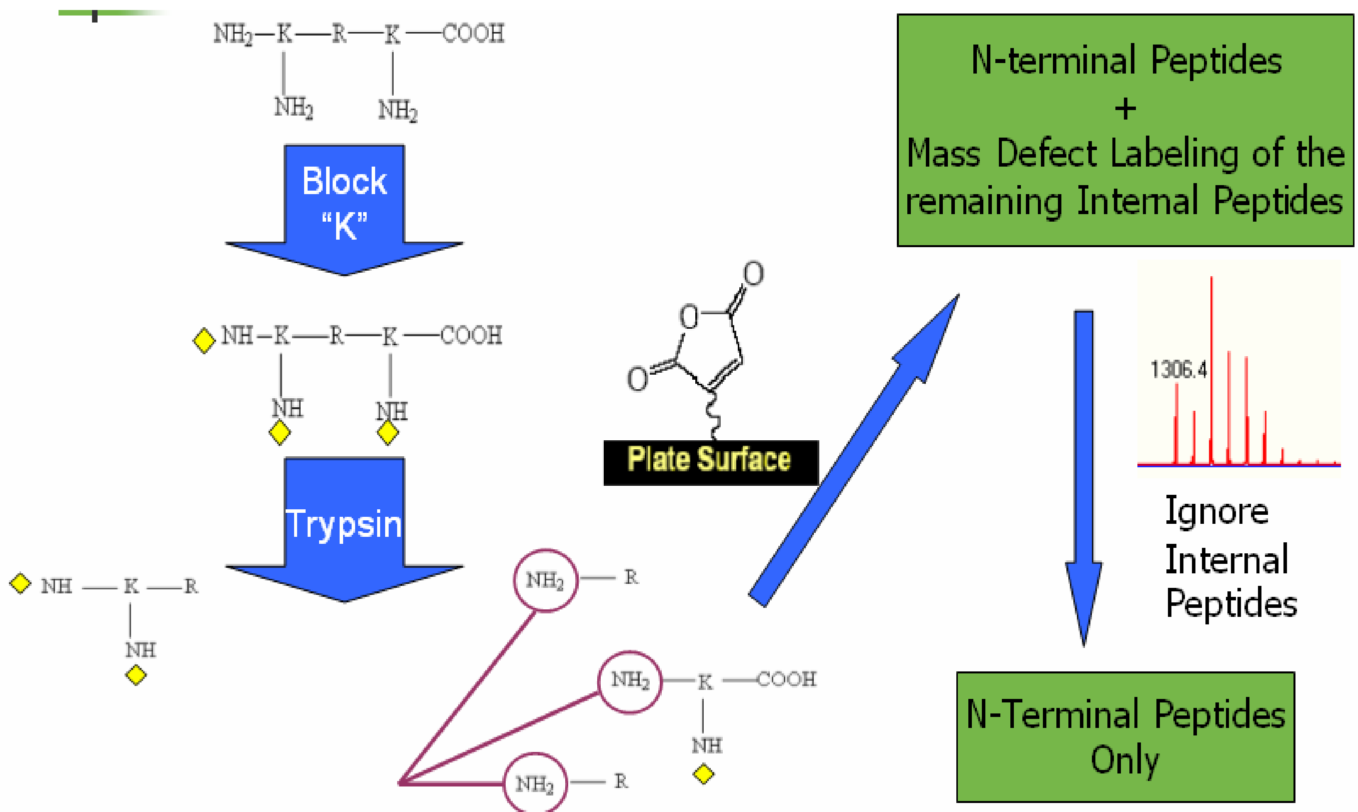


Figure 4.5 The proposed scheme describing isolation of N-terminal peptides using the reacti-bind maleic anhydride plates followed by mass defect labeling of the small portion of internal peptides.

at 10 µg/mL for a one hour reaction time, a 10 times more concentrated sample of the substance P mixture (at 100 µg/mL) is used for reaction in 10 different wells with a one hour reaction time in each well. At one hour intervals the sample is transferred by a pipette to a fresh well. After 10 hours of reaction (10 wells, one hour per well), the sample mixture is analyzed by mass spectrometry.

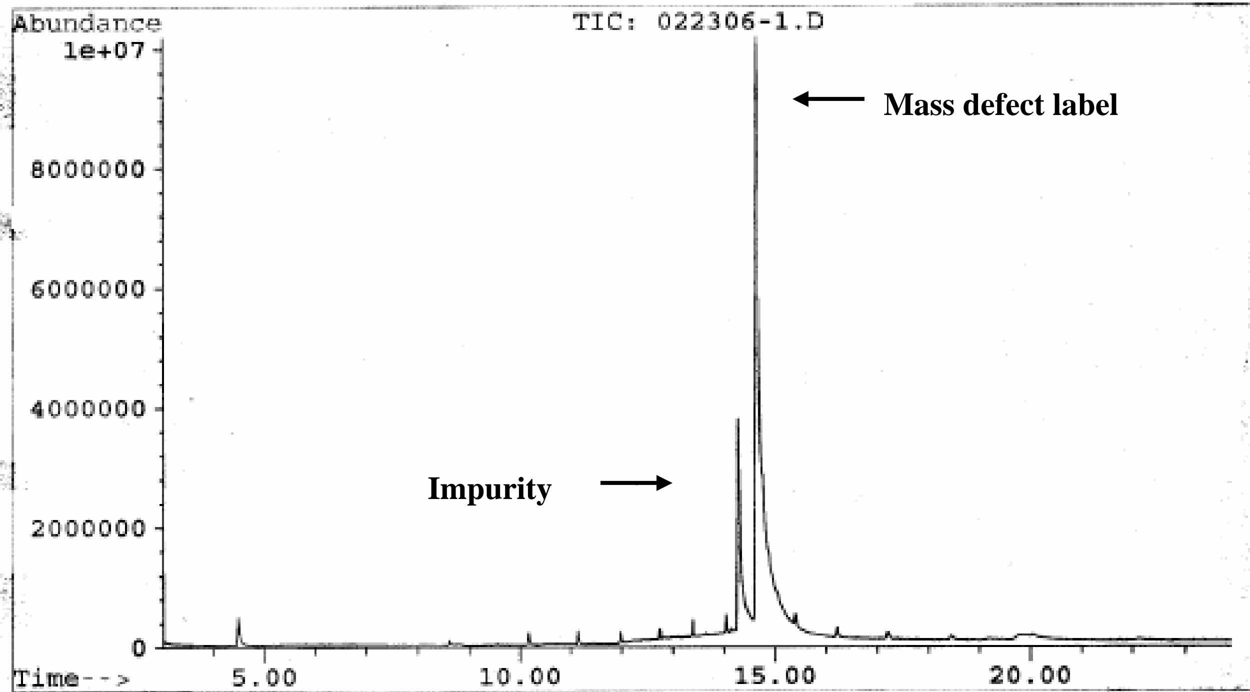
For mass defect labeling of internal peptides, the reagent N-2,5-dibromobenzoyloxy succinimide ester was synthesized by Dr. Robert Phillips (Department of Chemistry, University of Georgia). Figure 4.6 shows the result of gas-chromatography mass spectrometry (GC-MS) analysis of the compound. The GC-MS results show that although the presence of a small amount of impurity was observed, the synthesized mass defect label was substantially pure. 100 μL angiotensin II peptide was reacted with 30 μL of the mass defect label which was dissolved in dimethylformamide to make a saturated solution. After 24 hours of reaction the peptide was analyzed by mass spectrometry. The labeled peptides and peptide mixtures were analyzed on the 9.4 tesla BioApex FTICR mass spectrometer with an intermediate pressure Scout 100 MALDI source (Bruker Daltonics, Billerica, MA).

RESULTS AND DISCUSSION

To validate the use of isolated N-terminal peptides of *M. maripaludis* proteome by using reacti-bind maleic anhydride activated microplate and mass defect labeling of the internal peptides, the protocol was first tested on peptide standards. A mixture of substance P, Arg-Pro-Lys-Pro-Glu-Phe-Gly-Leu-Met-NH_2, and acetylated substance P, (N-Ac-Arg)-Pro-(ε-Ac-Lys)-Pro-Glu-Phe-Gly-Leu-Met-NH_2, was used to test the reactivity of the microplates. Substance P peptide has two free amines to bind to the microplate and behaves like an internal peptide. The acetylated Substance P has no free amines to bind to the microplate and behaves like a blocked N-terminal peptide. Figure 4.8 shows the mass spectrum of the mixture of peptides before and after 10 hours of reaction in the 10 microplate wells. The mass spectrum shows that although almost all of the underivatized substance P was removed, a small portion of the underivatized peptide still remains.

Scavenging resin beads that are commonly used in organic reactions to trap small molecules can serve as a medium to trap the internal peptides. The resin beads offer an advantage of greater surface area as compared to the well plates allowing higher loading capacity. 4-benzyloxybenzaldehyde polystyrene (EMD Biosciences, San Diego, CA), a polymer-supported scavenger resin for removing amines was tested for removal of the internal peptides. The resin has a loading capacity of 2-3 mM/g and is only compatible with organic solvents. 10 mg resin was soaked in excess dimethylformamide for 2 hours for substantial swelling. A mixture of underivatized angiotensin II, β-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe, and acetylated angiotensin II, (N-Ac-β-Asp)-Arg-Val-Tyr-Ile-His-Pro-Phe, prepared at a concentration of 0.5 mg/mL using a mixture of dimethylformamide and 100 mM ammonium bicarbonate (50/50 by volume) was added to the resin. The reaction was performed on the shaker at 50 ^{0}C for 24 hours. Figure 4.9 shows the mass spectrum of the peptide mixture before and after subjection to the resin. Though the mass spectrum shows that almost all of the underivatized angiotensin II was trapped, extra peaks showed up suggesting a side reaction. Also, using dimethylformamide, which is an integral requirement in the reaction scheme, is inconvenient because it is not protein friendly. Therefore, using 4-benzyloxybenzaldehyde was an unsuccessful attempt towards removal of the internal peptides.

To confirm the reactivity of the mass defect label with the free amine of an internal peptide, the label was tested for reaction on angiotensin II peptide. Angiotensin II, β-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe, has only one free amine for reactivity, namely the N-terminus. Figure 4.8 shows the reaction of the mass defect label with the primary amine and Figure 4.10 shows the mass spectrum of the angiotensin II peptide before and after reaction with the mass defect



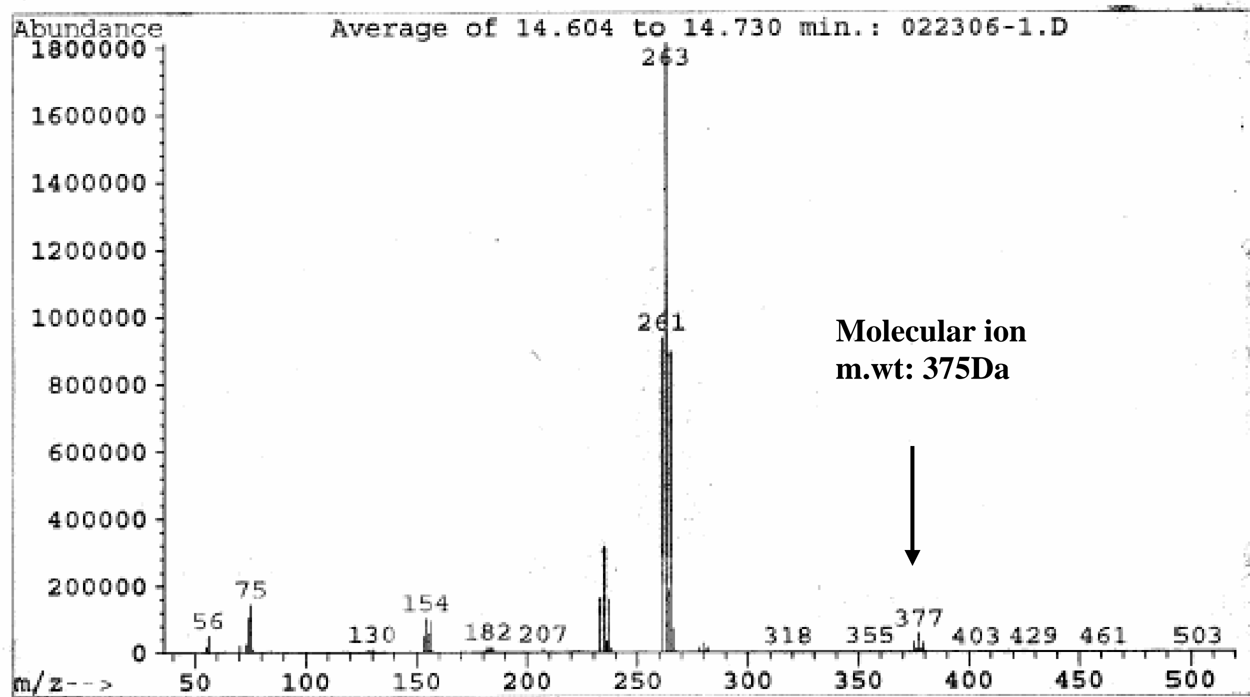


Figure 4.6 GC-MS of the substantially pure mass defect label.

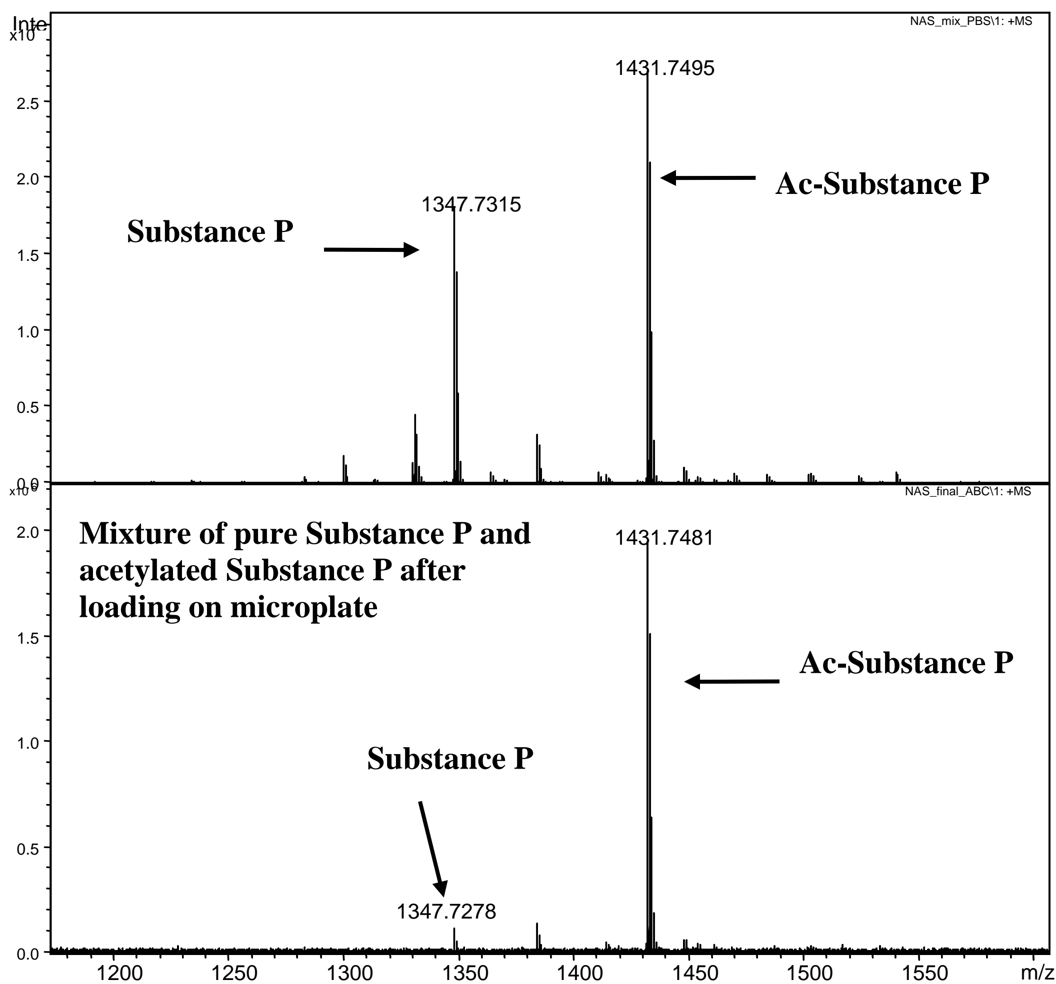


Figure 4.7 Use of reacti-bind plates to capture the internal peptides. The mass spectrum shows the mixture of acetylated and underivatized substance P before and after loading in the well plate. The reaction suggests that majority of the internal peptide was covalently attached to the walls of the plate.

label. The mass shift of 260 Da and the unique isotopic distribution suggests that the mass defect label reacted quantitatively with the free amine. However, due to the impurities present in the mass defect label, the chemistry was not clean and extra peaks were observed.

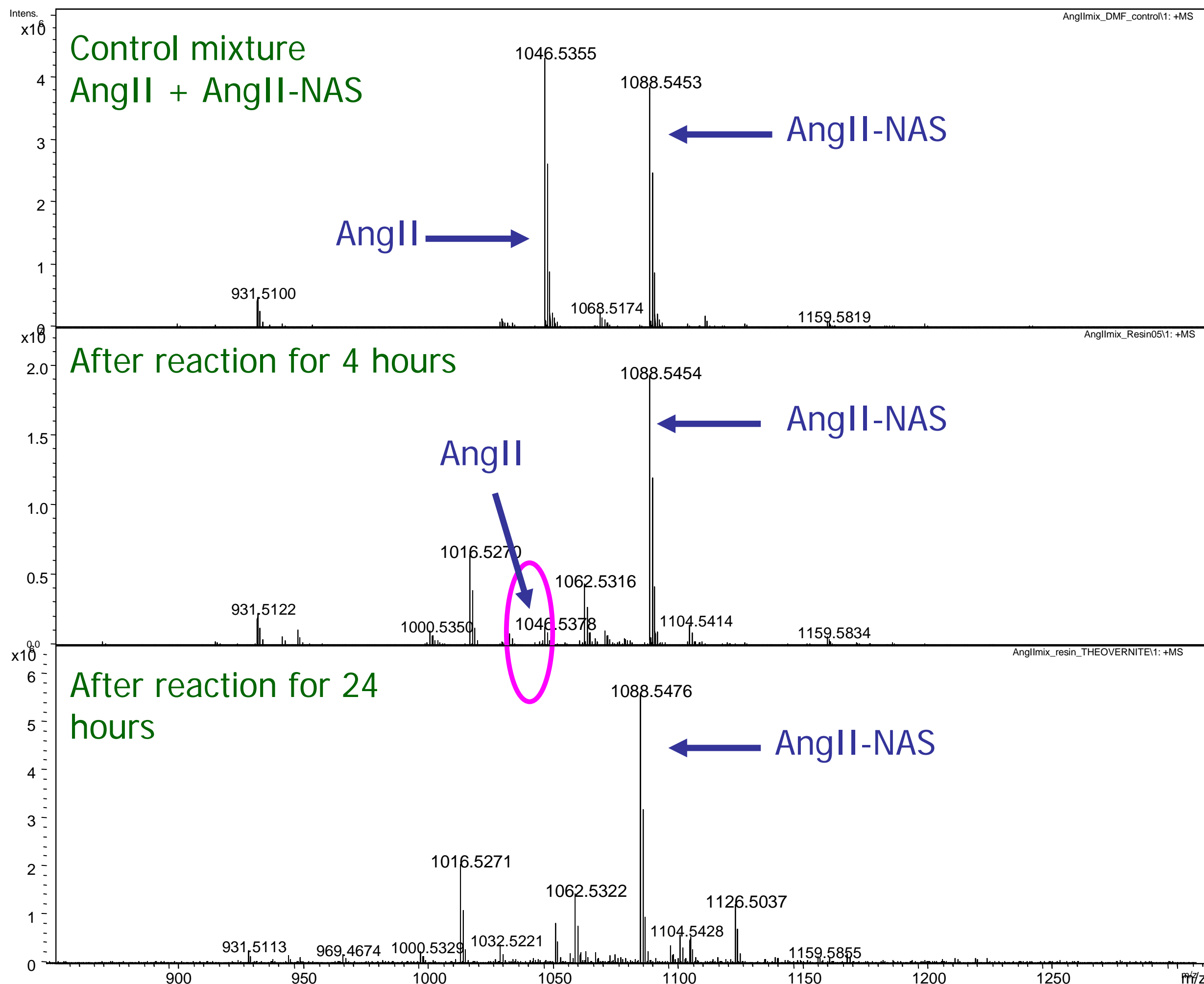
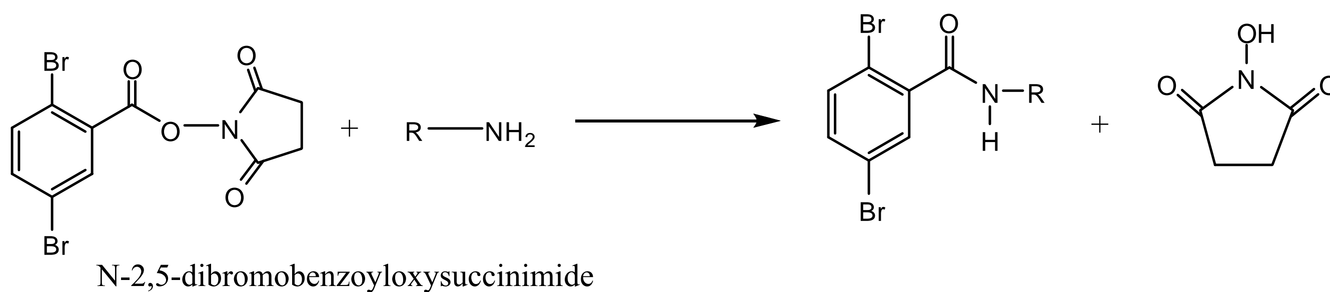


Figure 4.8 Use of 4-benzyloxy benzaldehyde polystyrene resin for trapping internal peptides. The mass spectrum shows the mixture of acetylated and underivatized angiotensin II before and after subjection to the resin. The reaction shows that after 24 hours almost all of the underivatized angiotensin II reacted with the resin.



N-2,5-dibromobenzoyloxysuccinimide

Figure 4.9 Reaction of the mass defect label with the free amine of the internal peptide.

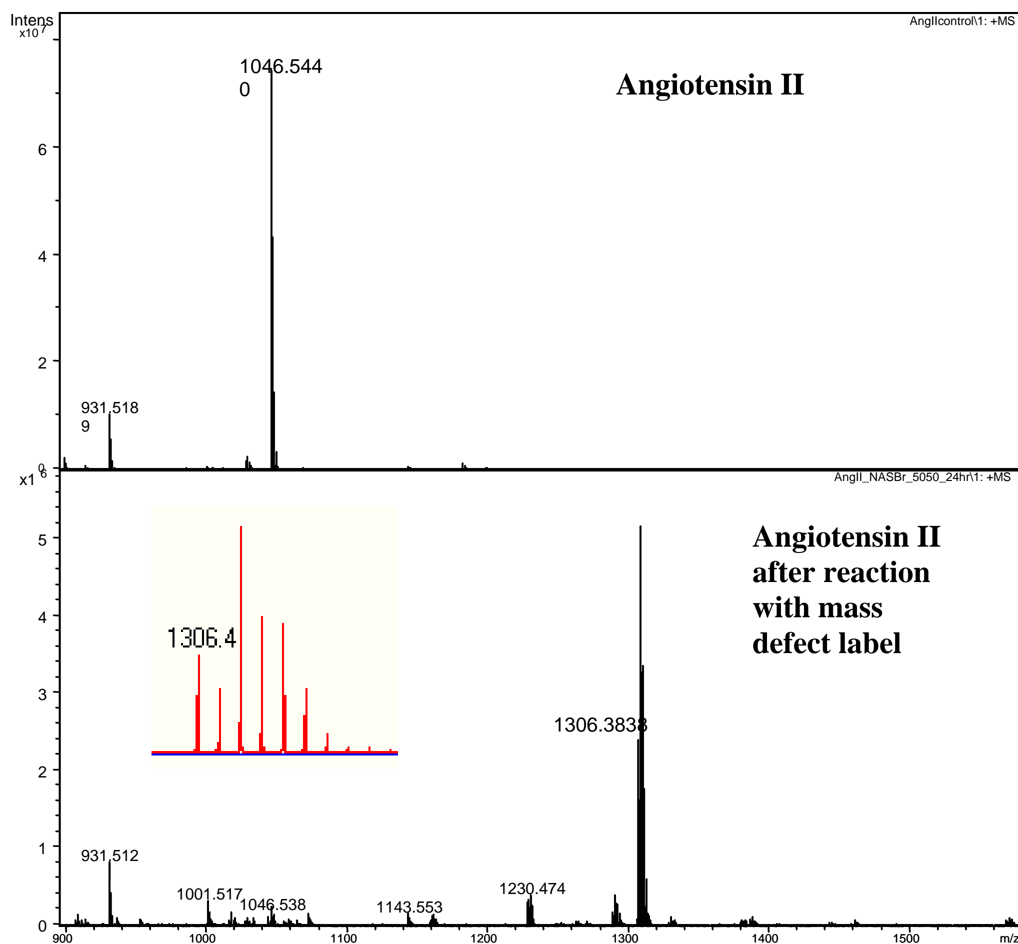


Figure 4.10 Mass defect labeling of the free amine of angiotensin II peptide. The shift of 260 Da suggests complete labeling. A mass scale expansion of a portion of the labeled peak displays the unique isotopic distribution due to the presence of two bromine atoms.

CONCLUSIONS

Positionally defined peptides such as the N-terminal and C-terminal peptides, yield substantial information gain in protein identification strategies. Shotgun proteomics studies using such informative peptides, along with accurate mass measurement reduce the complexity of the mixture, generate a small database, and considerably improve protein identification. This chapter introduces a new method for specific isolation of N-terminal peptides using N-acetoxysuccinimide for acetylation of the free amines of the proteins, commercially available reacti- bind plates for targeting the internal peptides and a mass defect label for improving confidence in isolation of the N-terminal peptides. However, this strategy requires an accurate sequence database of the *M. maripaludis* genome, specifically the ORF start sites and the N-terminal regions of proteins must be correct. To investigate the accuracy of the published sequence of the N-termini of all the proteins from *M. maripaludis*, tandem mass spectrometry should be carried out on the isolated N-terminal peptides to establish the true start site and sequence of each ORF.

Results of initial studies look promising and suggest that isolating N-terminal peptides of pure proteins should be pursued. A pure version of the mass defect label needs to be synthesized to increase specificity by eliminating side reactions. For quantitation, the stable isotope labeled form of N-acetoxysuccinimide can be used for acetylation.[14-16] This step will also assist in identifying and discriminating between chemically and naturally acetylated (modified) N-terminal peptides. On the whole, this method of isolating N-terminal peptides by use of mass defect labeling appears to be very promising and will increase sample throughput in quantitative proteomic studies.

References

(1) Gygi, S. P.; Rist, B.; Aebersold, R. *Curr. Opin. Biotechnol.* **2000**, *11*, 396-401.

(2) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. *Nat. Biotechnol.* **2003**, *21*, 566-569.

(3) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. *J. Mass Spectrom.* **2005**, *40*, 430-443.

(4) Easterling, M. L.; Mize, T. H.; Amster, I. J. *Anal. Chem.* **1999**, *71*, 624-632.

(5) McDonald, L.; Robertson, D. H. L.; Hurst, J. L.; Beynon, R. J. *Nature Methods* **2005**, *2*, 955-957.

(6) Zhang, H.; Yan, W.; Aebersold, R. *Curr. Opin. Chem. Biol.* **2004**, *8*, 66-75.

(7) Creighton, T. E. *2nd edition, W.H.Freeman and company* **1993**.

(8) Sherman, F.; Stewart, J. W.; Tsunasawa, S. *BioEssays* **1985**, *3*, 27-31.

(9) Yamaguchi, M.; Nakazawa, T.; Kuyama, H.; Obama, T.; Ando, E.; Okamura; Ueyama, N.; Norioka, S. *Anal. Chem.* **2005**, *77*, 645-651.

(10) Yu, Y.; Cui, J.; Wang, X.; Liu, Y.; Yang, P. *Proetomics* **2004**, *4*, 3112-3120.

(11) Hoang, V. M.; Conrads, T. P.; Veenstra, T. D.; Blonder, J.; Terunuma, A.; Vogel, J. C.; Fisher, R. J. *J Biomol Tech* **2003**, *14*, 216-223.

(12) Chakraborty, A.; Regnier, F. E. *J. Chromatogr. A* **2002**, *949*, 173-184.

(13) Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. *Anal. Chem.* **In press**.

(14) Regnier, F. E.; Riggs, L.; Zhang, R.; Xiong, L.; Liu, P.; Chakraborty, A.; Seeley, E.; Sioma, C.; Thompson, R. A. *J. Mass Spectrom.* **2002**, *37*, 133-145.

(15) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. *J. Chromatogr. B* **2000**, *745*, 197-210.

(16) Geng, M.; Ji, J.; Regnier, F. E. *J. Chromatogr. A* **2000**, *870*, 295-313.

CHAPTER 5

CONCLUSIONS

Proteomics is the global analysis of changes in the quantities of all proteins secreted by an organism. Such a wide-ranging analysis challenges the need for robust, automated, and sensitive high throughput technologies.[1-4] Direct analysis of a complex peptide mixture generated by digestion of an unseparated protein mixture (batch digestion) using liquid chromatography and mass spectrometry is known as shotgun proteomics.[2,5-7] Shotgun proteomics based on accurate mass measurement offers an alternative to tandem mass spectrometry analysis allowing high throughput studies.[8-10] Also, approaches using accurate mass measurement of peptide masses rather than tandem mass analysis information are highly useful for quantitative analysis.[11] This thesis describes two shotgun proteomics strategies to facilitate proteomics analysis by accurate mass measurement using FTICR-MS. Both the strategies discussed in this thesis seek to increase the rate of protein identification and to allow quantitative high throughput studies.

Chapter 3 describes a strategy to improve protein identification by generating longer peptides whose molecular weights are more specific for protein identification. ArgC digestion of a protein produces longer peptides than does trypsin, thereby increasing their specificity. However, ArgC is expensive and lacks specificity, displaying some trypsin-like behavior. The strategy described in Chapter 3 demonstrates that by acetylating all the lysine residues (using N-acetoxysuccinimide) and performing a tryptic digestion, proteolysis is blocked at lysine residues, and cleaves only after arginine residues, thus yielding ArgC-like peptides. This peptide generation technique was implemented on a metabolically labeled wild type/wild type (S2/S2) sample of *M. maripaludis* grown at early and late stages. 163 proteins were uniquely identified

using LC-MALDI-FTICR-MS and 166 proteins were identified using LC-ESI-FTICR-MS at a mass tolerance of 10 ppm and using number of nitrogens as a search constraint. Peptides resulting from ArgC-like digestion yielded higher protein coverage, especially for smaller proteins (up to 98%). Overall 239 proteins were identified by using both the techniques of which 90 were common to both. 48% of the peptides were unique and contributed to the identification of 14% proteins from *M. maripaludis*. 40% of the peptide masses that were measured did not correspond to any peptide in the database. These were from peaks of very low abundance (less than 2% relative intensity), and thus performing tandem mass spectrometry analysis on these peptides is a challenge. Removing the most abundant proteins (e.g. surface layer protein, ribosomal proteins) prior to experimentation, would help concentrating these low abundance peptides, thereby facilitating tandem mass analysis. However, ignoring these unidentifiable peptides, about 78% of peptides from *M. maripaludis* were unique which is in close agreement to the theoretical value of 89%. This difference is probably because the MALDI analysis favors lower mass peptides, which have lower specificity. Peptide specificity can be further improved with mass defect labeling of cysteine residues,[12] followed by ArgC-like digestion. This approach will produce more useful peptides and allow faster analysis with improved protein identification.

Isolating positionally defined peptides such as the N-terminal proteolytic peptide of each protein would provide substantial gain in assignment specificity, reduce the complexity of the peptide mixture, speed up data analysis, and allow high throughput studies.[13,14] Chapter 4 discusses a novel technique for improving protein identification by using isolation of N-terminal peptide and a mass defect labeling approach to remove stray internal peptides from analysis. For a small organism like *M. maripaludis* (1722 ORFs), at 10 ppm mass tolerance and 1 missed cleavage, the database of potential peptides will have only 6800 entries (includes methionine

cleavage at residue 1 for each protein). The small lookup table of potential peptides along with accurate mass measurement increases confidence in protein identification. N-acetoxysuccinimide can be used to acetylate all the free amines present in proteins.[15-18] A tryptic digestion of the treated proteins results in the formation of fully acetylated peptides from the N-terminus of the protein and internal peptides with a free amino group at their N-terminus. Results demonstrate that Reacti-bind™ maleic anhydride plates can be used to covalently attach the internal peptides so that the N-terminal peptides can be easily removed for mass analysis. Since such purifications never provide 100% purity, the expected small fraction of internal peptide that does not bind to the well can be treated with a mass defect label which is specific to the free amine of the internal peptide. A mass defect label, N-2,5-dibromobenzyloxysuccinimide, that reacts with free amines was synthesized. The mass defect label imparts a unique mass defect and isotopic distribution to the labeled internal peptides providing them with a signature that distinguishes them from the N-terminal peptides. This strategy however, relies heavily on the sequence database of the *M. maripaludis* genome, especially the ORF start sites and the N-terminal regions of the proteins. Tandem mass analysis on the isolated N-terminal peptides would help establish the true start site and sequence of each ORF. Further, quantitative studies can be performed using stable isotope labeled form of N-acetoxysuccinimde used for acetylation.[15,19,20] On the whole, isolating N-terminal peptides by use of mass defect labeling looks promising and its application can be extended to other biological samples.

References

(1) Quadroni, M.; James, P. *Electrophoresis* **1999**, *20*, 664-677.

(2) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Anal. Chem.* **2001**, *73*, 5683 -5690.

(3) Lane, C. S. *Cell. Mol. Life Sci.* **2005**, *62*, 848-869.

(4) Huber, L. A. *Nat Rev Mol Cell Biol* **2003**, *4*, 74-80.

(5) McDonald, W. H.; Yates, J. R. *Disease Markers* **2002**, *18*, 99-105.

(6) Gygi, S. P.; Rist, B.; Aebersold, R. *Curr. Opin. Biotechnol.* **2000**, *11*, 396-401.

(7) Peng, J.; Gygi, S. P. *J. Mass Spectrom.* **2001**, *36*, 1083-1091.

(8) Quenzer, T. L.; Emmett, M. R.; Hendrickson, C. L.; Kelly, P. H.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 1721-1725.

(9) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741-4750.

(10) Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999**, *71*, 2871-2882.

(11) Strittmatter, E. F.; Ferguson, P. L.; Tang, K.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 980-991.

(12) Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. *Anal. Chem.* **In press**.

(13) McDonald, L.; Robertson, D. H. L.; Hurst, J. L.; Beynon, R. J. *Nature Methods* **2005**, *2*, 955-957.

(14) Zhang, H.; Yan, W.; Aebersold, R. *Curr. Opin. Chem. Biol.* **2004**, *8*, 66-75.

(15) Regnier, F. E.; Riggs, L.; Zhang, R.; Xiong, L.; Liu, P.; Chakraborty, A.; Seeley, E.; Sioma, C.; Thompson, R. A. *J. Mass Spectrom.* **2002**, *37*, 133-145.

(16) Yu, Y.; Cui, J.; Wang, X.; Liu, Y.; Yang, P. *Proetomics* **2004**, *4*, 3112-3120.

(17) Hoang, V. M.; Conrads, T. P.; Veenstra, T. D.; Blonder, J.; Terunuma, A.; Vogel, J. C.; Fisher, R. J. *J Biomol Tech* **2003**, *14*, 216-223.

(18) Chakraborty, A.; Regnier, F. E. *J. Chromatogr. A* **2002**, *949*, 173-184.

(19) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. *J. Chromatogr. B* **2000**, *745*, 197-210.

(20) Geng, M.; Ji, J.; Regnier, F. E. *J. Chromatogr. A* **2000**, *870*, 295-313.