IMPROVING SHOTGUN PROTEOMICS BY: MASS DEFECT LABELING OF CYSTEINES AND AQUEOUS-ORGANIC SOLVENT SYSTEM FOR TRYPSINOLYSIS

by

HILDA BARRY

(Under the Direction of I. JONATHAN AMSTER)

ABSTRACT

Shotgun proteomics involves digestion of a protein mixture followed by separation of the peptides and subsequent analysis by mass spectrometry. Our approach to shotgun proteomics relies on accurate mass measurement of peptides by Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). This thesis describes two different strategies to improve proteomic analysis. The first approach uses mass defect labeling of cysteine residues in order to improve peptide assignment. The improvement originates from decongestion of the mass spectra due to the mass shift exhibited by the cysteine-containing peptides. The second approach targets a critical step for any shotgun proteomic analysis which is the proteolysis of the protein mixture. An aqueous-organic solvent system is proposed to improve the specificity of the peptides generated by trypsin and also to reduce significantly the time required for digestion.

INDEX WORDS: Mass spectrometry, Proteomics, FTICR MS, Accurate mass measurement, Mass defect labeling, trypsinolysis.

IMPROVING SHOTGUN PROTEOMICS BY: MASS DEFECT LABELING OF CYSTEINES AND AQUEOUS-ORGANIC SOLVENT SYSTEM FOR TRYPSINOLYSIS

by

HILDA BARRY

B.S, Universidad Central de Venezuela, Venezuela, 2001

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

IMPROVING SHOTGUN PROTEOMICS BY: MASS DEFECT LABELING OF CYSTEINES AND AQUEOUS-ORGANIC SOLVENT SYSTEM FOR TRYPSINOLYSIS

by

HILDA BARRY

Major Professor: I. Jonathan Amster

Committee: Ron Orlando
John Stickney

Electronic Version Approved:

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

To my parents, Jose and Hilda

ACKNOWLEDGMENT

First of all, I would like to thank Jehovah God for helping me achieve one of my goals in life. I also want to say thanks to my wonderful parents who supported my decision to move away in order to continue my education in a different language and, who continue to support and help me in every aspect of my life. My brothers, Beni and Roel, the most precious brothers in the whole world, who have helped me in every possible way since my very first day at UGA. To my wonderful husband, Kelly, who was always there when I needed to vent and talk non-stop about how my experiments were not working. Thank you for always listening to me and, for helping me realize that it really wasn't that bad. My friend Renita Okon who made me laugh when things were not that great. All the current members of my research group, Jeremy, Chunyan, Li, Richard and Liz, thanks for all your help. Also, former member Bryan Parks, thank you for teaching me my way around the lab and all the instruments.

Thanks to my committee members, Dr. John Stickney and Dr. Ron Orlando and special thanks to Dr. Robert Phillips and Dr. Vijay Gawandi for their help synthesizing different organic compounds for me. Lastly, I would like to thank my advisor, Dr. Jon Amster, for making my research possible and for supporting me during these years at UGA.

TABLE OF CONTENTS

Page

CHAPTER 1

LITERATURE REVIEW

The following chapter is intended to give an insight into the field of proteomics, and the important role that mass spectrometry plays as the technique of choice for the analysis of biomolecules. We will describe the basis of two of the most important ionization methodologies developed to the present day, electrospray and matrix-assisted laser desorption ionization. Also, the very sensitive mass spectrometer FT-ICR will be explained, as well as important information on tandem MS recent advances.

Protein Analysis, Proteomics: The field of protein analysis comprises a wide variety of purposes; it defines the identities, quantities, structures and functions of proteins, and it also characterizes how these properties vary in different cellular contexts. Protein analysis has experienced major changes from its origin, which can be traced back around 60 years, to the present day.[1] In the 1950's, it was known that different proteins were composed of different amino acids and that they exhibited different biological functions.[2] However, it was not clear how these large biomolecules could be synthesized and what exactly was their composition.[2] At that time, Frederick Sanger[3] developed a method known as dinitro-phenyl-labeling of N-terminal amino acids of peptides, which consisted of the reaction of 1,2,4-fluorodinitrobenzene (FDNB) with amino groups; after complete acid hydrolysis of the dinitrophenyl (DNP)-protein, the DNP groups remained attached to the N-terminal amino acid and can be isolated and identified.[2, 3] He used this method to determine the amino acid sequence of insulin, one of the few proteins available in pure form and large quantities at the time, and successfully achieved his goal of obtaining the first protein sequence ever reported. This achievement earned him a Nobel prize, and it would motivate other scientists to focus their work on the determination of protein structures.[4] Sanger further developed

methods for studying biopolymers, turning his attention to the nucleic acids, RNA and DNA. His work culminated in the development of the “dideoxy” technique for DNA sequencing in 1975. He later used this technique to sequence the genome of the bacteriophage f*x* 174, the first fully sequence genome that earned him a second Nobel prize a few years later.[5] Earlier the same year, Maxam and Gilbert,[6] reported another methodology for DNA sequencing sufficiently reliable for the translation to protein sequences, this was actually the first published method for DNA sequencing.[7]

Roughly at the same time that F. Sanger studies started, Pehr Edman carried out his pioneering research that later led to the sequencing method known as “Edman degradation”, this methodology was used until recent years as the only way to determine the primary structure of proteins.[4, 8] This technique allows the identification of the amino acids from a purified protein that are sequentially cleaved from the amino-terminal residue.[8, 9] In the mid 1970’s, a technique for separation of individual components from protein mixtures was introduced by O’Farrell.[10] This methodology, known as two dimensional polyacrylamide gel electrophoresis (2D-PAGE), allows separation of proteins by their isoelectric point in one dimension and molecular weight in the second dimension.[10, 11] Later, in 1979, H. Towbin[12] developed a technique that would successfully be used in conjunction with 2D-PAGE. This methodology, known as “electroblotting”, offered a simple way for detection of biomolecules such as proteins and nucleic acids. After separation in a electrophoresis gel, molecules are transferred by means of an electrical field onto a membrane which immobilizes the molecules, where they are further processed by Edman degradation in order to identify and

characterize them.[12] Although this methodology served its purpose for some time, it presented some drawbacks, such as extensive times of analysis and low sensitivity.[9] During the same time period, mass spectrometry (MS) also started emerging as one of the most reliable and versatile methods for analysis of proteins. In the 1990's, the introduction of peptide mass fingerprinting (PMF) represented a significant improvement in protein analysis, allowing identification of proteins from their proteolytic fragments.[13, 14] Basically, if a pure protein is digested with a protease that cleaves at predictable locations, the result will be a unique collection of different peptides, each with a characteristic mass, that could be accurately measured by mass spectrometry.[13] The collection of characteristic masses from protein digest, a peptide mass fingerprint, could identify a protein in sequence database. This way, peptide mass fingerprinting, together with mass spectrometry, quickly displaced Edman degradation from the front page of protein analysis methodologies since it exhibited better sensitivity and the times required to perform analysis could be decrease tremendously. In addition, MS could identify blocked or modified proteins unlike Edman degradation.[9]

During this breakthrough era, the combination of these improvements enabled a new field to come to life, the field known nowadays as "Proteomics".[15] The proteome has been defined as "the protein complement of the genome".[16] This proteome reflects the external conditions encountered by the cell, since the types of expressed proteins, their abundance, modifications and other characteristics depend on the physiological state of the cell or tissue.[17] The number of possible proteins for any given cell or tissue can be daunting due to the fact that they are estimated to be expressed over a dynamic

range of at least six orders of magnitude.[17] Fortunately, current technology and genomic DNA databases permit rapid and automated protein identification.

Since 2D-PAGE was developed, there have been recognized limitations to this technique that can significantly affect proteome analysis. Specific classes of proteins are known to be absent or underrepresented in 2D-PAGE gel patterns. In addition, low abundance proteins are often not observed and therefore not identified.[18, 19] These and other limitations presented by 2D-PAGE technology made researchers focus on further developments of mass spectrometric methods in order to overcome these drawbacks. As a consequence, recent developments in MS have enabled more accurate measurements of molecular weights, better sensitivity and, other advances that allow analysis of low abundance proteins. In addition to this, greater automation and efficiency in data acquisition is achieved by the use of tandem mass spectrometry. Specifically two technical breakthroughs in MS, the ionization techniques electrospray (ESI) and matrix assisted laser desorption ionization (MALDI), have led to a powerful convergence of the fields of mass spectrometry and protein chemistry.[17] The number of laboratories around the world that have adopted mass spectrometry in preference to 2D gel technology for protein separation and quantification continues to increase at an exponential pace.[11]

Two general and different approaches can be applied in MS-only based proteome analysis technology, “bottom up” and “top down”.[20] “Bottom up” involves cleaving the protein with a specific protease to allow protein identification from the resulting peptides. “Top down” consists of the analysis of individual proteins without

enzymatic proteolysis. Nevertheless, there are many different approaches within each one of these methodologies.[18] The strategy used by our laboratory involves the “bottom up” approach known as “shotgun proteomics”, which consists of performing enzymatic digestion of the protein mixtures for their subsequently separation and MS analysis.[21-23] Figure 1.1 shows a schematic representation of a shotgun proteomic analysis. This method is capable of producing high-throughput protein identification provided that proteolytic digestion is consistent between experiments as well as protein expressions.[21] Shotgun proteomic analyses are widely used for identification of proteins in biological systems and it is often combined with differential stable isotopic labeling technology which provides relative quantification of multiple proteins simultaneously.[24-26]

Matrix-Assisted Laser Desorption Ionization (MALDI): During the 1980’s various research groups attempted to solve the volatilization/ionization problem of mass spectrometry using laser light as the energy source. Letokhov, demonstrated that small polar molecules could be partially vaporized by focusing a light beam onto them and yet, chemical degradation could be avoided.[27] This approached was further developed by M. Karas and F. Hillenkamp.[28] In the mid 1980’s, this group had showed the use of an absorbing matrix to help the volatilization of small molecules. In 1988, the application of this method to large biomolecules was reported by both K. Tanaka,[29] and by Karas and Hillenkamp,[30] who showed that gaseous ions could be formed from macromolecules using a low energy laser at long UV wavelength (>330 nm) in order to avoid absorption by the aromatic amino acids in proteins that could lead to fragmentation.[29, 31] Further developments of this approach led to the technique known as MALDI, matrix-assisted

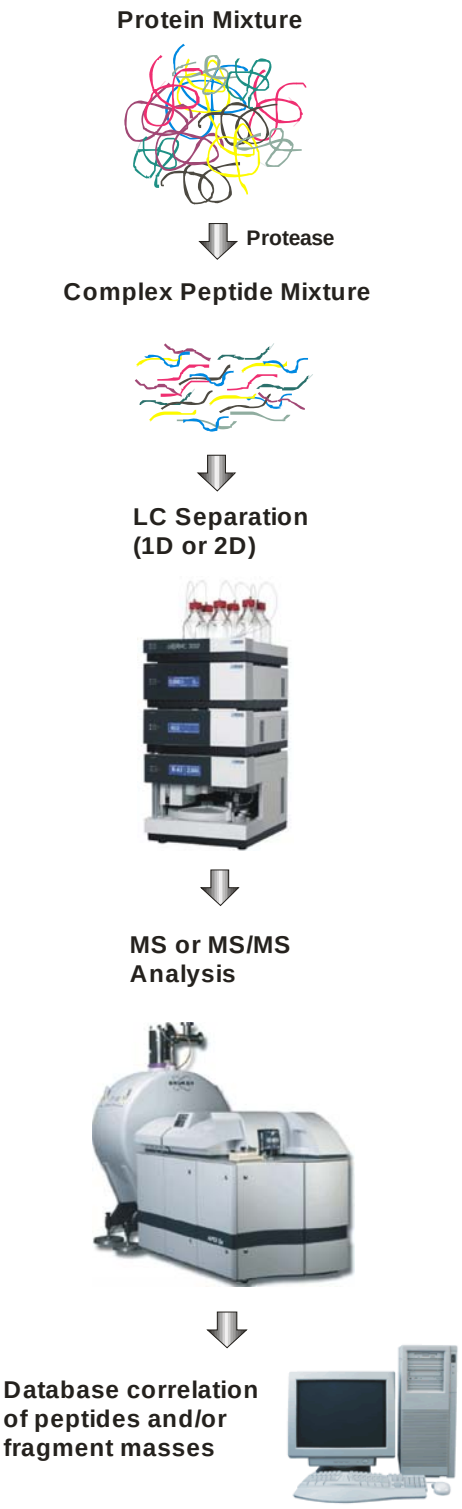

Figure 1.1 Diagram of a shotgun proteomic analysis

laser desorption ionization, which incorporates the molecules of interest in a crystalline matrix of low molecular weight.[30, 32, 33]In MALDI, sample molecules are irradiated with a pulse laser beam to promote desorption ionization of the sample. The sample is mixed prior to analysis with a molar excess of a highly absorbing matrix, which typically consist of an aromatic acid with a chromophore that strongly absorbs the laser wavelength.[34, 35] In most commercially available MALDI mass spectrometer, a nitrogen laser of wavelength 337 nm is used.[29, 32] The irradiation by the laser induces an explosive evaporation of the matrix crystals into the gas phase, entraining intact analyte in the expanding matrix plume.[36] Different models have been proposed to explain desorption of the matrix-sample material from the crystal surface, but the origin of ions produced in MALDI is still not fully understood.[37] The most widely accepted ion formation mechanism involves gas-phase proton transfer reactions of the sample molecules with the photoionized matrix molecules in the expanding plume (Figure 1.2).

The MALDI matrix must meet a number of requirements simultaneously. It should be able to embed and isolate analytes, be highly absorbing at the laser wavelength, be compatible with the vacuum requirements of an ion source, be soluble in solvents compatible with the analyte, and lastly, promote analyte ionization. It is believed that compounds with exchangeable protons, such as carboxylic acids, are good MALDI matrices for peptides and proteins because they easily protonate the neutral analyte molecules in the expanding plume.[38] However, when denaturation of the tertiary structure of biomolecules is unwanted, the use of nonacidic matrices is recommended.[39] A number of matrices have been tested throughout the years and selection of the best matrix usually depends on both the mass analyzer used as well as the nature of the

sample.[34, 40-45] Another aspect of the MALDI analysis is the sample preparation step, basically, mixing of the matrix with the analyte of interest. The original simple sample preparation procedure known as the “dried-droplet” method is still widely used and it has remained almost intact since its introduction in 1988 by Karas et. al. Several other methods have been reported since then and they are found to be very useful for analysis of biomolecules.[30, 46-49] Factors such as the choice and concentration of the matrix, the organic solvent used, and the sample deposition technique can strongly influence the spectrum profile. MALDI, however, is more tolerant to the presence of buffers, salts and detergents than other ionization techniques.

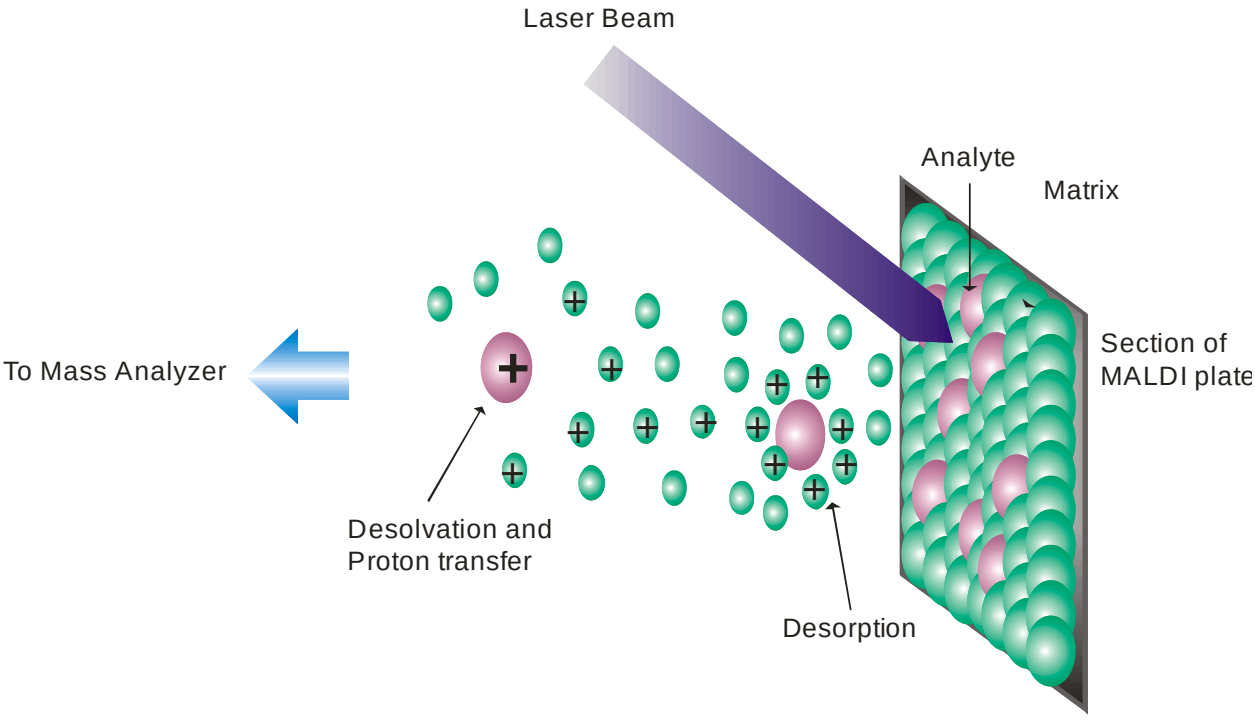

Figure 1.2 A schematic of the MALDI process

Conventional MALDI sources are operated under high vacuum in order to attain a large mean free path of ions in the spectrometer. However, high vacuum conditions

can promote metastable decay of ions produce in the MALDI source which increases the complexity and limits usefulness of the spectrum obtained.[50, 51] In past years, several researchers have developed MALDI sources that can be effectively operated at atmospheric pressure, thus reducing the chances of production of metastable ions.[50-56] Recently, MALDI sources operated at intermediate pressures have been used.[57, 58] These pressures range from 10^{-2} to 1 Torr, that is pressures which are at least 10,000 times higher than for a conventional ion source. This type of operational conditions promotes collisional thermalization of the analyte ions, which reduces the degree of ion fragmentation typically found in conventional vacuum MALDI.[57-62] Due to MALDI's strengths, this technique has been enjoyed wide utilization as a tool in proteomic studies, to identify proteins from a variety of samples.[63] Other advantages worth mentioning are: [63-66]

- Spectral simplicity due to mostly singly charged molecules
- High mass range (up to 1M Da) [67]
- High sensitivity (femtomole down to attomole range, depending on mass analyzer)
- High tolerance to the presence of common components of biological buffers
- Unlike electrospray ionization (vide infra), can interrupt data acquisition with no detrimental consequences

However, MALDI as any analytical technique presents several limitations which are listed below: [63-66]

- Working mass range limited by matrix interference below 700 Da
- Low shot to shot reproducibility
- Analyte signal exhibits strong dependence on sample preparation method
- Photodegradation of some analytes
- Low degree of charging limits utility for high molecular weight analytes

Electrospray Ionization (ESI): The development of electrospray for ionizing macromolecules dates back to the 1960's when Malcolm Dole, due to the pressing need for a viable method to analyze high molecular weight biomolecules developed a method to generate small droplets containing single macromolecules.[68-73] Some decades passed before the introduction of the combination of this ionization technique with a mass spectrometer by John Fenn who demonstrated its utility for biomolecule analysis.[74] The electrospray ionization source was first presented by Yamashita and Fenn at a symposium in San Francisco in 1988.[74] Nowadays, this technique is widely used by scientists around the world in order to obtained detailed information about molecular weights and structures from a variety of samples; it is also naturally compatible with many types of separation techniques which makes it the technique of choice for on-line separation/analysis of analytes.[75, 76] In the electrospray process, a sample dissolved in an aqueous / polar organic solution is pushed through a capillary which is held at a high potential. This high electric field creates an elongated tip of liquid

at the exit of the capillary known as “Taylor cone”. From the cone emerges a mist of highly charged droplets, which subsequently evaporate to produce lone ions that are then analyzed by the mass spectrometer.[77-82] Although a number of researchers have focused in finding a mechanism to explain this phenomenon, questions remain regarding the mechanism by which the charged droplets evaporate to ultimately produce gas-phase ions.[73, 77, 83, 84] Two models are accepted as possible mechanism for the production of gaseous ions, ion evaporation model proposed by Malcon Dole and the charged residue model by Iribarne and Thomson.[73, 85-87] Figure 1.3 shows a representation of the electrospray process.

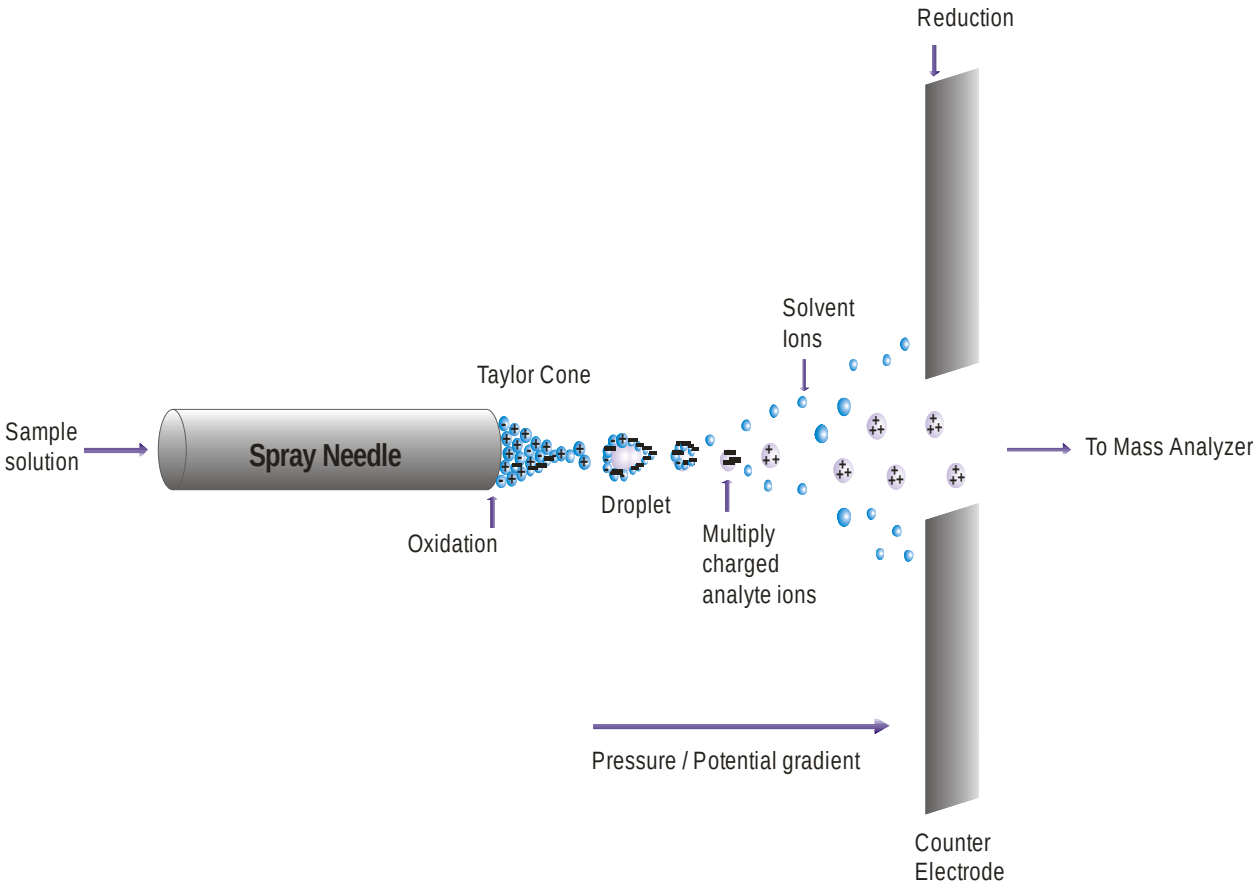

Figure 1.3 A schematic of the ESI process

In the electrospray processes, the ions observed are typically ionized by the addition of a proton, or other cation, or the removal of a proton. This technique is able to produce multiply charged ions which enables the analysis of large biomolecules at relatively low mass-to-charge ratios.[38, 70, 82, 88-90] Multiply charged ions are also ideal for characterization by tandem mass spectrometry (MS/MS).[91-95] It has been found that the sensitivity of the electrospray process is enhanced by lowering the flow from μL/min rates to nL/min rates and, a number of laboratories have devoted efforts in that direction making possible to achieve sensitivities of attomole levels.[75, 96-98] Working at the lower flow rates of nL/min is commonly referred to as “nanospray”. An additional advantage of this type of electrospray ionization over “conventional” ESI is that is more tolerant to a wide range of liquid compositions.[97, 98]

In general, ESI’s advantages can be summarized as:

- Soft ionization, which for the most part do not degrade the molecule during the ionization process
- Multiple charging allows the detection of high mass compounds at m/z ranges easily determined by most mass spectrometers
- Readily coupled to liquid separations for on-line analysis
- Enables MS/MS of high MW molecules
- Very low chemical background which leads to excellent detection limits

Of course, ESI also exhibits disadvantages, the most common being:

- Presence of buffers, salts reduces sensitivity dramatically

- Complexity of spectra increases due to multiple charges species for each analyte
- Ion suppression occurs for mixture, requiring on-line separation for analysis of complex mixtures

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS): This mass spectrometric method was inspired by the early development in conventional ICR spectroscopy and Fourier transform nuclear magnetic resonance (FT-NMR).[99] In the late 1920s, Ernest O. Lawrence and his students began working on a device known as the "cyclotron". A couple of early prototype models were constructed in 1929, but it was not until 1932 that one of the models gave evidence of full functionality.[100] That year, Lawrence and his student, M. Stanley Livingston, demonstrated that a charged particle moving perpendicular to a uniform magnetic field is limited to a circular orbit in which the angular frequency of the particle's motion is independent of the particle's orbital radius. This can be visualized in equation 1, known as the cyclotron equation:

Equation 1 $$\omega_c = \frac{q \cdot \mathbf{B}}{m}$$

where, ω_c is angular frequency, q is the particle charge, **B** is the magnetic field and m is the particle mass. The significance of Lawrence's development was that a particle could be excited by use of only small electric field strength to very large kinetic energy. Also, since different ions would have different cyclotron frequencies, this would allow their characterization with no need for prior separation of the ions. This device was used for

atom-smashing experiments, but by 1950 it was also applied to measuring mass spectra of different species.[101] Approximately at the same time, conventional nuclear magnetic resonance (NMR) spectroscopy was first demonstrated and it was of special interest for chemist to know that the NMR frequency of a particular magnetic nucleus varied from compound to compound.[102] This allowed the development of the NMR technique for chemical applications by the mid 1960's. In 1966, Richard Ernst and Wes Anderson developed a methodology to perform Fourier transform (FT) - NMR experiments in order to overcome the low sensitivity problem presented by conventional NMR spectroscopy.[103] This technique was quickly commercialized by the 1970s. Recognizing the enhancements that result from combining FT technology with an analytical spectroscopy method, studies were focused on combining FT and ICR into a powerful technique that would allow broadband excitation of analytes and conversion of the broadband ICR signal current to a signal voltage with a broadband RC circuit.[99, 104, 105] In 1974, M. Comisarow and A. Marshall described FT-ICR applied to mass spectrometry for the first time.[106, 107] Since then, many research labs have devoted efforts to further develop this technique, which exhibits ultra high resolution and an impressive wide mass range.[105, 108-118] In 1990, the first combination of the ionization technique electrospray and FT-ICR was reported by Hunt and McLafferty and, since then, many advances have been performed greatly facilitating the field of proteomics research.[90, 119-122] A few years later, another important development took place, the combination of the ionization technique MALDI with FT-ICR mass spectrometry.[123-125] The first experiments were reported by Hettich and Buchanan and, as with ESI-FTICR

mass spectrometry, proteomics research has benefited from this development and further advanced.[123-127]

General Principles: FT-ICR mass spectrometry or FT-MS is a technique based on measuring cyclotron frequencies of ions in a spatially uniform static magnetic field. It consists of simultaneously exciting all ions by a rapid scan over a wide frequency range.[38] This induces the ions to travel in a circular orbit perpendicular to the magnetic field (Figure 1.4),[118] with a frequency give by equation 1. The signal is subsequently detected on a pair of plates as an image current which is induced by the ions moving in phase, as they execute their cyclotron motion (Figure 1.5). The resulting time domain signal contains all the cyclotron frequencies that have been excited. Fourier transformation of the time domain signal provides a frequency spectrum, which is converted into a mass spectrum based on the inverse proportionality between frequency and the mass/charge ratio. Figure 1.6 shows a detailed diagram of the steps involved in a FT-MS experiment.

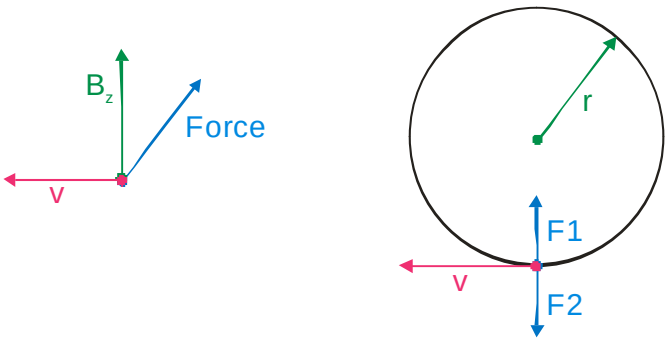

Figure 1.4 Ion cyclotron motion. Lorentz magnetic force induces ions to travel in a circular trajectory perpendicular to the magnetic field.

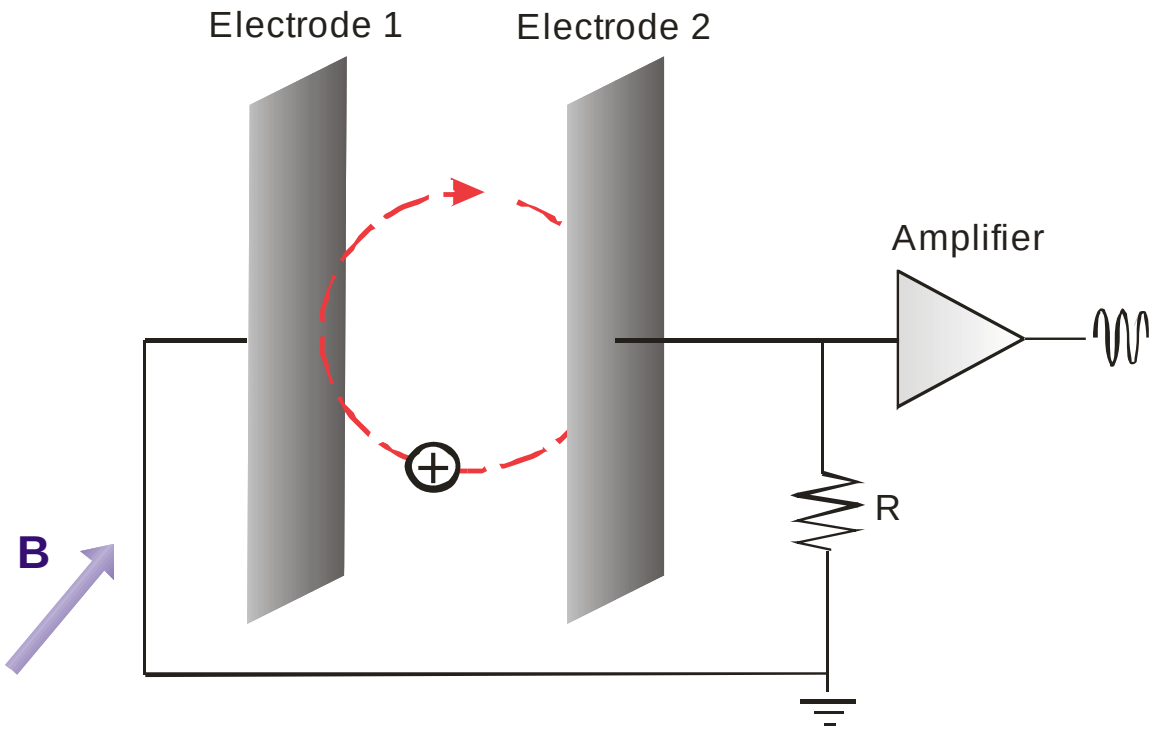

Figure 1.5 Diagram of an ion packet cycling in the magnetic field between two electrode plates. As the positively charged ions circles between both electrodes, migration of electrons between the detection electrodes produces an image current, which is converted to a voltage and amplified, producing a sinusoidal signal.

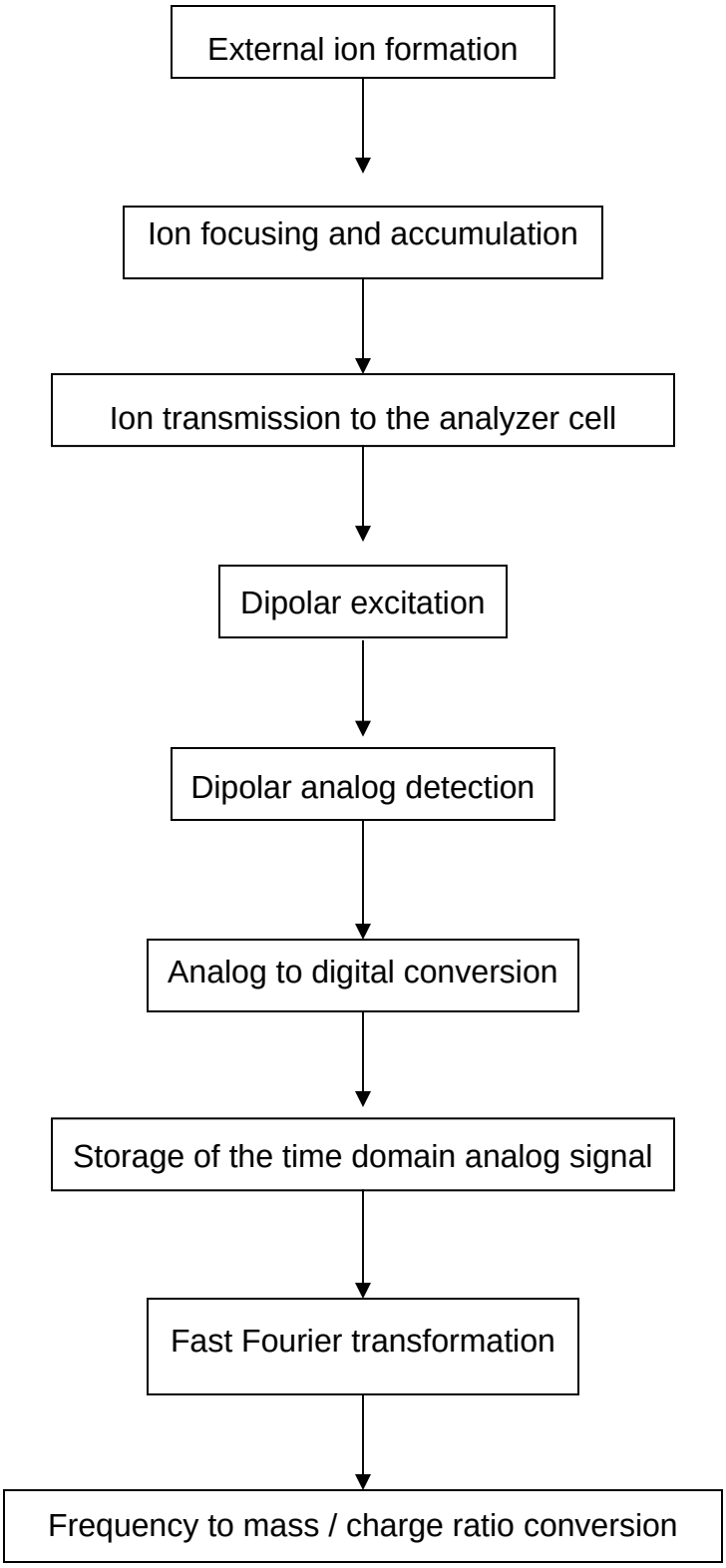

Figure 1.6 Series of events generally found in a FT-MS experiment.[128]

Ion Motion: The main motion considered for FT-MS is ion cyclotron motion.[118] When an ion moves perpendicular to the direction of the uniform magnetic field, the Lorentz force causes the ion to be deflected into a circular orbit. This force is describe in equation 2, where *q* is the charge of the ion, v is the velocity and **B** represents the magnetic field.

Equation 2 $$F = q \cdot \left(v \otimes \mathbf{B} \right)$$

This force exerted on the ion is made up of two components, the inward force on the ion (F1 in Figure 1.4), and the outward force on the ion (F2 in Figure 1.4). For the ion to maintain a stable circular orbit the magnetic force and the centrifuge force must exactly equal each other. By equating this two forces, the cyclotron frequency of an ion of mass *m*, and charge *q* can be derived (Equation 1).[110] The resulting frequency has no dependence of the velocity of the ion, thus allowing ions of a given mass to have the same cyclotron frequency regardless of the time the individual ions enter the cell. The ions also experience two other types of motion, known as trapping and magnetron motion, which are discussed elsewhere.[118]

Ion Excitation and Detection: One of the main components of FT-MS is the analyzer cell, where the excitation and detection of ions occur. The ions are formed outside of the cell and are subsequently trapped in the cell located in the homogeneous field region of a large superconducting magnet, and it is in this place where they experience the motions mentioned above. The cell is the 'heart' of the FT-MS

instrument and several different geometries have been described. [118, 129, 130] Regardless of the cell's geometry, they all exhibit trapping electrodes adjacent to the excite and detect electrodes to create a trapping cell for the ions. Figure 7 shows a cylindrical cell whose principal axis is aligned with the magnetic field. In this configuration the end cap electrodes are replaced with open cylinders, and the center cylinder is divided into four electrodes which function as excitation and detection plates. This type of configuration exhibits advantages over other types of cell, provides easy access to the interior of the trap to facilitate loading and ejection of charge particles, and introducing laser beams. Also, the effect of charging and contamination of the trapping plates is eliminated.[118, 131]

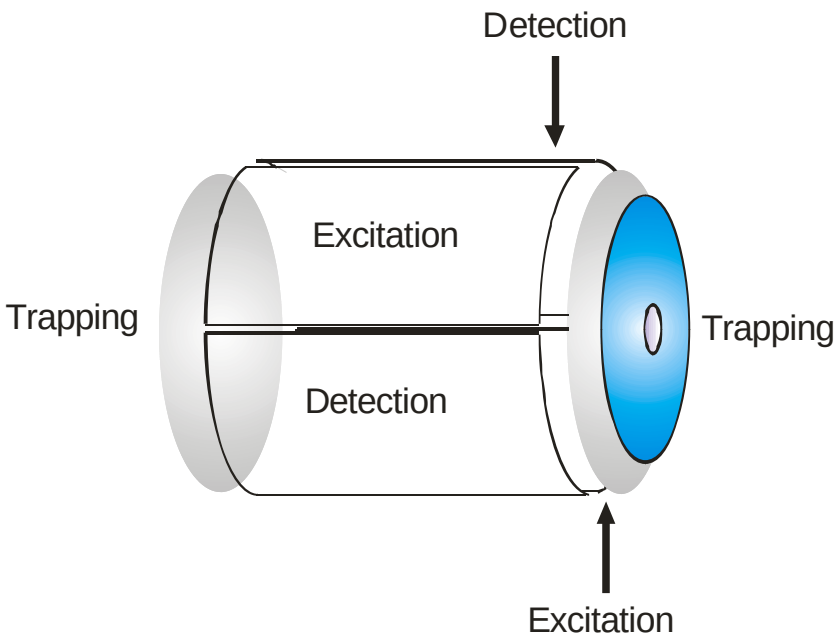

Figure 1.7 Cylindrical cell

Once ions are trapped, they undergo excitation by applying a sinusoidal voltage to the excite plates, which makes them move coherently to the same cyclotron radius.[116] Unlike other mass analyzers, FT-MS does not use any type of electron multiplier

detectors. FT-MS makes use of a resonance method to detect the image current signal generated by the ions in the cell. This signal is then digitized and converted to the frequency domain by Fourier transformation.[116, 118] The length of the signal depends on the time that the ions remain moving in coherence in the cell, and this is also directly related to the resolution achieved by this technique.[116, 132] The longer the duration of the transient, the better the resolution will be. Equation 3 shows this relationship between transient length *T*, and resolving power ***R,*** where f_c represents the cyclotron frequency.

Equation 3 $$R = \frac{f_c \cdot T}{2}$$

Performance of an FT-MS instrument depends on several factors, some of them are:[133]

- Lower pressures decreases collisional frequency, which in time provides longer transients
- Homogeneity of the electric and magnetic fields
- High magnetic field strength increases ion retention, resolution and high mass capability.

Figure 1.8 shows how the magnetic field strength affects different parameters of the FT-MS experiment. FT-MS provides high masses detection, high resolution and mass accuracy which are usually far better than those obtain by other mass analyzers. [96, 112, 134, 135] However, as any analytical tool, FT-MS also presents some limitations:

- Requires higher vacuum than other mass spectrometers (10^{-9} Torr, or lower)
- Slower measurements are required for good resolution (1 s vs 100 μs for a time-of-flight (TOF) measurement)
- Generally requires a higher level of expertise for instrument operation

Also, an important consideration in this technology is the phenomenon known as "space charge", which originates from the influence of the electric field on the ions trapped in the cell. This brings as consequence mass shifts that affects the mass accuracy of the experiment.[136, 137]

Tandem Fourier Transform - Mass Spectrometry (MS/MS): Tandem MS or MS/MS is based on the selection of a precursor ion from a mixture of masses, followed by its dissociation or reaction to produce smaller ions that are mass analyzed.[121] Tandem MS has become a powerful technique for obtaining protein sequence information. It allows the generation of peptide sequence tags that are used for highly refined database searches.[138] Advances in tandem MS introduced in the past years have produced new applications targeted to the biochemistry and medicinal fields. These include location of post-translational modifications and binding sites, in addition to sequence information and identification of proteins.[121, 139] For tandem MS applications, the ionization technique commonly used is ESI rather than MALDI, as ESI forms multiply charged ions that facilitates ion dissociation and the production of useful fragments,.[38, 70, 121]

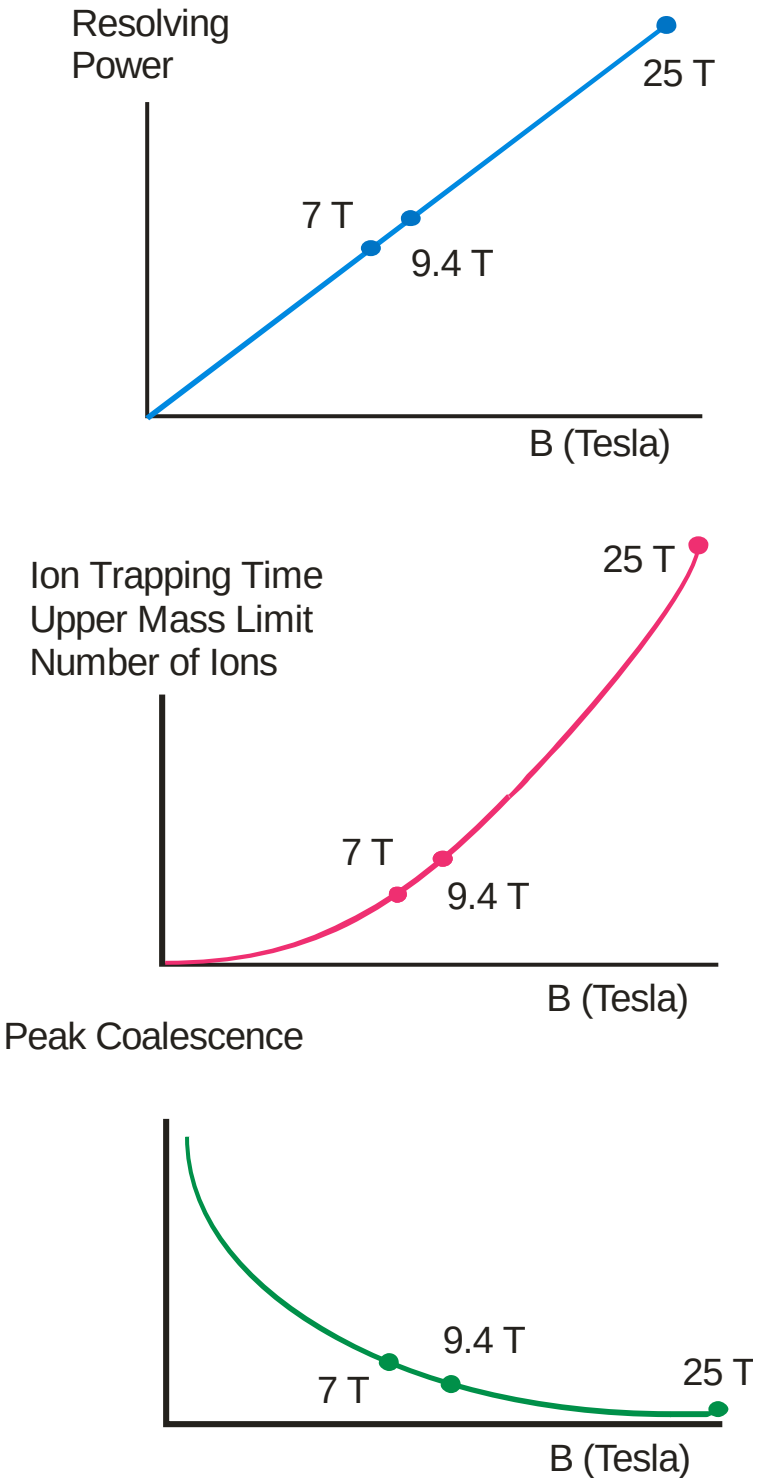

Figure 1.8 FTMS performance parameters as a function of applied magnetic field strength (adapted from Marshall [133])

A tandem mass spectrometer can be conceived either in space or in time; tandem MS in space involves coupling of two or more mass analyzers, and tandem MS in time is based on sequence of events performed in an ion storage or trap device.[38] In FT-MS, tandem MS can be performed as a series of events in the analyzer cell, since this is an efficient magnetic ion trap in which ions can be manipulated in various ways. There are several ion dissociation techniques for fragmentation of molecules used in FT-ICR tandem MS, some of these include, sustained off-resonance irradiation collision-induced dissociation (SORI-CID),[140, 141] on-resonance excitation CID (RE-CID),[142] infrared multiphoton dissociation (IRMPD),[143, 144] blackbody infrared radiative dissociation (BIRD),[145] electron capture dissociation (ECD),[146] and electron detachment dissociation (EDD).[147, 148] Each one of these techniques fragments the selected ions in a different fashion, by either low or high energy fragmentation pathways, and generates distinct product ions useful for obtaining detailed structural features of the molecules analyzed.[141, 149]

Hybrid Quadrupole / Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (Qq-FTMS): In the past years, advances in mass spectrometry technology have resulted in the introduction of hybrid instruments which give even more detailed sample characterization than standard tandem MS instruments.[149-152] One such hybrid instrument is the Qq-FTMS, in which ions can be mass selected by an external quadrupole, allowed to dissociate by CAD in a collision cell, and the products detected in the FT-ICR cell.[150] Often, an accumulating RF-only multipole device precedes the Qq stage. Initial external trapping of ions has proved to be advantageous to ESI-FTMS experiments, enhancing the signal-to-noise ratio and mass resolving power.[153] Selective

external ion accumulation exhibits several advantages over conventional accumulated trapping, such as: 1) higher efficiency of ion trapping, due to displacement of the trapping process from the FT-ICR cell to an external trap that can manage higher pressures; 2) improvement in the duty cycle of the FT-ICR instrument, since an external ion trap serves as a buffer between the source and the FT-ICR cell, duty cycle could reach 100% in optimized conditions; 3) expansion of the effective dynamic range of the instrument by ejecting higher abundance species and accumulating lower abundance ones for longer time.[150, 154, 155] Figure 1.9 shows a schematic diagram of the external accumulation interface and FT-ICR cell typically found in a hybrid Qq-FTMS instrument.

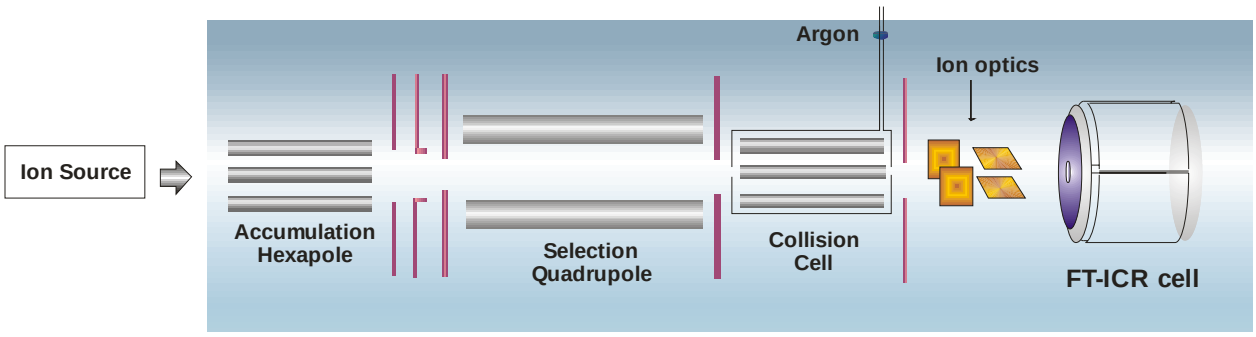

Figure 1.9 A schematic diagram of external accumulation interface and FT-ICR cell. The rf-only traps could be quadrupoles, hexapoles or octopoles.

FT-MS in Proteomics: FT-MS instrumentation provides high sensitivity and resolution, and excellent mass accuracy which makes them very useful for resolving individual peptides from complex samples, which is a very important goal for proteome studies.[156] Accurate measurement of single peptides measured by FT-MS along with easily obtainable constraints have been used to identify proteins by sequence database

searching.[157] FT-MS is also able to identify and locate post-translational modifications when it is combined with Electron Capture Dissociation (ECD), a fragmentation technique in which low energy electrons are introduced to interact with the present multiply protonated ions. This interaction produces radical cations that readily fragment.[158, 159] Since ECD produces significantly different types of fragment ions, it can be combined with other fragmentation methods as infrared multiphoton dissociation (IRMPD) and collision-induced dissociation (CID) to provide an even more complete analysis of complex protein mixtures.[160]

Chromatographic coupling: Proteomic studies usually deals with very complex mixtures of many peptides. Regardless of the very high mass resolution and mass accuracy of FT-MS, the large range of protein relative abundances can present a major difficulty for proteomic analyses. The maximum dynamic range for a single FT-MS mass spectrum is often limited to around 10^3. On the other hand, the dynamic range and sensitivity of FT-MS are limited by the charge capacity of either the cell or the external accumulation multipole.[20, 136] However, when FT-MS is combined with high resolution - liquid chromatography separations, the dynamic range of measurements increases to at least 10^6.[161, 162] The combination of liquid chromatography with FT-MS has been a very important advancement in proteomics. Further developments in separation techniques continue to improve the dynamic range and sensitivity of analyses.[20, 163, 164]

Scope of the present thesis: The next chapter describes the instrumentation and protocols used for the analyses performed. Chapter 3 describes the improvement of peptide identification in shotgun proteomics by use of cysteine mass defect labeling as

a constraint. Chapter 4 describes a different approach to trypsinolysis of proteins, an aqueous/organic solvent system which seems to improve the specificity of the digestion and, it also reduces significantly the time required for protein digestion. Finally Chapter 5 recounts conclusions pertinent to the work carried out and also shows future development planned for our approach to shotgun proteomics.

References

(1) Werner, T. Proteomics and regulomics: The yin and yang of functional genomics. *Mass Spectrom. Rev.* **2004**, *23*, 25-33.

(2) Stretton, A. O. W. The first sequence: Fred Sanger and insulin. *Genetics* **2002**, *162*, 527-532.

(3) Sanger, F. The Free Amino Groups of Insulin. *Biochem. J.* **1945**, *39*, 507-515.

(4) Biemann, K. The coming of age of mass spectrometry in peptide and protein chemistry. *Protein Sci.* **1995**, *4*, 1920-1927.

(5) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463-5467.

(6) Maxam, A.; Gilbert, W. A new method of sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 560-564.

(7) Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **2000**, *16*, 16-23.

(8) Edman, P. Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.* **1950**, *4*, 277-282,283-293.

(9) Steen, H.; Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nature Rev. Mol. Cell Biol.* **2004**, *5*, 699-711.

(10) O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, *250*, 4007-4021.

(11) Godovac-Zimmermann, J.; Brown, L. R. Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom. Rev.* **2001**, *20*, 1-57.

(12) Towbin, H.; Staehelin, T.; Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 4350-4354.

(13) Griffin, P.; Maccoss, M.; Blevins, R.; Aaronson, J.; Eng, J.; John R. Yates, I. Direct database searching with MALDI-PSD spectra of peptide. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1546-1551.

(14) Pappin, D.; Hojrup, P.; Bleasby, A. Rapid identification of proteins by peptide mass fingerprinting. *Curr. Biol.* **1993**, *3*, 327-332.

(15) Patterson, S. D. How much of the proteome do we see with discovery-based proteomics methods and how much do we need to see? *Curr. Proteomics* **2004**, *1*, 3-12.

(16) Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. genet. eng.* **1996**, *13*, 19-50.

(17) Goodlett, D. R.; Yi, E. C. Proteomics without polyacrilamide: qualitative and quantitative uses of tandem mass spectrometry in proteome analysis. *Funct. Integr. Genomics* **2002**, *2*, 138-153.

(18) Kettman, J. R.; Frey, J. R.; Lefkovits, I. Proteome, transcriptome and genome: top down or bottom up analysis? *Biomol. Eng* **2001**, *18*, 207-212.

(19) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994-999.

(20) Bogdanov, B.; Smith, R. D. Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.* **2005**, *24*, 168-200.

(21) Liu, H.; Sadygov, R. G.; John R. Yates, I. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76*, 4193-4201.

(22) John R. Yates, I. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **1998**, *33*, 1-19.

(23) Brock, A.; Horn, D. M.; Peters, E. C.; Shaw, C. M.; Ericson, C.; Phung, Q. T.; Salomon, A. R. An automated matrix-assisted laser desorption/ionization quadrupole Fourier transform ion cyclotron resonance mass spectrometer for "bottom-up" proteomics. *Anal. Chem.* **2003**, *75*, 3419-3428.

(24) Schirmer, E. C.; Florens, L.; Guan, T.; Yates III, J. R.; Gerace, L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **2003**, *301*, 1380-1382.

(25) Ideker, T.; Thorsson, V.; Ranis, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science* **2001**, *292*, 929-934.

(26) Gygi, S. P.; Han, D. K. M.; Gingras, A.-C.; Sonenberg, N.; Aebersold, R. Protein analysis by mass spectrometry and sequence database searching: Tools for cancer research in the post-genomic era. *Electrophoresis* **1999**, *20*, 310-319.

(27) Letokhov, V. S. Laser Photoionization Spectroscopy of Single Atoms and Molecules. *J. Mod. Opt.* **1985**, *32*, 1191-1210.

(28) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53-68.

(29) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.

(30) Karas, M.; Hillenkamp, F. Laser desorption of proteins with molecular mass exceeding 10,000 daltons. *Anal. Chem.* **1988**, *60*, 2299-2301.

(31) Yoshida, T.; Tanaka, K.; Ido, Y.; Akita, S.; Yoshida, Y. Detection of high mass molecular ions by laser desorption time-of-flight mass spectrometry. *J. Mass Spec. Soc. Jap.* **1988**, *36*, 59-69.

(32) Bahr, U.; Karas, M.; Hillenkamp, F. Analysis of biopolymers by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *Fresenius. J. Anal. Chem.* **1994**, *348*, 783-791.

(33) Beavis, R. C. Matrix-assisted ultraviolet laser desorption: Evolution and principles. *Org. Mass Spectrom.* **1992**, *27*, 653-659.

(34) Karas, M.; Bahr, U.; Gieumann, U. Matrix-assisted laser desorption ionization mass spectrometry. *Mass Spectrom. Rev.* **1991**, *10*, 335-357.

(35) Chen, X.; Carroll, J. A.; Beavis, R. C. Near-ultraviolet-induced matrix-assisted laser desorption/ionization as a function of wavelength. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 885-891.

(36) Gluckmann, M.; Karas, M. The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization. *J. Mass Spectrom.* **1999**, *34*, 467-477.

(37) Zenobi, R.; Knochenmuss, R. Ion formation in MALDI mass spectrometry. *Mass Spectrom. Rev.* **1998**, *17*, 337-366.

(38) Hoffmann, E. d.; Stroobant, V. *Mass Spectrometry: Principles and Applications*, Second Edition ed.: West Sussex, England, (2002).

(39) Fitzgerald, M. C.; Parr, G. R.; Smith, L. M. Basic matrixes for the matrix-assisted laser desorption/ionization mass spectrometry of proteins and oligonucleotides. *Anal. Chem.* **1993**, *65*, 3204-3211.

(40) Burton, R. D.; Watson, C. H.; Eyler, J. R.; Lang, G. L.; Powell, D. H.; Avery, M. Y. Proton affinities of eight matrices used for matrix-assisted laser desorption/ionization. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 443-446.

(41) Dale, M.; Knochenmuss, R.; Zenobi, R. Two-phase matrix-assisted laser desorption/ionization: matrix selection and sample pretreatment for complex anionic analytes. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 136-142.

(42) Schurenberg, M.; Dreisewerd, K.; Hillenkamp, F. Laser desorption/ionization mass spectrometry of peptides and proteins with particle suspension matrixes. *Anal. Chem.* **1999**, *71*, 221-229.

(43) Beavis, R. C.; Chaudhary, T.; Chait, B. T. alpha-cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **1992**, *27*, 156-158.

(44) Strupat, K.; Karas, M.; Hillenkamp, F. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1991**, *111*, 89-102.

(45) Sze, E. T. P.; Chan, T.-W. D.; Wan, G. Formulation of matrix solutions for use in matrix-assisted laser desorption/ionization of biomolecules. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 166-174.

(46) Vorm, O.; Roepstorff, P.; Mann, M. Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Anal. Chem.* **1994**, *66*, 3281-3287.

(47) Perera, I. K.; Perkins, J.; Kantartzoglou, S. Spin-coated samples for high resolution matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of large proteins. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 180-187.

(48) Dai, Y.; Whittal, R. M.; Li, L. Two-Layer Sample Preparation: A Method for MALDI-MS Analysis of Complex Peptide and Protein Mixtures. *Anal. Chem.* **1999**, *71*, 1087-1091.

(49) Xiang, F.; Beavis, R. C.; Ens, W. A method to increase contaminant tolerance in protein matrix-assisted laser desorption/ionization by the fabrication of thin protein-doped polycrystalline films. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 199-204.

(50) Zhang, J.; LaMotte, L.; Dodds, E. D.; Lebrilla, C. B. Atmospheric Pressure MALDI Fourier Transform Mass Spectrometry of Labile Oligosaccharides. *Anal. Chem.* **2005**, *77*, 4439 - 4447.

(51) Ho, Y.-P.; Fenselau, C. Metastable decay of peptide ions on a Fourier transform mass spectrometer equipped with an external ion source. *J. Mass Spectrom.* **2000**, *35*, 183-188.

(52) Stemmler, E. A.; Buchanan, M. V.; Hurst, G. B.; Hettich, R. L. Analysis of modified oligonucleotides by matrix-assisted laser desorption/ionization Fourier transform mass spectrometry. *Anal. Chem.* **1995**, *67*, 2924 - 2930.

(53) Cancilla, M. T.; Penn, S. G.; Carroll, J. A.; Lebrilla, C. B. Coordination of Alkali Metals to Oligosaccharides Dictates Fragmentation Behavior in Matrix Assisted Laser Desorption Ionization/Fourier Transform Mass Spectrometry. *J. Am. Chem. Soc.* **1996**, *118*, 6736 - 6745.

(54) Laiko, V. V.; Moyer, S. C.; Cotter, R. J. Atmospheric Pressure MALDI/Ion Trap Mass Spectrometry. *Anal. Chem.* **2000**, *72*, 5239-5243.

(55) Daniel, J. M.; Laiko, V. V.; Doroshenko, V. M.; Zenobi, R. Interfacing liquid chromatography with atmospheric pressure MALDI-MS. *Anal. Bioanal. Chem.* **2005**, *383*, 895-902.

(56) Baykut, G.; Jertz, R.; Witt, M. Matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry with pulsed in-source collision gas and in-source ion accumulation. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1238-1247.

(57) O'Connor, P. B.; Costello, C. E. A high pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry ion source for thermal stabilization of labile biomolecules. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1862-1868.

(58) Garrett, T. J.; Yost, R. A. Analysis of intact tissue by intermediate-pressure MALDI on a linear ion trap mass spectrometer. *Anal. Chem.* **2006**, *78*, 2465-2469.

(59) Kellersberger, K. A.; Tan, P. V.; Laiko, V. V.; Doroshenko, V. M.; Fabris, D. Atmospheric pressure MALDI-Fourier transform mass spectrometry. *Anal. Chem.* **2004**, *76*, 3930-3934.

(60) Baldwin, M. A.; Medzihradszky, K. F.; Lock, C. M.; Fisher, B.; Settineri, T. A.; Burlingame, A. L. Matrix-assisted laser desorption/ionization coupled with quadrupole/orthogonal acceleration time-of-flight mass spectrometry for protein discovery, identification, and structural analysis. *Anal. Chem.* **2001**, *73*, 1707 - 1720.

(61) Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M.; Ens, W.; Standing, K. G. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1047-1057.

(62) Krutchinsky, A. N.; Loboda, A. V.; Spicer, V. L.; Dworschak, R.; Ens, W.; Standing, K. G. Orthogonal injection of matrix-assisted laser desorption/ionization ions into a time-of-flight spectrometer through a collisional damping interface. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 508-518.

(63) Rappsilber, J.; Moniatte, M.; Nielsen, M. L.; Podtelejnikov, A. V.; Mann, M. Experiences and perspectives of MALDI MS and MS/MS in proteomic research. *Int. J. Mass Spectrom.* **2003**, *226*, 223-237.

(64) Donegan, M.; Tomlinson, A. J.; Nair, H.; Juhasz, P. Controlling matrix suppression for matrix-assisted laser desorption/ionization analysis of small molecules. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1885-1888.

(65) Go, E. P.; Prenni, J. E.; Wei, J.; Jones, A.; Hall, S. C.; Witkowska, H. E.; Shen, Z.; Siuzdak, G. Desorption/Ionization on Silicon Time-of-Flight/Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 2504-2506.

(66) Hatsis, P.; Brombacher, S.; Corr, J.; Kovarik, P.; Volmer, D. A. Quantitative analysis of small pharmaceutical drugs using a high repetition rate laser matrix-assisted laser/desorption ionization source. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2303-2309.

(67) Wenzel, R. J.; Matter, U.; Schultheis, L.; Zenobi, R. Analysis of megadalton ions using cryodetection MALDI time-of-flight mass spectrometry. *Anal. Chem.* **2005**, *77*, 4329-4337.

(68) Fenn, J. B. Ion formation from charged droplets: roles of geometry, energy, and time. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 523-612.

(69) Fenn, J. B.; Mann, M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64.

(70) Cole, R. B. *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications*: Canada, (1997).

(71) Yamashita, M.; Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* **1984**, *88*, 4451 - 4459.

(72) Yamashita, M.; Fenn, J. B. Negative ion production with the electrospray ion source. *J. Phys. Chem.* **1984**, *88*, 4671-4675.

(73) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular beams of macroions. *J. Chem. Phys.* **1968**, *49*, 2240-2249.

(74) Meng, C. K.; Mann, M.; Fenn, J. B. Electrospray ionization of some polypeptides and small proteins. *Proceedings of the 36th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, San Francisco, CA* **1988**, 773.

(75) Gaskell, S. J. Electrospray: Principles and Practice. *J. Mass Spectrom.* **1997**, *32*, 677-784.

(76) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization-principles and practice. *Mass Spectrom. Rev.* **1990**, *9*, 37-70.

(77) Cech, N. B.; Enke, C. G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **2001**, *20*, 362-387.

(78) Mora, J. F. d. l.; Berkel, G. J. V.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. Electrochemical processes in electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2000**, *35*, 939-952.

(79) Loo, J. A.; Udseth, H. R.; Smith, R. D. Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. *Anal. Biochem.* **1989**, *179*, 404-412.

(80) Ikonomou, M. G.; Blades, A. T.; Kebarle, P. Investigations of the electrospray interface for liquid chromatography/mass spectrometry. *Anal. Chem.* **1990**, *62*.

(81) Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. New developments in biochemical mass spectrometry: electrospray ionization. *Anal. Chem.* **1990**, *62*, 882-899.

(82) Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R. Principles and practice of electrospray ionization - mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev.* **1991**, *10*, 359-452.

(83) Enke, C. G. A Predictive Model for Matrix and Analyte Effects in Electrospray Ionization of Singly-Charged Ionic Analytes. *Anal. Chem.* **1997**, *69*, 4885-4893.

(84) Berkel, G. J. V.; Zhou, F.; Aronson, J. T. Changes in bulk solution pH caused by the inherent controlled-current electrolytic process of an electrospray ion source. *Int. J. Mass Spectrom. Ion Processes* **1997**, *162*, 55-67.

(85) Iribarne, J. V.; Thomson, B. A. Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* **1976**, *64*, 2287-2294.

(86) Cooks, R. G.; Ouyang, Z.; Takats, Z.; Wiseman, J. M. Ambient mass spectrometry. *Science* **2006**, *311*, 1566-1570.

(87) Kebarle, P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J. Mass Spectrom.* **2000**, *35*, 804-817.

(88) Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. Fourier-transform mass spectrometry of large molecules by electrospray ionization. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86*, 9075-9078.

(89) Bruce, J. E.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. Large Molecule Characterization Based upon Individual Ion Detection with Electrospray Ionization-FTICR Mass Spectrometry. *Anal. Chem.* **1998**, *70*, 519-525.

(90) Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. Fourier-Transform Mass-Spectrometry of Large Molecules by Electrospray Ionization. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 9075-9078.

(91) Mendes, M. A.; Souza, B. M. d.; Santos, L. D. d.; Palma, M. S. Structural characterization of novel chemotactic and mastoparan peptides from the venom of the social wasp Agelaiapallipes pallipes by high-performance liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 636-642.

(92) Mendes, M. A.; Souza, B. M. d.; Santos, L. D. d.; Santos, K. S.; Palma, M. S. Analyzing glycerol-mediated protein oligomerization by electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 2636-2642.

(93) Edmonds, C. G.; Joseph, A. L.; Loo, R. R.; Udseth, H. R.; Barinaga, C. J.; Smith, R. D. Application of electrospray ionization mass spectrometry and tandem mass spectrometry in combination with capillary electrophoresis for biochemical investigations. *Biochem. Soc. Trans.* **1991**, *19*, 943-947.

(94) Chitta, R. K.; Gross, M. L. Electrospray Ionization-Mass Spectrometry and Tandem Mass Spectrometry Reveal Self-Association and Metal-Ion Binding of Hydrophobic Peptides: A Study of the Gramicidin Dimer. *Biophysical Journal* **2004**, *86*, 473-479.

(95) Loo, J. A.; G., E. C.; Smith, R. D. Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* **1990**, *248*, 201-204.

(96) Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. Attomole-Sensitivity Electrospray Source for Large-Molecule Mass Spectrometry. *Anal. Chem.* **1995**, *67*, 3802-3805.

(97) Jurascheka, R.; Dülcksa, T.; Kara, M. Nanoelectrospray—more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 300-308.

(98) Wilm, M. S.; Mann, M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167-180.

(99) Comisarow, M. B.; Marshall, A. G. The early development of fourier transform ion cylcotron resonance (FT-ICR) spectroscopy. *J. Mass Spectrom.* **1996**, *31*, 581-585.

(100) Lawrence, E. O.; Livigston, M. S. The production of high speed light ions without the use of high voltages. *Phys. Rev.* **1932**, *40*, 19-35.

(101) Hipple, J. A.; Sommer, H.; Thomas, H. A. A Precise Method of Determining the Faraday by Magnetic Resonance. *Phys. Rev.* **1949**, *76*, 1877-1878.

(102) Bloch, F.; Hansen, W. W.; Packard, M. Nuclear Induction. *Phys. Rev.* **1946**, *69*, 127.

(103) Ernst, R. R.; Anderson, W. A. Application of Fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instrum.* **1966**, *37*, 93.

(104) Marshall, A. G. Ion Cyclotron Resonance and Nuclear Magnetic Resonance Spectroscopies: Magnetic Partners for Elucidation of Molecular Structure and Reactivity. *Acc. Chem. Res.* **1996**, *29*, 307-316.

(105) Marshall, A. G. Milestones in fourier transform ion cyclotron resonance mass spectrometry technique development. *Int. J. Mass Spectrom.* **2000**, *200*, 331-356.

(106) Comisarow, M. B.; Marshall, A. G. Resolution-enhanced Fourier transform ion cyclotron resonance spectroscopy. *J. Chem. Phys.* **1975**, *62*, 293-295.

(107) Marshall, A. G.; Comisarow, M. Fourier transform ion cyclotron resonance [FT-ICR] spectroscopy. *Chem. Phys. Lett.* **1974**, *25*, 282-283.

(108) Comisarow, M. B.; Marshall, A. G. Theory of Fourier transform ion cyclotron resonance mass spectroscopy. I. Fundamental equations and low-pressure line shape. *J. Chem. Phys.* **1976**, *64*, 110-119.

(109) Dienes, T.; Pastor, S. J.; Schürch, S.; Scott, J. R.; Yao, J.; Cui, S.; Wilkins, C. L. Fourier transform mass spectrometry - advancing years (1992-mid. 1996). *Mass Spectrom. Rev.* **1996**, *15*, 163-211.

(110) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.

(111) Guan, S.; Marshall, A. G. Ion traps for Fourier transform ion cyclotron resonance mass spectrometry: principles and design of geometric and electric configurations. *Int. J. Mass Spectrom. Ion Processes* **1995**, *146-147*, 261-296.

(112) McLafferty, F. W. High-resolution tandem FT mass spectrometry above 10 kDa. *Acc. Chem. Res.* **1994**, *27*, 379-386.

(113) Shi, S. D.-H.; Drader, J. J.; Hendrickson, C. L.; Marshall, A. G. Fourier transform ion cyclotron resonance mass spectrometry in a high homogeneity 25 tesla resistive magnet. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 265-268.

(114) Sparkman, D. Mass Spectrometry PittCon 2006. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 873-884.

(115) Zhang, L.-K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. Accurate mass measurements by Fourier transform mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, 286-309.

(116) Park, Y.; Lebrilla, C. B. Application of Fourier transform ion cyclotron resonance mass spectrometry to oligosaccharides. *Mass Spectrom. Rev.* **2005**, *24*, 232-264.

(117) Vartanian, V. H.; Anderson, J. S.; Laude, D. A. Advances in trapped ion cells for Fourier transform ion cyclotron resonance mass spectrometry. *Mass Spectrom. Rev.* **1995**, *14*, 1-19.

(118) Amster, I. J. Fourier transform mass spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1325-1337.

(119) Lorenz, S. A.; Maziarz, E. P.; Wood, T. D. Electrospray ionization Fourier transform mass spectrometry of macromolecules: The first decade. *Appl. Spectrosc.* **1999**, *53*, 18A-36A.

(120) Smith, R. D. Evolution of ESI-mass spectrometry and Fourier transform ion cyclotron resonance for proteomics and other biological applications. *Int. J. Mass Spectrom.* **2000**, *200*, 509-544.

(121) Williams, E. R. Tandem FTMS of large biomolecules. *Anal. Chem.* **1998**, *70*, 179A-185A.

(122) Winger, B. E.; Hofstadler, S. A.; Bruce, J. E.; Udseth, H. R.; Smith, R. D. High-Resolution Accurate Mass Measurements of Biomolecules Using a New

Electrospray-Ionization Ion-Cyclotron Resonance Mass-Spectrometer. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 566-577.

(123) Hettich, R.; Buchanan, M. Structural Characterization of Normal and Modified Oligonucleotides by Matrix-Assisted Laser Desorption Fourier-Transform Mass-Spectrometry. *J. Am. Soc. Mass Spectrom.* **1991**, *2*, 402-412.

(124) Hettich, R. L.; Buchanan, M. V. Matrix-Assisted Laser Desorption Fourier-Transform Mass-Spectrometry for the Structural Examination of Modified Nucleic-Acid Constituents. *Int. J. Mass Spectrom. Ion Processes* **1991**, *111*, 365-380.

(125) Hettich, R. L.; Buchanan, M. V. Investigation of Uv Matrix-Assisted Laser Desorption Fourier-Transform Mass-Spectrometry for Peptides. *J. Am. Soc. Mass Spectrom.* **1991**, *2*, 22-28.

(126) McIver, R. T.; Li, Y. Z.; Hunter, R. L. Ftms Method for High-Resolution Matrix-Assisted Laser-Desorption. *Int. J. Mass Spectrom. Ion Processes* **1994**, *132*, L1-L7.

(127) Koster, C.; Castoro, J. A.; Wilkins, C. L. High-Resolution Matrix-Assisted Laser Desorption Ionization of Biomolecules by Fourier-Transform Mass-Spectrometry. *J. Am. Chem. Soc.* **1992**, *114*, 7572-7574.

(128) Marshall, A. G.; Hendrickson, C. L. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *Int. J. Mass Spectrom.* **2002**, *215*, 59-75.

(129) Campbell, V. L.; Guan, Z.; Vartanian, V. H.; Laude, D. A. Cell geometry considerations for the Fourier transform ion cyclotron resonance mass spectrometry remeasurement experiment*. Anal. Chem.* **1995**, *67*, 420-425.

(130) Beu, S. C.; David A. Laude, J. Open trapped ion cell geometries for Fourier transform ion cyclotron resonance mass spectrometry*. Int. J. Mass Spectrom. Ion Processes* **1992**, *112*, 215-230.

(131) Frankevich, V.; Zenobi, R. Dynamic ion trapping in a cylindrical open cell for Fourier transform ion cyclotron resonance mass sepctrometry*. Int. J. Mass Spectrom.* **2001**, *207*, 57-67.

(132) Grosshans, P. B.; Marshall, A. G. Theory of ion cyclotron resonance mass spectrometry: resonant excitation and radial ejection in orthorhombic and cylindrical ion traps*. Int. J. Mass Spectrom. Ion Processes* **1990**, *100*, 347-379.

(133) Marshall, A. G.; Guan, S. Advantages of High Magnetic Field for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*. Rapid Commun. Mass Spectrom.* **1996**, *10*, 1819-1823.

(134) Chen, R.; Cheng, X.; Mitchell, D. W.; Hofstadler, S. A.; Wu, Q.; Rockwood, A. L.; Sherman, M. G.; Smith, R. D. Trapping, detection, and mass determination of coliphage T4 DNA ions of 108 Da by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **1995**, *67*, 1159 - 1163.

(135) Li, Y.; Robert T. McIver, J. High-accuracy molecular mass determination for peptides and proteins by Fourier transform mass spectrometry*. Anal. Chem.* **1994**, *66*, 2077-2083.

(136) Easterling, M. L.; Mize, T. H.; Amster, I. J. Routine Part-per-Million Mass Accuracy for High- Mass Ions: Space-Charge Effects in MALDI FT-ICR. *Anal. Chem.* **1999**, *71*, 624-632.

(137) Taylor, P. K.; Amster, I. J. Space charge effects on mass accuracy for multiply charged ions in ESI–FTICR. *Int. J. Mass Spectrom.* **2003**, *222*, 351-361.

(138) Roepstorff, P. Mass spectrometry in protein studies from genome to function. *Curr. Opin. Biotechnol.* **1997**, *8*, 6-13.

(139) Schwartz, J. C.; Wade, A. P.; Enke, C. G.; Cooks, R. G. Systematic delineation of scan modes in multidimensional mass spectrometry. *Anal. Chem.* **1990**, *62*, 1809-1818.

(140) Senko, M. W.; Speir, J. P.; McLafferty, F. W. Collisional activation of large multiply charged ions using Fourier transform mass spectrometry. *Anal. Chem.* **1994**, *66*, 2801-2808.

(141) Herrmann, K. A.; Somogyi, Á.; Wysocki, V. H.; Drahos, L.; Vékey, K. Combination of Sustained Off-Resonance Irradiation and On-Resonance Excitation in FT-ICR. *Anal. Chem.* **2005**, *77*, 7626-7638.

(142) Cody, R. B.; Freiser, B. S. High-resolution detection of collision-induced dissociation fragments by Fourier transform mass spectrometry. *Anal. Chem.* **1982**, *54*, 1431-1433.

(143) Zhang, J.; Schubothe, K.; Li, B.; Russell, S.; Lebrilla, C. B. Infrared Multiphoton Dissociation of O-Linked Mucin-Type Oligosaccharides. *Anal. Chem.* **2005**, *77*, 208-214.

(144) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994**, *66*, 2809-2815.

(145) Price, W. D.; Schnier, P. D.; Williams, E. R. Tandem Mass Spectrometry of Large Biomolecule Ions by Blackbody Infrared Radiative Dissociation. *Anal. Chem.* **1996**, *68*, 859-866.

(146) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* **2000**, *72*, 563-573.

(147) Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A. Electron detachment dissociation of peptide di-anions: an electron–hole recombination phenomenon. *Chem. Phys. Lett.* **2001**, *342*, 299-302.

(148) Nielsen, M. L.; Budnika, B. A.; Haselmanna, K. F.; Zubarev, R. A. Tandem MALDI/EI ionization for tandem Fourier transform ion cyclotron resonance mass spectrometry of polypeptides. *Int. J. Mass Spectrom.* **2003**, *226*, 181-187.

(149) Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312*, 212-217.

(150) Belov, M. E.; Nikolaev, E. N.; Anderson, G. A.; Auberry, K. J.; Harkewicz, R.; Smith, R. D. Electrospray ionization-Fourier transform ion cyclotron mass spectrometry using ion preselection and external accumulation for ultrahigh sensitivity. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 38-48.

(151) Patrie, S. M.; Charlebois, J. P.; Whipple, D.; Kelleher, N. L.; Hendrickson, C. L.; Quinn, J. P.; Marshall, A. G.; Mukhopadhyay, B. Construction of a hybrid quadrupole/Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for versatile MS/MS above 10 kDa. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1099-1108.

(152) Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. A top-down method for the determination of residue-specific solvent accessibility in proteins. *J. Mass Spectrom.* **2004**, *39*, 322-328.

(153) Senko, M. W.; Hendrickson, C. L.; Emmett, M. R.; Shi, S. D. H.; Marshall, A. G. External accumulation of ions for enhanced electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 970-976.

(154) Bruce, J. E.; Anderson, G. A.; Smith, R. D. "Colored" noise waveforms and quadrupole excitation for the dynamic range expansion of Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **1996**, *68*, 534-541.

(155) Wang, Y.; Shi, S. D. H.; Hendrickson, C. L.; Marshall, A. G. Mass-selective ion accumulation and fragmentation in a linear octopole ion trap external to a Fourier transform ion cyclotron resonance mass spectrometer. *Int. J. Mass Spectrom.* **2000**, *198*, 113-120.

(156) Witt, M.; Fuchser, J.; Baykut, G. Fourier transform ion cyclotron resonance mass spectrometry with nanolc/microelectrospray ionization and matrix-assisted laser desorption/ionization: analytical performance in peptide mass fingerprint analysis. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 553-561.

(157) Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal. Chem.* **2000**, *72*, 1112 - 1118.

(158) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of American Chemical Society* **1998**, *120*, 3265-3266.

(159) Syrstada, E. A.; Turecek, F. Toward a general mechanism of electron capture dissociation. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 208-224.

(160) Constans, A. Putting a new spin on FT-MS. *The Scientist* **2003**, *17*, 46.

(161) Shen, Y.; Tolic', N.; Masselon, C.; Paša-Tolic', L.; Camp II, D. G.; Lipton, M. S.; Anderson, G. A.; Smith, R. D. Nanoscale proteomics. *Anal. Bioanal. Chem.* **2004**, *378*, 1037-1045.

(162) Smith, R. D.; Shen, Y.; Tang, K. Ultrasensitive and Quantitative Analyses from Combined Separations-Mass Spectrometry for the Characterization of Proteomes. *Acc. Chem. Res.* **2004**, *37*, 269-278.

(163) Shen, Y.; Toli, N.; Zhao, R.; Pasa-Tolic, L.; Li, L.; Berger, S. J.; Harkewicz, R.; Anderson, G. A.; Belov, M. E.; Smith, R. D. High-Throughput Proteomics Using High-Efficiency Multiple-Capillary Liquid Chromatography with On-Line High-Performance ESI FTICR Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 3011-3021.

(164) Belov, M. E.; Anderson, G. A.; Mark A. Wingerd; Udseth, H. R.; Tang, K.; Prior, D. C.; Swanson, K. R.; Buschbach, M. A.; Strittmatter, E. F.; Ronald J. Moore; Smith, R. D. An Automated High Performance Capillary Liquid Chromatography-Fourier

Transform Ion Cyclotron Resonance Mass Spectrometer for High-Throughput Proteomics. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 212-232.

CHAPTER 2

EXPERIMENTAL

This chapter will provide information about the methodologies and instruments used to perform the experiments whose results are shown and discussed in chapter 3 and 4. Specifically, the chapter describes methods for derivatizing proteins, performing proteolytic digestion, HPLC separation of peptides, MALDI – FTMS and HPLC-ESI-FTMS/MS experiments.

I. Sample Preparation: This is an essential stage in the analysis process. Peptide analysis by mass spectrometry requires adequate sample for signal detection and little background from contaminating peptides. Also, samples that are undergoing liquid chromatographic (LC) separation need to contain low amounts of organic modifier and salts.[1] Different conditions and approaches were tested in order to optimize: denaturing of protein mixture, sample removal of low molecular compounds and, derivatization of cysteines by 2,4-dibromo-(2-iodo) acetanilide. The following proteins standards were used for the aforementioned purpose and they were analyzed by MALDI-FTMS:

- Bovine Serum Albumin (Calbiochem, San Diego, CA): MW: 66433.4 Da. 35 cysteine residues.
- β-Lactoglobulin (Sigma-Aldrich, St. Louis, MO): MW: 19883,4 Da. 7 cysteine residues.
- Ovalbumin (Sigma-Aldrich, St. Louis, MO): MW: 42881.5 Da. 6 cysteine residues.
- Carbonic Anhydrase (Sigma-Aldrich, St. Louis, MO): MW: 28678.7 Da. 0 cysteine residues. Used as a negative control.

First of all, denaturing of the sample was performed by addition of 6 M guanidine hydrochloride (Fisher Scientific, Pittsburgh, PA) and / or application of heat. No significant differences were found on protein coverage or intensity of the signal obtained between the samples denatured by heat only and those that contained the denaturing agent. Therefore, it was determined for the treatment of all subsequent samples that the denaturing step would consist of heating only at 95 °C for 15 minutes. Secondly, removal of low molecular components and large particle size material is an important step to prevent clogging of the LC column during the separation process and to reduce the background signal obtained by mass spectrometry.[1] Two different techniques were used in order to determine which one would suit better our experiments. The first one consisted of the use of Microcon® centrifugal filters to remove particles smaller than 3,000 Da and the second approach used centrifugal size exclusion chromatography by allowing the sample to pass through a column filled with Sephadex G-25 (Aldrich, St. Louis, MO), an inert media composed of macroscopic beads which retains molecules with a molecular weight lower than 5,000 Da. Analysis of the data obtained showed better signal and higher protein coverage for the samples cleaned by centrifugal size exclusion chromatography than those filtered by Microcon®. Consequently, this approach was included in the protocol used for preparation of the samples.

Regarding derivatization of cysteine residues by the labeling reagent 2,4-dibromo-(2-iodo) acetanilide, several reactions conditions were tested, including amount of reagent used, pH, temperature and length of reaction and, organic solvent used to dissolve the reagent. The best results were obtained when a 100-fold molar excess of the reagent dissolved in methanol was added to samples, at pH 8 for 90 minutes in the

dark at room temperature. Figure 2.1 shows the sample preparation protocol used for all proteome samples of *Methanococcus maripaludis* analyzed and discussed in chapter 3.

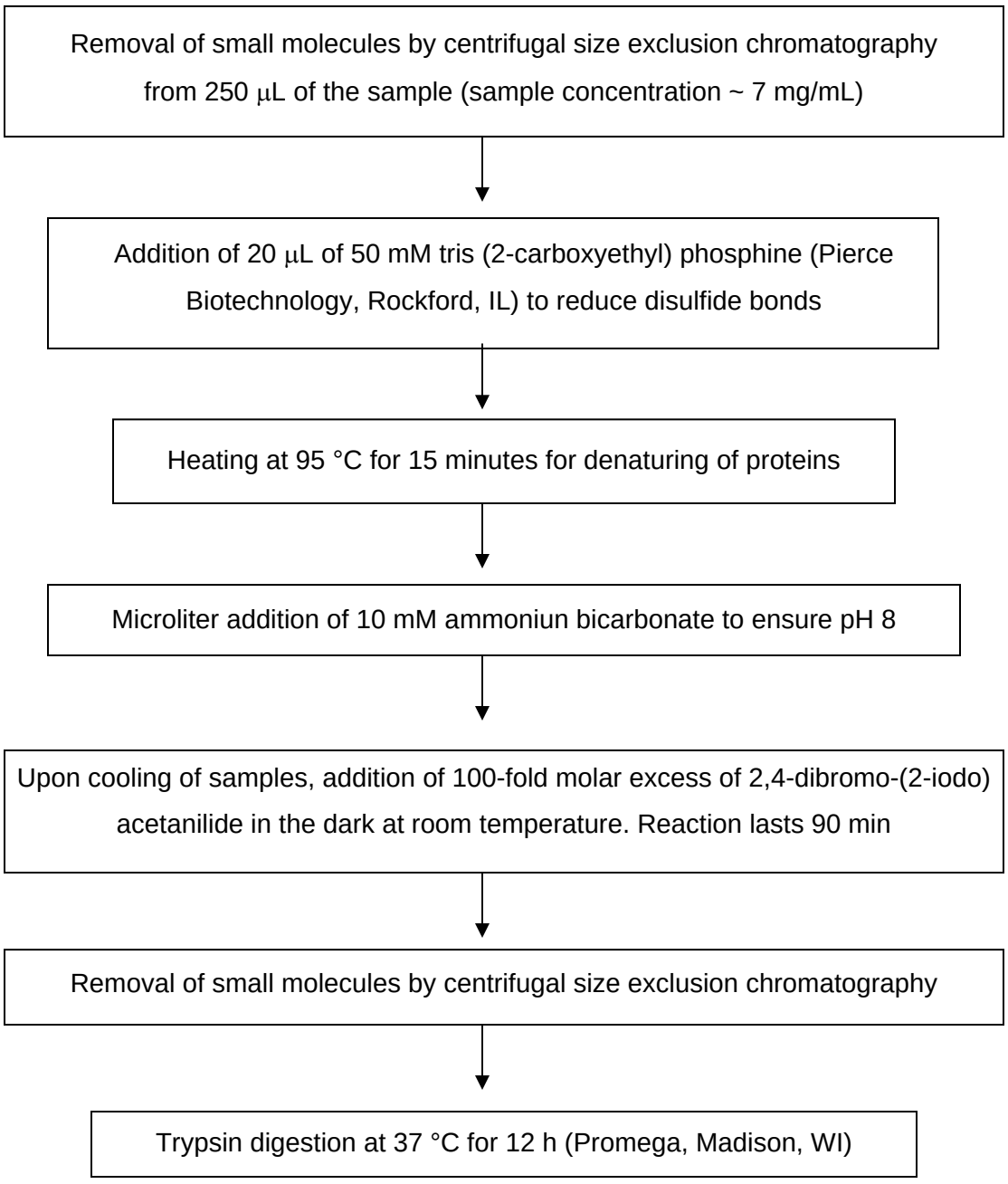

Figure 2.1. Sample preparation protocol for mass defect labeling experiments

Instrumentation, MALDI – FTMS analysis: Analysis of peptides mixtures by MALDI-FTMS required the prior separation off-line of the sample by high performance liquid chromatography. We will first describe the experimental conditions and instrumentation used for the separation followed by the correspondent explanation for the MALDI-FTMS experiment.

II. High Performance Liquid Chromatography (HPLC): Separation of peptide mixtures was performed using an UltiMate Plus, FAMOS system by Dionex (Sunnyvale, CA). Figure 2.2 shows the actual instrument used for the experiments, which consist of four modules: FAMOS™, Switchos™, UltiMate™ and Probot™. The FAMOS™ module (Figure 2.3) is a fully automated micro autosampler which was used to inject volumes of 20 μL of sample by an injection loop of the same volume. The sample tray is equipped with Peltier cooling which keeps samples at approximately 6° C. The Switchos™ module is used for pre-concentration and clean-up of the sample. It consists of two 10 port valves with a high precision loading pump and a four channel solvent selection valve. UltiMate™ module is in charge of mixing the mobile phase, which is achieved by using two high -precision reciprocating pumps that generate flow rates ranging from 50 nL/min up to 200 μL/min. It is also equipped with a quaternary low pressure gradient former that allows mixing of up to three solvents. The solvents used for HPLC separation are filtered prior to use by 0.45 μm membrane disc filters (Pall Corporation, Ann Arbor, MI) plus once the solvents were in the UltiMate^{TM} module, they were degassed by a helium sparging device.

Figure 2.2 UltiMate^{TM} HPLC system

Figure 2.3 Micro autosampler FAMOS^{TM}

A flow rate of 0.3 μL / min was used to elute the peptides from the analytical column. Two different stationary-phase, analytical columns were used: a) 75 μm i.d. x 15 cm, C18 PepMap100, 3 μm, 100Å; and b) 75 μm i.d. x 15 cm, C8 PepMap100, 3 μm, 100Å by LC Packings – Dionex (Sunnyvale, CA). The gradient used for off-line separation consisted of two mobile phases, mobile phase A: water/acetonitrile/ trifluoroacetic acid (98:2:0.1 by volume); mobile phase B: acetonitrile.

Mobile phase B increased concentration from 0 to 100% over 90 min, and elution at 100% B continued for 30 minutes to ensure total elution of the components of the sample. This HPLC system is also equipped with an UV detector which allows multi UV-trace recording. The UV-traces were recorded at wavelength 214 nm. The eluate was collected onto a stainless steel MALDI target at 60 second intervals using the Probot™ module (LC Packings – Dionex) which is a high precision x / y / z robot for micro fraction collection of the column eluate (Figure 2.4). After the fractions were collected, deposition of the MALDI matrix onto each spot was performed. The choice of matrix is known to affect the response obtained by MALDI mass spectrometry.[2] The matrix used for all MALDI experiments was 2,5-dihydrohybenzoic acid which, unlike other matrices, is soluble in water as well as organic solvents and is more tolerant towards contaminations such as salts and/or detergents.[3, 4] This matrix solution is made 1 M in a mix of acetonitrile / water / trifluoroacetic acid (50 / 50 / 0.1 by volume). The deposition method used is known as dried droplet method, which consists of mixing of the matrix and sample solutions, deposition onto plate and, air drying of solvents.[5]

Figure 2.4 Micro fraction collector Probot™

III. Intermediate Pressure Matrix-Assisted Laser Desorption Ionization – Fourier Transform Mass Spectrometry (IP-MALDI-FTMS), 7.0 Tesla: Analyses of protein standards and proteome samples described in chapter 3 were performed using an IP-MALDI-FTMS system from Bruker Daltonics, Inc. (Billerica, MA) equipped with a 7.0 Tesla magnet (Figure 2.5). Once the MALDI target plate is spotted with the different fractions of the sample, it is then place into the SCOUT 100 MALDI source, in which desorption of ions is induced by a pulsed nitrogen laser at 337 nm. This desorption of ions occurs at intermediate pressure (~ 1 mTorr) by pulsing of argon gas into the source in order to suppress metastable decomposition.[6] Every spot in the MALDI target plate can be analyzed by the X-Y manipulator which allows the precise alignment of each MALDI target spot with the laser beam. The laser is fired 12 times for each scan and, in a typical experiment, 12 scans are co-added for each recorded spectrum.[7] After the laser desorption event (12 shots), the pulsed valve closes to allow a 100 msec pump-

down, and ions are accumulated in the hexapole. Then, the ions are extracted from the hexapole and transferred to the analyzer cell by electrostatic ion optics. As the ions enter the ICR cell, Sidekick™ is applied in order to push the ions off-axis, therefore, improving trapping efficiency and detection.[8]

Figure 2.5. Bruker MALDI-FTMS 7.0 Tesla

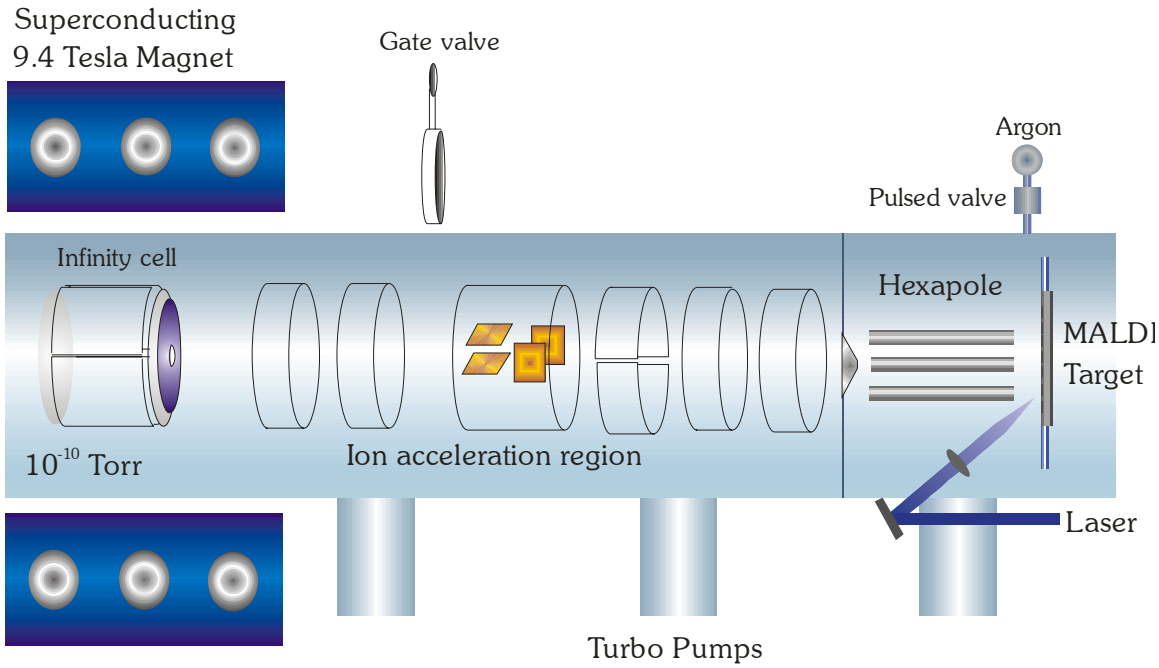

Figure 2.6. Components of MALDI-FTMS 9.4 Tesla

IV. Intermediate Pressure Matrix-Assisted Laser Desorption Ionization – Fourier Transform Mass Spectrometry (IP-MALDI-FTMS), 9.4 Tesla: Shotgun proteomics studies utilizing matrix assisted laser desorption ionization (MALDI) as the ionization technique described in chapter 4 were performed on the 9.4 T BioApex FT-ICR mass spectrometer (Figure 2.7), also equipped with an intermediate pressure SCOUT 100 source from Bruker Daltonics (Billerica, MA). Figure 2.6 shows a schematic of the components of the IP-MALDI-FTMS 9.4 T instrument. Operational conditions are similar to those described for the 7.0 Tesla instrument.

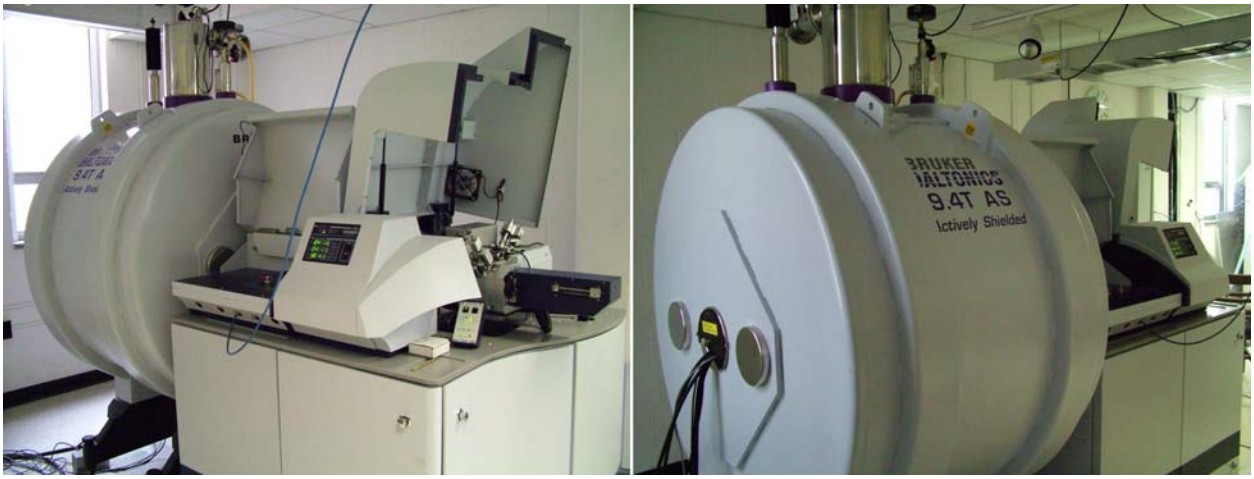

Figure 2.7. Bruker MALDI-FTMS 9.4 Tesla

VI. Electrospray Ionization – Quadrupole – Fourier Transform Mass Spectrometry (ESI-Qh-FTMS), 7.0 Tesla: Attempts for MS/MS experiments described in chapter 4 were performed by an Apex-Qe-FTMS from Bruker Daltonics Inc. (Billerica, MA), (Figure 2.9). This instrument is equipped with an Apollo II electrospray ionization source and a quadrupole mass analyzer collision cell which allows precursor selection and collisional activated dissociation of ions. An schematic of this instrument is shown in figure 2.10. For a typical MS experiment the sample is introduced into the source from the HPLC system and, charged electrosprayed droplets are generated by applying a high voltage difference between a stainless steel spray needle and the stainless steel capillary. Once the charged particles are formed, they go through the nozzle skimmer region in which excess neutrals particles are removed. Then, ions move to the first hexapole in which they experience ion accumulation, subsequently they move into the quadrupole, which can serve as a mass filter or as an ion guide, and then to the second hexapole that can function as a collision cell or simply as an ion guide when operated in an rf-only mode. Finally, the ions are guided by electrostatic ion optics to the ICR cell

which is operated at pressures ~ 10^{-10} Torr. For a standard MS/MS experiment, the ions undergo collision with argon gas in the collision cell which produces fragments from the precursor ions and then these ions are transferred to the analyzer ICR cell.

Figure 2.9. Bruker ESI-FTMS 7.0 Tesla

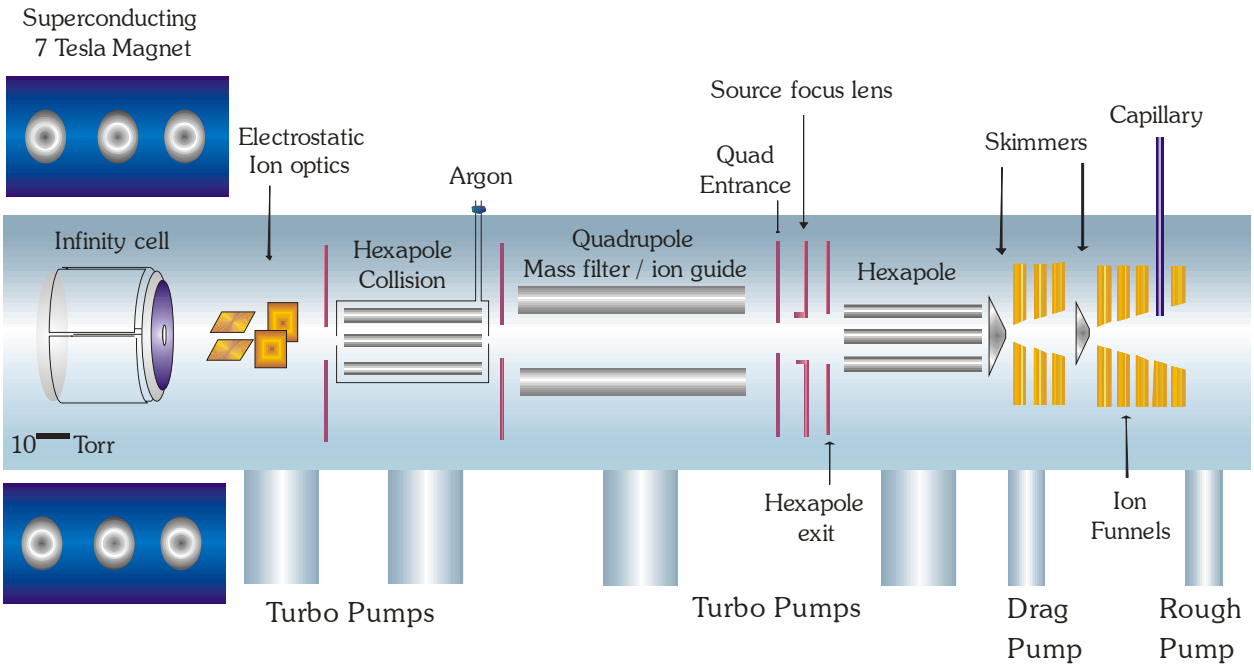

Figure 2.10. Components of ESI-FTMS 7.0 Tesla

References

(1) Visser, N. F. C.; Lingeman, H.; Irth, H. Sample preparation for peptides and proteins in biological matrices prior to liquid chromatography and capillary zone electrophoresis. *Analytical & Bioanalytical Chemistry* **2005**, *382*, 535-558.

(2) Cohen, S. L.; Chait, B. T. Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Analytical Chemistry* **1996**, *68*, 31-37.

(3) Strupat, K.; Karas, M.; Hillenkamp, F. 2,5-dihydrohybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry. *International journal of mass spectrometry and ion processes* **1991**, *111*, 89-102.

(4) Gusev, A. I.; Wilkinson, W. R.; Proctor, A.; Hercules, D. M. Improvement of signal reproducibility and matrix / comatrix effects in MALDI analysis. *Analytical Chemistry* **1995**, *67*, 1034-1041.

(5) Karas, M.; Hillenkamp, F. Laser desorption of proteins with molecular mass exceeding 10,000 daltons. *Analytical Chemistry* **1988**, *60*, 2299-2301.

(6) Garrett, T. J.; Yost, R. A. Analysis of intact tissue by intermediate-pressure MALDI on a linear ion trap mass spectrometer. *Analytical Chemistry* **2006**, *78*, 2465-2469.

(7) Wong, R. L.; Amster, I. J. Combining low and high mass ion accumulation for enhancing shotgun proteome analysis by accurate mass measurement. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 205-212.

(8) Hofstadler, S. A.; Sannes-Lowery, K. A.; Griffey, R. H. m/z-selective infrared multiphoton dissociation in a Penning trap using sidekick trapping and an rf-tickle pulse*. Rapid Communications in Mass Spectrometry* **2001**, *15*, 945-951.

CHAPTER 3

MASS DEFECT LABELING OF CYSTEINE FOR IMPROVING PEPTIDE ASSIGNMENT IN SHOTGUN PROTEOMIC ANALYSES

Introduction: The primary goal of a proteomic analysis is to be able to systematically identify and quantify the majority of proteins expressed in a cell or tissue.[1, 2] The conventional approach for conducting proteome-wide studies is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE),[3] where a large number of proteins can be separated on the basis of their isoelectric point and molecular weight. Although 2D-PAGE technology has been the chief technology for proteomic analysis to date, it has recognized limitations, such as a bias toward the most abundant proteins and dynamic range and protein solubility issues that complicate the detection and separation of low-abundance and hydrophobic proteins.[4] In recent years, a number of researchers have focused on improving proteomic analyses via the development of shotgun proteomic methods.[5-9] These methods identify and quantify proteins which have not been separated prior to digestion. The basis of this approach is to perform a batch digestion of an unseparated protein mixture, to separate the resulting peptides by one or more dimensions of liquid chromatography, and to identify the proteins from which the peptides derive by mass spectrometry analysis.[8]

Two mass spectrometry approaches for shotgun proteomic analysis have been reported. First is the use of tandem mass spectrometry to generate fragmentation data which can be used by search engines to identify the protein origin of the peptides.[2, 6, 8, 10, 11] These methods are able to detect and identify a wide variety of protein classes including those with extremes in isoelectric point, molecular weight, abundance, and hydrophobicity. However, these methods are time consuming and produce very large data sets, as they require the generation of a fragmentation spectrum for each peptide

in a mixture that contains thousands of components. A second approach is the use of accurate mass measurement to identify proteins. If the molecular masses of the peptides from a batch digest are measured with high enough mass measurement accuracy (MMA), a reasonable fraction of their masses can uniquely identify them by comparison to a list of masses for all of the possible proteolytic peptides predicted from an *in silico* digest of the genome. Other experimental information can be used to increase the fraction of identified peptides, for example, HPLC retention time.[11] Methods that combine MMA with the MS/MS capabilities have also been reported.[11, 12]

In this paper, we describe a new method for improving the specificity of protein identification by accurate mass measurement of peptides. The improvement is based upon the derivatization of a specific amino acid with a reagent that changes the mass defect of the peptide. For the purpose of discussion, we refer to the mass defect as the difference between the exact monoisotopic mass of a compound and its nominal molecular weight, that is the weight based on the nucleon values of the most abundant isotope of each element, e.g. 12 amu for C, 16 amu for O, etc. Peptides are composed principally of elements from the first two rows of the periodic table. These elements have mass defects that lie in the range of +/- 0.008 amu. The mass defect of peptide molecules is approximately +0.05 amu per 100 amu of molecular weight, i.e. a 1 kDa peptide has a mass defect of approximately +0.5 amu, and a 2 kDa peptide has a mass defect of 1 amu. The positive mass defect is a result of the high stoichiometric proportion of hydrogen atoms in a peptide molecular formula (the hydrogen mass defect is +0.0078 amu). Although peptide molecules have significant mass defects because of the large number of atoms from which they are assembled, the distribution of mass

defects is generally narrow, causing peptide molecular weights at any given nominal mass to occupy only a small portion of a unit mass. This is illustrated in Figure 3.1, which shows a histogram of monoisotopic masses for the 125 possible tryptic peptides (up to 1 missed cleavage) with molecular weights between 1500 and 1503 that one predicts for all proteins in the sequence database for the organism *Methanococcus maripaludis*. This organism has 1722 open reading frames, which is about average for a single cell organism, and has approximately 95,700 predicted tryptic peptides above with molecular weights 700 amu (allowing up to 1 missed cleavage), and the predicted peptide mass distribution is similar to that of any organism. Because of the narrow distribution of mass defects for peptides, their molecular weights cluster into one-third of the total mass space causing masses to overlap, and reducing the specificity of a peptide mass for identifying the protein origin. Greater specificity would be possible if the peptide masses were distributed more evenly across the mass scale.

The narrow distribution of mass defects for a compound class has been noted previously by other researchers in mass spectrometry. Perfluoroalkanes have distinctly different mass defects that do not overlap those of most other organic compounds, and have long been employed as internal calibrants for exact mass measurements.[13] The components of complex mixtures of small molecules can be assigned a Kendrick mass defect value, which allows homologous series to be assigned to various compound classes. [14-19] Labeling the N-terminus of a protein with a compound that alters the mass defect is used to distinguish the N-terminal peptide fragments from C-terminal and internal fragments produced by nozzle-skimmer dissociation of intact proteins, and is the basis of a commercial reagent (IDBEST™) and process.[20, 21] We report here a

method for altering the mass defects of a selected fraction of the peptides in a batch digest of a proteome so that the resulting peptides can be more readily identified by accurate mass measurement.

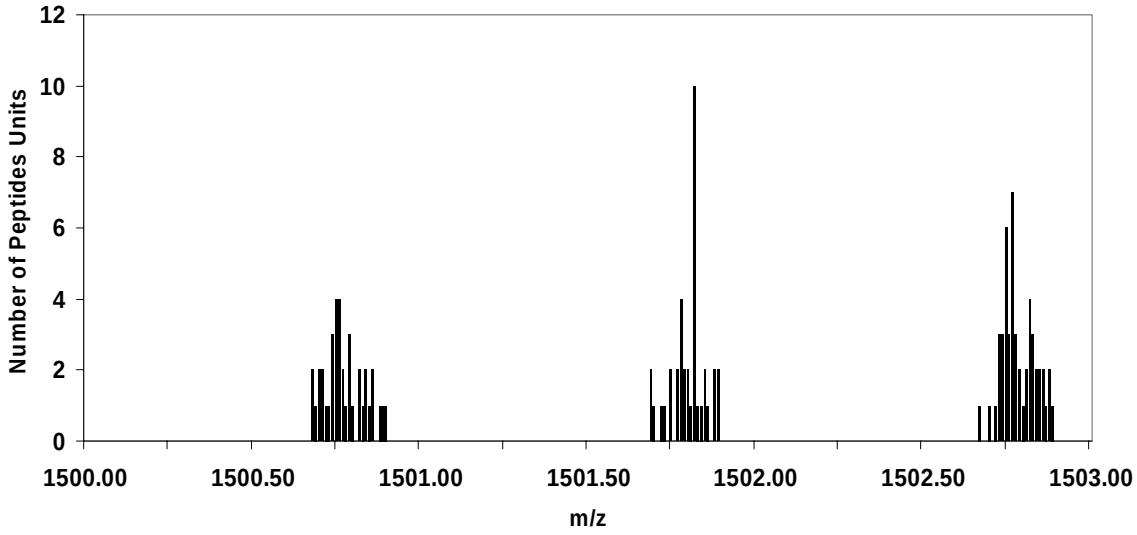

Figure 3.1. Histogram of the molecular weight distribution of the predicted tryptic peptides of *M. maripaludis* over the range 1500-1503 Da, Illustrating the distribution of mass defects of peptides. The bin size is 0.01 amu. Peptide masses are observed to cluster in approximately one-third of the available mass space.

Experimental Section:

Reagent Synthesis: The cysteine-alkylating reagent, 2,4-dibromo-(2'-iodo) acetanilide, was prepared by addition of 5.4 mmol (0.46 mL) of oxalyl chloride (Acros Organics, Morris Plains, NJ) in 2.7 mL of dry dichloromethane to 1 equivalent (1 g) of 2-iodoacetic acid (Acros Organics) in 4 mL of dry dichloromethane. This mixture was stirred for 3 h at 0 ^{o}C under nitrogen to yield a pink solution (2-iodoacetylchloride.) This

solution was added dropwise with stirring to 1 equivalent (1.3 g) of 2,4-dibromoaniline (Acros Organics) in 10 mL of dry dichloromethane. A white crude solid appeared as a precipitate, and was collected by filtration and purified by recrystallization from hot water to give the final product in 70% yield. The structure of the purified 2,4-dibromo-(2'-iodo) acetanilide was confirmed by ^{1}H-NMR and mass spectrometry (NMR and MS spectra are included as supplementary data.) All reagents and solvents were used as purchased without further purification.

Protein Labeling: The labeling of cysteine before versus after tryptic digestion was compared for a number of proteins. We consistently find that the best results are obtained by labeling before digestion, as it is easier to remove the excess labeling reagent from a protein solution than from a solution of lower molecular weight peptides. Each protein standard was dissolved in alkaline solution (10 mM ammonium bicarbonate) to make a 1 mg/mL solution, and denatured by heating at 95 ^{o}C. Disulfide bonds were reduced by addition of tris (2-carboxyethyl) phosphine (Pierce Biotechnology, Rockford, IL). The protein then underwent reaction with a 100 fold molar excess of 2,4-dibromo-(2'-iodo) acetanilide at pH 8 for 90 minutes in the dark at room temperature. Prior to trypsin digestion, the derivatized protein was subjected to centrifugal size exclusion chromatography using a 3 mL spin column packed with Sephadex G-25 (Aldrich, St. Louis, MO) to remove excess 2,4-dibromo-(2'-iodo) acetanilide. Trypsin digestion was performed under standard conditions (Promega, Madison WI), i.e. at 37 ^{o}C, pH 7, for 18 hours.

Proteome Labeling: Whole cell lysates were extracted from *Methanococcus maripaludis* that was grown on minimal media with ammonium sulfate as the sole source of nitrogen. Cells were grown using ammonium sulfate both with the naturally occurring isotopic composition (99.6% ^{14}N, 0.4% ^{15}N) and with 98% ^{15}N-enrichment. The cells were concentrated by centrifugation at 10000 rpm for 30 minutes; lysis of the cells was performed with a French press. DNA was digested and removed from the extract by adding DNAase to the sample followed by centrifugation. Equal amounts of protein extracts were mixed together before batch trypsinolysis. Prior to denaturing and labeling of the proteome, small molecules were removed by centrifugal size exclusion spin columns packed with Sephadex G-25. Subsequent treatment of the sample followed the procedure described above for labeling of the protein standards. Total protein concentrations were determined spectrophotometrically measuring at 562 nm using a bicinchoninic acid protein assay kit (Pierce, Rockford, IL).

Mass Spectrometry: Samples were analyzed by matrix assisted laser desorption/ionization (MALDI) Fourier transform ion cyclotron resonance (FTICR) mass spectrometry using a 7 Tesla magnet (Bruker Daltonics Inc, Billerica, MA). This instrument is equipped with a SCOUT 100 MALDI source which desorbs ions at elevated pressure (~ 1 mTorr) to suppress metastable decomposition. Conditions for operation of the FTICR MS were similar to those reported previously,[22] and external mass calibration was established using a peptide mixture generated by tryptic digestion of chicken egg albumin (Sigma, St. Louis, MO). The MALDI matrix was 2,5-dihydroxybenzoic acid (DHB) (Lancaster, Pelham, NH).

High Performance Liquid Chromatography: Separations of peptide mixtures were performed on an UltiMate™ Plus, FAMOS by Dionex (Sunnyvale, CA). Reverse-phase columns used were: 1) 75 µm i.d. x 15 cm, C18 PepMap100, 3 µm, 100Å; and 2) 75 µm i.d. x 15 cm, C8 PepMap100, 3 µm, 100Å (LC Packings–Dionex). Mobile phase A was water/acetonitrile/ trifluoroacetic acid (98:2:0.1 by volume), and mobile phase B was acetonitrile. A gradient from 0 - 100% B over 90 min was used at an approximate column flow of 300 nL/min; the total run time was 120 minutes. The eluate was collected onto a stainless steel MALDI target at 60 second intervals using a Probot™ Micro Fraction Collector (LC Packings–Dionex). The MALDI matrix was added after the fraction collection was completed, requiring resuspension of the dried, fractionated peptides in 0.5 µL of the matrix solution (1 M DHB in 50:50:0.1 water:acetonitrile:trifluoroacetic acid.)

Protein Identification: The molecular weight of the peptides and their nitrogen stoichiometry were determined from the MALDI-FTICR mass spectrum. The number of nitrogen atoms in each peptide was determined from the mass separation between the monoisotopic peak of the peptide containing the natural distribution of ^{14}N / ^{15}N and the monoisotopic peak of the ^{15}N-enriched counterpart. The data was analyzed using software that was developed in-house to identify the proteins from which the peptides were derived. The software compares the experimentally determined molecular weight and nitrogen stoichiometry with values in a look-up table that is populated with the predicted tryptic fragments (up to 1 missed cleavage) for all protein sequences for the organism in question. A peptide is considered to be identified when there is only one

predicted peptide that meets the following match criteria: the predicted peptide has a mass that lies within a specified mass tolerance of the measured molecular weight, and it has the same nitrogen stoichiometry as the measured value. Peptide identifications were made using a mass tolerance of 10 ppm.

Results and Discussion:

Mass Defect Labels: The narrow distribution of mass defects that is characteristic of peptides arises in part from the small mass defect of their component elements, and from the uniform stoichiometry of peptides. Table 3.1 shows the mass defect of the elements which comprise proteins. As can be seen, their mass defects are small (less than 10 mmu for H, C, N, and O, and around 28 mmu for S). The average elemental ratio for an amino acid residue is $C_{4.9384}H_{7.7583}N_{1.3577}O_{1.4773}S_{0.0417}$.[23] Given that nitrogen (mass defect = +3.1 mmu) and oxygen (mass defect = -5.1 mmu) have comparable stoichiometric values, their mass defects tend to cancel in a peptide. One can see that the mass defect of a peptide is principally due to hydrogen, and that the distribution of mass defects comes from the narrow distribution of elemental stoichiometries. Figure 3.1 suggests that the distribution of mass defects at any nominal mass is roughly one third of an amu. One can calculate the distribution of mass defects at each nominal mass for the tryptic peptides of all proteins in a database, and we have done this for peptides with masses from 700-3000 that derive from the proteins in the *M. maripaludis* database. The composite distribution of mass defects around the average value at each nominal mass is shown in Figure 3.2. As can be seen, peptides masses

occupy only one-third of the available mass scale, which causes some of the predicted masses to overlap, even at a mass tolerance of 10 ppm. To shift some of the peptide masses to the region of the mass scale that is unpopulated, we alter the mass defects of a portion of the peptides by derivatizing a less frequently occurring amino acid, cysteine, with a reagent which introduces a large mass defect. This is accomplished by introducing a heavy element with a large mass defect into the elemental composition, in this case, bromine.

Element	Mass Defect (amu)
^{12}C	0
^{1}H	0.0078
^{16}O	-0.0051
^{15}N	0.0031
^{32}S	-0.0279

Table 3.1. Mass difference from nucleon value of the most abundant isotope of the elements found in proteins.

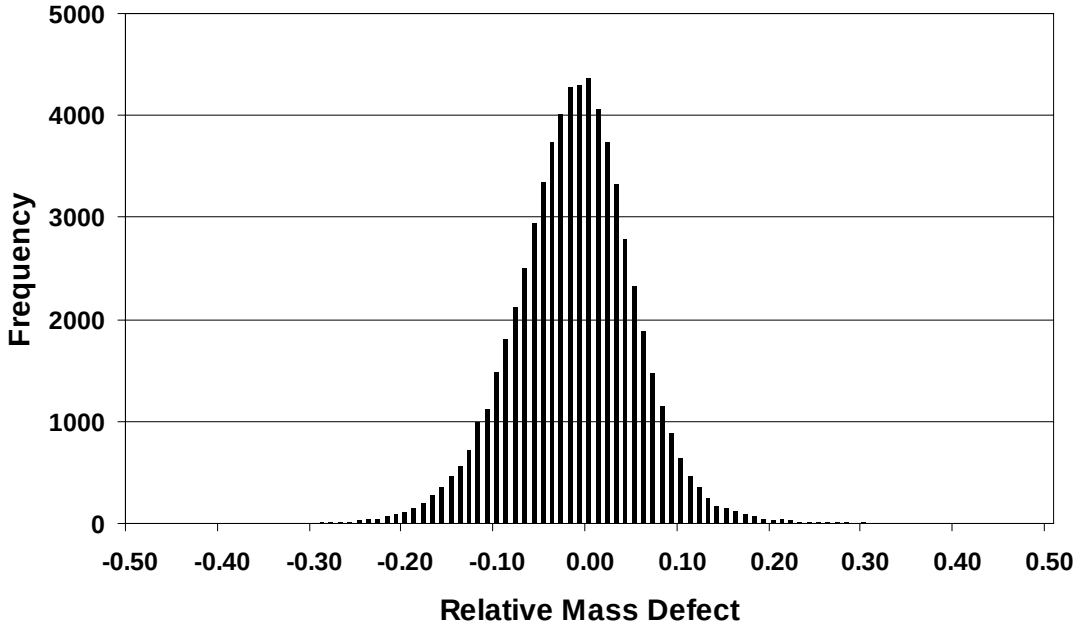

Figure 3.2. The composite distribution of mass defects for all tryptic peptides of *M. maripaludis* with molecular weights between 700-3500 amu. The horizontal axis is the mass difference (amu) between a peptide's mass defect and the average mass defect for all peptides of the same nominal mass.

Derivatization of a specific amino acid with a compound that affects the mass defect will yield two sets of peptides; unlabeled peptides with typical mass defects, and labeled peptides with masses that lie in a region of the mass scale that is unoccupied by underivatized peptides. To achieve this end, we have synthesized a reagent that we refer to as a mass defect label (MDL) which derivatizes a specific type of amino acid and which changes the mass of the resulting product in a manner that makes it easy to distinguish derivatized peptides from other peptides of the same nominal mass. The

ideal tagging reagent will (1) have high reaction specificity for a low abundance amino acid such as cysteine or tryptophan, (2) introduce a mass defect shift of 0.3–0.6 amu, (3) be stable to the chemical and physical conditions necessary for derivatization and mass spectral characterization, and (4) have no deleterious effects on peptide solubility or ionization efficiency. The MDL reported here is a derivative of iodoacetamide and reacts specifically with cysteine, as shown in Figure 3.3.

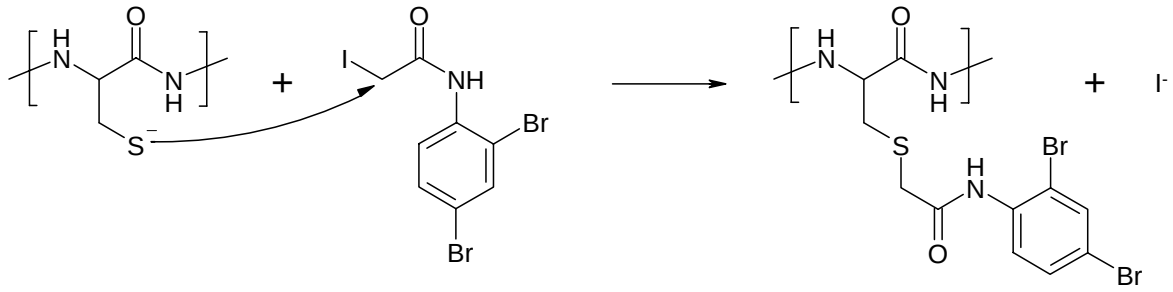

Figure 3.3. Mechanism of an alkylation reaction of a cysteine-containing peptide with 2,4-dibromo-(2'-iodo) acetanilide.

Figure 3.4 illustrates the change in the mass defect distribution for the tryptic peptides that is expected from cysteine-derivatization of all the proteins in the *M. maripaludis* sequence database. The derivatized peptide masses occupy a region in which no unlabeled peptides are found, approximately 0.3 amu below the unlabeled peptides. Because only 15-20% of tryptic peptides contain cysteine, fewer peptides will occupy the new region of mass, and therefore there is a lower probability of mass

overlap for predicted peptides. This suggests that a higher proportion of derivatized peptides can be identified by their mass compared to underivatized peptides.

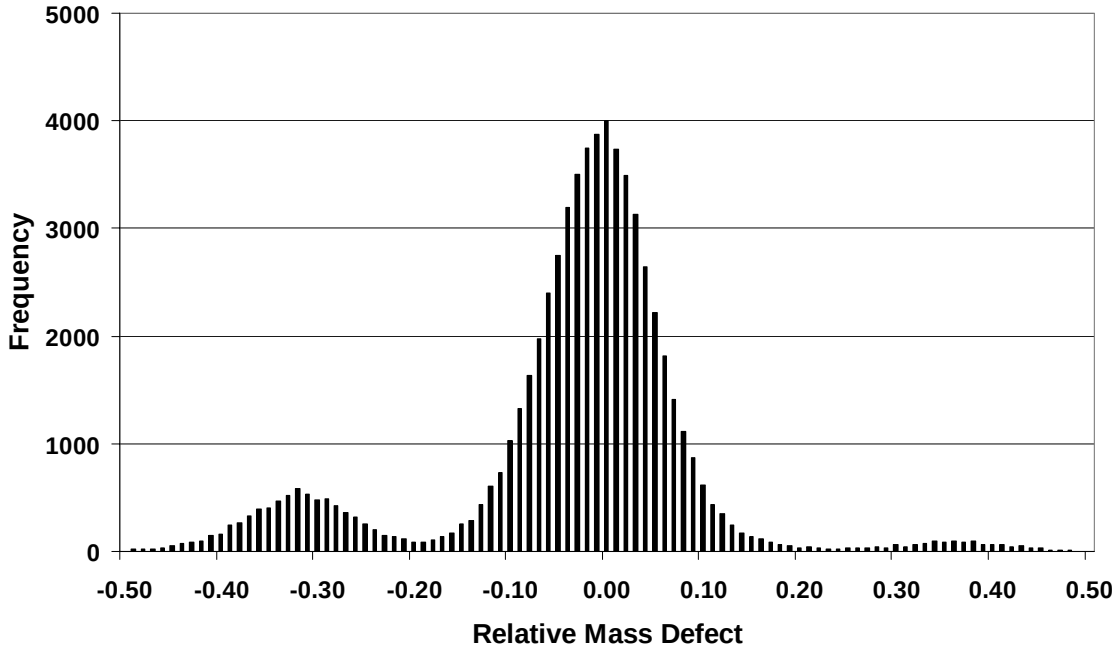

Figure 3.4. The composite distribution when all the cysteine-containing peptides have been labeled. The central distribution corresponds to all peptides that do not contain cysteine. All singly labeled cysteine-containing peptides appear in the smaller distribution centered at -0.30 amu. Doubly labeled cysteine-containing peptides appear at +0.40 amu.

Labeling of Protein Standards: Several protein standards were tested, including bovine serum albumin, β-lactoglobulin, ovalbumin, and carbonic anhydrase. These proteins underwent derivatization of their cysteine residues with the MDL,

digestion by trypsin, and analysis by MALDI-FTMS. Mass defect labeled peptides could be identified both by their mass defect values and by the isotope pattern that is characteristic of the presence of two bromine atoms.

Figure 3.5 shows the calculated isotopic distribution of a peptide (BSA 445-458) that is labeled by the mass defect reagent and compares the distribution to that of the corresponding unlabeled peptide. The use of chlorine isotope patterns to identify derivatized cysteine-containing peptides in a proteomics assay has been reported previously.[24] Here, we do not use the isotopic pattern to establish that derivatization has occurred. The mass defect of the resulting peptide provides this information. However, it is important that the unusual isotopic pattern is taken into consideration when assigning the monoisotopic peak. Figure 3.6 shows a mass spectrum of the tryptic peptides of bovine serum albumin that has been derivatized with the MDL; peaks corresponding to labeled peptides are identified with a square. As can be seen in the mass spectrum, many of the abundant peaks in the mass spectrum are from derivatized peptides, demonstrating that the MDL does not adversely affect the detectability of the peptides. No non-derivatized cysteine-containing peptides were found in the mass spectra for any of the protein tryptic digests that were tested, suggesting that the derivatization reaction was complete. Bovine serum albumin contains 35 cysteines, and 32 labeled cysteine residues were observed in the mass spectra of the tryptic peptides. Interestingly, in the underivatized control spectrum, only five cysteine-containing peptides were observed, suggesting that this derivatization increases the detectability of the cysteine-containing peptides. For β-lactoglobulin, 5 out of 7 possible cysteines were observed in their labeled state, and for ovalbumin, 3 out of 6 labeled cysteines were observed. Bovine

carbonic anhydrase II, which does not have a cysteine residue, served as a negative control. No labeled peptides were found in the mass spectrum of its tryptic digest. Overall, these data suggest that the reaction of the MDL reagent is specific for cysteine residues, quantitative in reactivity (no underivatized cysteines was observed), and has no adverse effect on the detectability of the derivatized peptides by MALDI mass spectrometry.

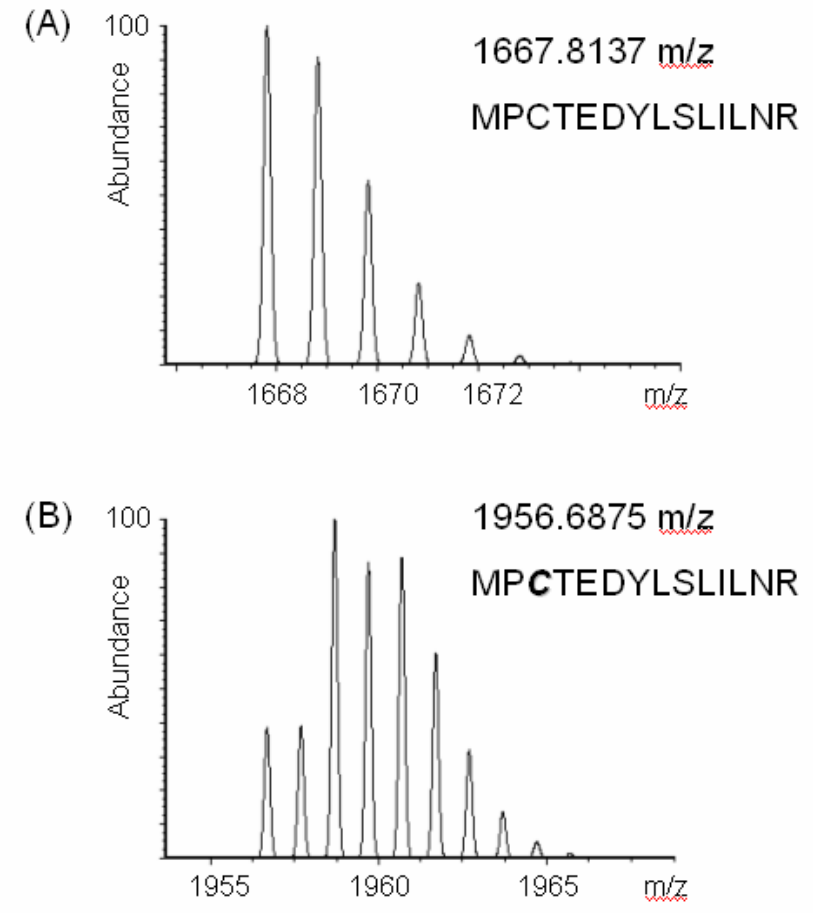

Figure 3.5. Calculated isotopic pattern for the peptide MPCTEDYLSLILNR from bovine serum albumin (residues 445-458), (A) without and (B) with the dibromoacetanilide mass defect label.

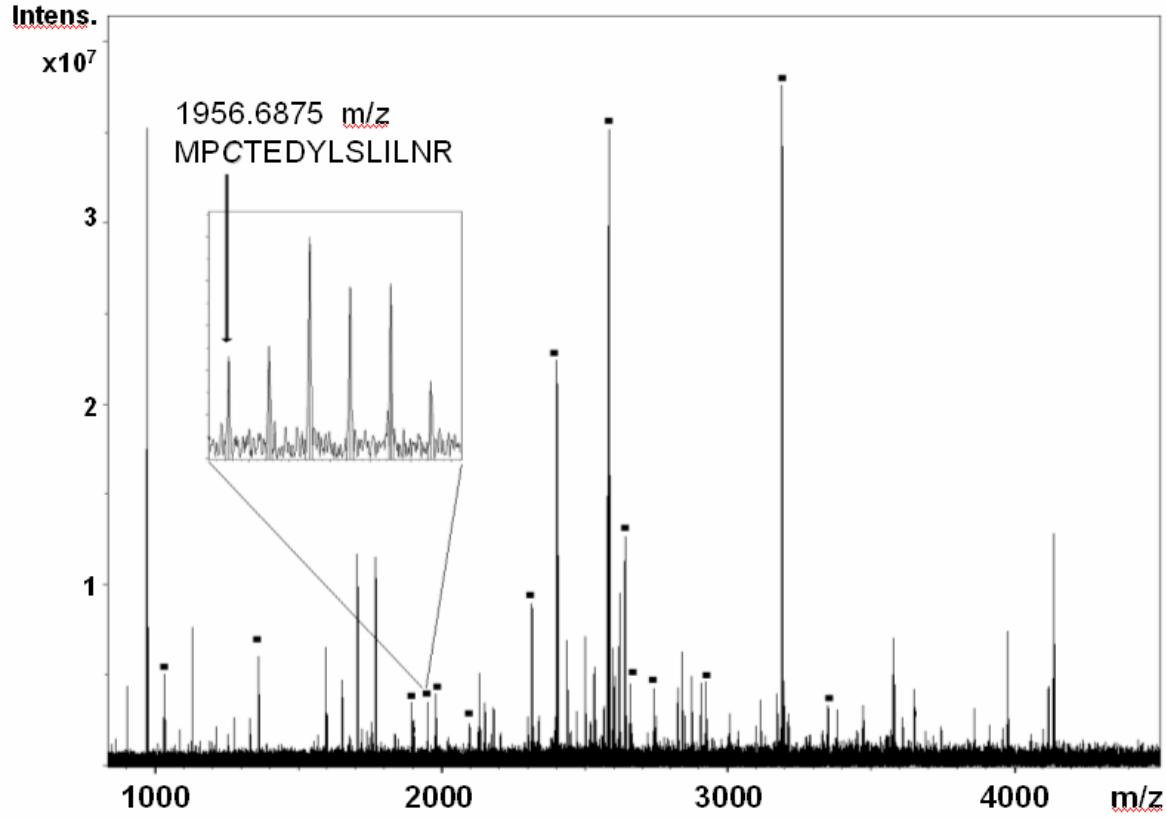

Figure 3.6. MALDI-FTICR mass spectrum obtained of a bovine serum albumin digest; mass defect labeled-peptides are denoted with a box. Inset shows a mass scale expansion of the peaks near m/z 1957, identified as the peptide MPCTEDYLSLILNR, whose predicted isotope pattern is shown in Figure 3.5.

Protein Identification: To test the effectiveness of this method for improving protein identification we derivatized a proteome sample from the organism *M. maripaludis*. For this experiment, we also use endogenous ^{15}N labeling of protein mixtures to improve the specificity of the protein identification. All proteins from two

whole cell lysates are isolated from two identical cultures, one grown using a nitrogen source (ammonium sulfate) with the natural abundance of ^{15}N and the other with 98% ^{15}N. Equal amounts of protein are then collected from each culture, and combined.[25] This method is a useful tool to assist with protein identification. Previously, we have found a significant improvement in the ability to identify peptides by accurate mass measurement when nitrogen stoichiometry is used as a search constraint. (Parks, B.A.; Amster, I.J. *manuscript in preparation*).

M. maripaludis contains 1,722 open reading frames (ORF's), [26] and 18% of the 95,719 predicted tryptic peptides with up to 1 missed cleavage contain cysteine. The utility of this approach (^{15}N and MDL labeling) to protein identification by accurate mass measurement has been estimated for this organism at a mass search tolerance of 10 ppm; the fraction of unique peptides increases from 8 percent (unlabeled peptides) to 43 percent (labeled peptides) when all the possible peptides up to *m/z* 3500 are taken into account. If only the mass defect labeled cysteine-containing peptides are considered, 75% of the masses are unique (database searching with 10 ppm mass tolerance and using the nitrogen stoichiometry as a search constraint).

Increasing the percentage of identified peptides should increase the number of identified proteins. This was examined for the *M. maripaludis* proteome. Whole cell lysates from *M. maripaludis* were derivatized and digested by trypsin and subsequently fractionated by nano-LC using a C18 column. The fractions were analyzed by MALDI-FTMS. Analysis of the spectra resulted in the assignment of 1449 non-redundant peptides masses. Out of these, 156 (11%) were found to be mass defect labeled

peptides. Using these data, a search was made against a list of predicted *M. maripaludis* tryptic peptides masses. Using a mass tolerance of 10 ppm, this resulted in the identification of 304 proteins using both nitrogen stoichiometry and mass defect labeling, which is an improvement of 14% over the 268 proteins identified when the search is made against a list that does not include the MDL-peptides. We have previously analyzed the same proteome (but without mass defect labeling or cysteine alkylation) several times under similar conditions, and we typically identify 275 ± 25 proteins. We attribute the improvement in proteome coverage to the fact that mass defect labeling increases both the detectability and the identification specificity of cysteine-containing peptides. To check the effect of the MDL on the detectability of cysteine-containing peptides, we have made MALDI-FTMS measurements of the tryptic digest products of bovine serum albumin (BSA) prepared using three different methods; (1) with no alkylation of cysteine; (2) with alkylation by iodoacetamide (carbamidomethylation); (3) with alkylation by the mass defect label. Each of the three digests were analyzed four times. Of the 35 cysteine residues in BSA, we observe 4-6 (average equals 5) when the cysteines are not alkylated, 8-15 (average 11.3) when cysteines are alkylated by iodoacetamide, and 12-20 (average 15.5) when the mass defect label is used. These data show that the MDL procedure provides 50% better detectability for cysteine-containing peptides compared to carbamidomethylation, and 300% improvement compared to peptides with unalkylated cysteines.

Detailed analysis of the data gave some insight into the hydrophobicity of the labeled peptides**.** Figure 3.7a shows a graph of the percentage of labeled peptides found per fraction versus the retention time. Most of the labeled peptides eluted from

the column after the gradient reaches 50% organic composition. These data suggest that the labeled cysteine-containing peptides are more hydrophobic, consistent with the structure of the mass-defect label. Earlier elution and better separation of this sample can be achieved by using a column with a less hydrophobic stationary phase. We have examined the same labeled proteome using a C8 column. Analysis of the data resulted in assignment of 1195 pairs of non-redundant peptides masses, of which 126 (11%) were mass defect labeled peptides. Figure 3.7b shows the percentage of labeled peptides per fraction versus retention time with the C8 column. The labeled peptides are found to be distributed more evenly throughout the LC separation with the C8 column. Nevertheless, the total number of identified proteins shows a slight decrease when compared with the data obtained using a C18 column (279 versus 307 identified proteins). Based on the results obtained it appears that earlier elution of mass defect labeled peptides does not seem to positively affect the total number of those peptides observed by MALDI -FTMS.

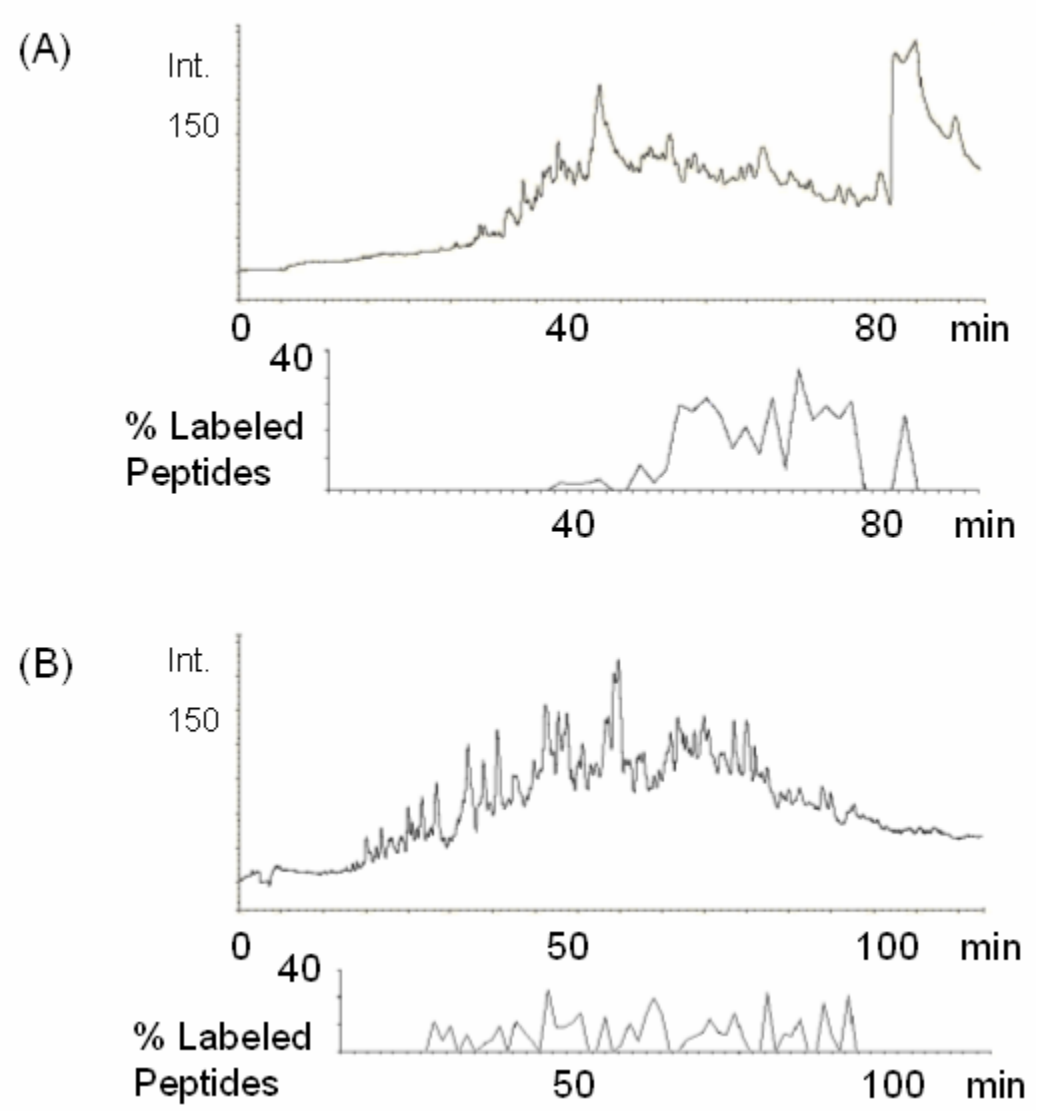

Figure 3.7. Chromatogram and plot of percentage of labeled peptides versus elution time for: (A) C18 column proteome separation and (B) C8 column proteome separation. Percent of labeled peptides was calculated using the total number of peptides observed and the number of MDL peptides found for each fraction collected, and analyzed by MALDI-FTICR mass spectrometry.

Combining both sets of data, the total number of observed peptide pairs ($^{14}N/^{15}N$) is 6146; 475 of these were found to be labeled with the cysteine-specific reagent. It is useful to use this large data set to examine the improvement in database searching that results from mass defect labeling and metabolic ^{15}N labeling. For peptides without a mass defect label, the fraction of unique peptides goes from 7 percent when using only the molecular weight to search the database (i.e. no nitrogen stoichiometry data used in search) to 27 percent for the non-MDL peptides when the nitrogen stoichiometry constraint is used. For the mass defect labeled proteome, the number of unique peptides increases to 2108, which represents 34 percent of the total number of peptides. If one considers only the peptides labeled by 2,4-dibromoacetanilide, 47% of the peptides are identified. Having a higher percentage of unique peptides increases the number of identified proteins. Indeed, identification of proteins shows that if only the non-labeled peptides are used, 377 proteins are identified compared to 425 proteins identified when all the found peptides masses are used. These “extra” 48 proteins are not usually identified from the complex mixture of proteins from *M. maripaludis* by the standard protocol (no cysteine alkylation), demonstrating that better protein coverage is obtained by using the accurately measured masses of mass defect labeled cysteine-containing peptides to identify proteins.

We anticipate significant improvement in this method by refinement of this technique. For example, we note that the percentage of identified peptides obtained in these experiments is lower than one would predict from a statistical analysis of the proteome. The expected identification specificity mentioned above (43% identification for non-MDL peptides, searching at 10 ppm mass tolerance and using the nitrogen

stoichiometry as a constraint; 75% identification for MDL-peptides) was calculated using all the possible tryptic peptides in the mass range of 700-3500 amu. Figure 3.8 shows a plot of the number of peptides observed versus their mass-to-charge for the experiment using a C18 analytical column. Most peptides are found in the range between 700 and 2500 amu. The calculated fraction of unique peptides for the tryptic peptides within this mass range is 36% which corresponds well with the observed experimental result of 34%. Detection of higher mass peptides can be achieved by optimizing the operational conditions of the instrument MALDI-FTMS. For the instrument used in these studies, by optimizing the higher mass region of the mass range, the sensitivity of the lower mass region is reduced. Recently, it has been demonstrated in our laboratory that by combining data collected using two different sets of tuning conditions the dynamic range for the analysis of a proteome can be improved.[22] Another approach to increasing the number of mass defect labeled peptides observed is analyzing them by ESI-MS. It has been found in previous studies that ESI is more favorable for the ionization and detection of hydrophobic peptides than is MALDI.[27, 28] Therefore, more mass defect labeled cysteine-containing peptides are expected to be observed by using ESI compared to MALDI. This will be the subject of future studies in our laboratory. Combining both MALDI and ESI results could lead to gaining the most information possible out of a particular sample due to their complementary nature.[29]

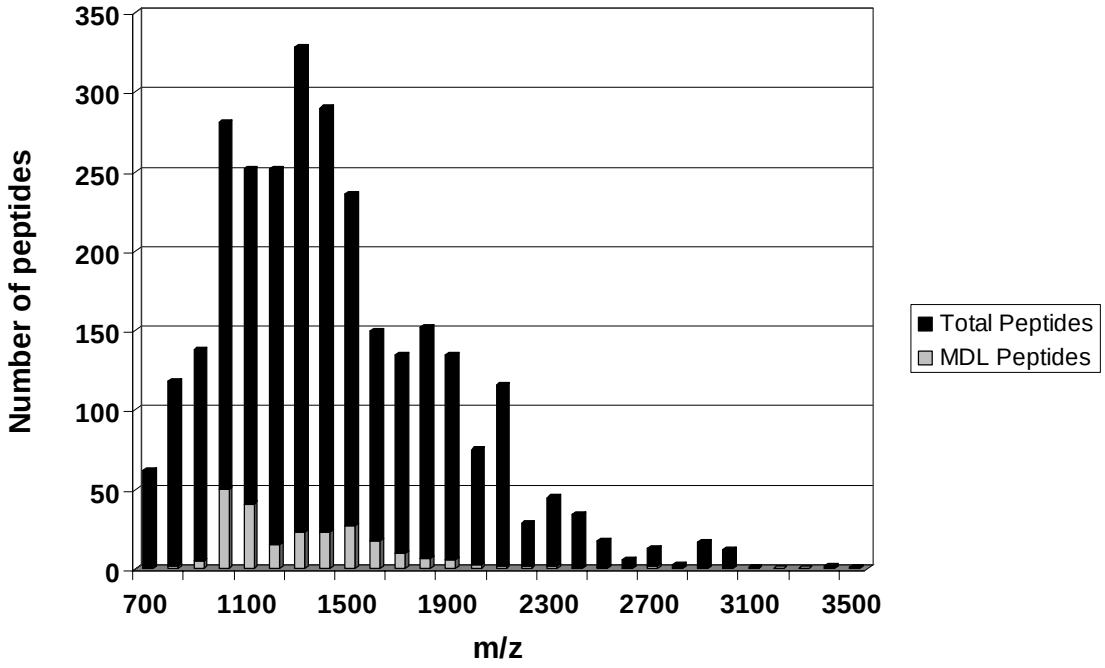

Figure 3.8. Histogram for all possible tryptic peptides from *M. maripaludis* within 700 and 3500 amu. Gray bars represent the number of MDL peptides and black bars the total number of peptides for each 100 amu mass bin.

Conclusions: The method presented here provides a way to improve the specificity of peptide identification based on accurate mass measurement, which leads to an increase in the number of proteins that can be identified in an organism with small genome (< 5000 ORF's). This approach has several significant differences from methods that use derivatives with affinity tags, such as ICAT reagents.[30] First, both unlabeled and mass defect labeled peptides are analyzed simultaneously, which eliminates the need for separation prior to analysis and allows the detection of proteins that do not contain cysteine. Second, improvement in specificity arises from the decongestion of the mass spectrum, meaning that regions of the mass space that were

previously unoccupied will be populated by the labeled cysteine-containing peptides. Another important advantage of using this approach constitutes the identification of proteins usually missed by other methods; in this case 48 extra proteins were identified by adding a mass defect tag to the cysteine-containing peptides, as this is found to improve both their detectability and their identification specificity. In addition, the analysis of these samples was performed by MALDI-FTMS without requiring the use of tandem MS which demands the acquisition of much larger data sets, and requires significantly more computational analysis of the data. This approach can be extended to the labeling of other amino acids that occur with lower than average frequency, such as tryptophan or histidine, by using labeling reactions that are specific for these amino acids. Such work is currently under investigation in our laboratory.[31]

References

(1) Gygi, S. P.; Rist, B.; Aebersold, R. Measuring gene expression by quantitative proteome analysis*. Curr. Opin. Biotechnol.* **2000**, *11*, 396-401.

(2) Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics*. Chemical Reviews* **2001**, *101*, 269-295.

(3) Peng, J.; Gygi, S. P. Proteomics: the move to mixtures*. Journal of Mass Spectrometry* **2001**, *36*, 1083-1091.

(4) Beranova-Giorgianni, S. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations*. Trends Anal. Chem.* **2003**, *22*, 273-281.

(5) Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Pasa-Tolic, L.; Smith, R. D. Utility of accurate mass tags for proteome-wide protein identification*. Anal. Chem.* **2000**, *72*, 3349-3354.

(6) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III Direct analysis of protein complexes using mass spectrometry*. Nature America* **1999**, *17*, 676-682.

(7) Nepomuceno, A. I.; Muddiman, D. C.; Bergen, R.; Craighead, J. R.; Burke, M. J.; Caskey, P. E.; Allan, J. A. Dual Electrospray Ionization Source for Confident Generation of Accurate Mass Tags Using Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*. Anal. Chem.* **2003**, *75*, 3411-3418.

(8) Wolters, D. A.; Washburn, M. P.; Yates, J. R., III An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683-5690.

(9) McDonald, H.; Yates, J. R., III Shotgun proteomics and biomarker discovery. *Disease Markers* **2002**, *18*, 99-105.

(10) Eng, J. K.; McCormack, A. L.; Yates, J. R., III An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976-989.

(11) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2002**, *2*, 513-523.

(12) Lipton, M. S.; Pasa-Tolic, L.; Anderson, G. A.; Anderson, D. J.; Auberry, D. L.; Battista, J. R.; Daly, M. J.; Smith, R. D. Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11049-11054.

(13) Beynon, J. H. *Advances in Mass Spectrometry*; Pergamnon Press: New York, (1963).

(14) Kendrick, E. A mass scale based on CH2 = 14.0000 for high resolution mass spectrometry of organic compounds. *Anal. Chem.* **1963**, *35*, 2146-2154.

(15) Wu, Z. G.; Rodgers, R. P.; Marshall, A. G. Two- and three-dimensional van Krevelen diagrams: A graphical analysis complementary to the Kendrick mass plot for sorting elemental compositions of complex organic mixtures based on

ultrahigh-resolution broadband Fourier transform ion cyclotron resonance mass measurements. *Anal. Chem.* **2004**, *76*, 2511-2516.

(16) Marshall, A. G.; Rodgers, R. P. Petroleomics: The next grand challenge for chemical analysis. *Acc. Chem. Res.* **2004**, *37*, 53-59.

(17) Stenson, A. C.; Marshall, A. G.; Cooper, W. T. Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **2003**, *75*, 1275-1284.

(18) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. N. Kendrick mass defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **2001**, *73*, 4676-4681.

(19) Jones, J. J.; Stump, M. J.; Fleming, R. C.; Jackson O. Lay, J.; Wilkins, C. L. Strategies and data analysis techniques for lipid and phospholipid chemistry elucidation by intact cell MALDI-FTMS. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1665-1674.

(20) Hall, M. P.; Schneider, L. V. Isotope-differentiated binding energy shift tags (IDBEST) for improved targeted biomarker discovery and validation. *Expert Review of Proteomics* **2004**, *1*, 421-431.

(21) Schneider, L. V.; Hall, M. P.; Petesch, R.; Target Discovery: U.S., 2002.

(22) Wong, R. L.; Amster, I. J. Combining low and high mass ion accumulation for enhancing shotgun proteome analysis. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 205-212.

(23) Senko, M. W.; Beu, S. C.; McLafferty, F. W. Automated Assignment of Charge States from Resolved Isotopic Peaks for Multiply Charged Ions. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 52-56.

(24) Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal. Chem.* **2000**, *72*, 1112-1118.

(25) Oda, Y.; Cross, F. R.; Cowbum, D.; Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6591-6596.

(26) 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. Complete Genome Sequence of the Genetically Tractable Hydrogenotrophic Methanogen Methanococcus maripaludis. *Journal of Bacteriology* **2004**, *186*, 6956-6969.

(27) Medzihradszky, K. F.; Leffler, H.; Baldwin, M. A.; Burlingame, A. L. Protein identification by in-gel digestion, high-performance liquid chromatography, and mass spectrometry: peptide analysis by complementary ionization techniques. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 215-221.

(28) Bodnar, W. M.; Blackburn, R. K.; Krise, J. M.; Moseley, M. A. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage*. J. Am. Soc. Mass Spectrom.* **2003**, *14*, 971-979.

(29) Stapels, M. D.; Barofsky, D. F. Complementary use of MALDI and ESI for the HPLC MS/MS analysis of DNA binding proteins*. Anal. Chem.* **2004**, *76*, 5423-5430.

(30) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. Proteome analysis of low abundance proteins using multidimensional chromatography and isotope-coded affinity tags*. Journal of Proteome Research* **2002**, *1*, 47-54.

(31) Li, C.; Gawandi, V.; Protos, A.; Phillips, R. S.; Amster, I. J. A MALDI compatible reagent for tagging tryptophan residue*. European Journal of Mass Spectrometry* **In press**.

CHAPTER 4

TRYPSIN DIGESTION IN MIXED AQUEOUS-ORGANIC SOLVENT SYSTEM: AN APPROACH TO IMPROVE SPECIFICITY OF THE PROTEOLYSIS FOR PROTEIN ANALYSIS

Introduction: Proteomic analysis has relied heavily upon separation of proteins by gel electrophoresis for many years until recently, when new developments have decreased the need for separation of protein mixtures prior to protein identification.[1-5] The approach known as “shotgun proteomics” has gained significant popularity in the proteomic field.[6, 7] Shotgun proteomics refers to digestion of the unseparated protein mixture followed by chromatographic separation of the resulting peptides and subsequent analysis by mass spectrometry.[3, 8] Therefore, identification of the proteins present in the sample is based on the identification of their corresponding proteolytic fragments or peptides.[8, 9] These proteolytic fragments are commonly generated by digestion by the enzyme trypsin, which generates peptides with masses suitable for analysis by most of the routinely used mass spectrometers.[10, 11] Trypsin cleaves specifically at the C-terminus of arginine and lysine residues unless they precede a proline residue, generating a majority of fragments in the molecular weight range under 4000 Da.[10, 12, 13] Trypsin is believed to have very high specificity for the sites of cleavages but, practical experience also shows that for any analysis of a tryptic digest, there is a number of peptides masses that cannot be correlated to predicted tryptic fragments.[13, 14] Since trypsin sometimes cannot access every expected cleavage site, search engines also provide information for peptides exhibiting missed cleavages.[15-17] However, in most cases, many of the unmatched masses remained unidentified and these fragments are often attributed to peptides resulting from nonspecific cleavages or peptides containing residues carrying post-translational or chemical modifications.[18] Usually, if the peptide signal intensity is low enough, the mass is simply not taken into

consideration for the analysis, however, sometimes they can make up for a significant percentage of the total masses obtained for an specific study.[19]

Several different approaches have been reported in order to enhance the efficiency of cleavage of proteins, including use of immobilized trypsin reactors,[20-22] addition of organic solvents to aid protein denaturation,[22-24] chemical cleavage at residues other than lysine and arginine,[25] high-temperature proteolytic digestion,[26] enhancing trypsin digestion by microwave energy,[27] among many other methods.[28, 29] This chapter will describe an approach to tryptic digestions that makes use of the organic solvent acetonitrile to assist denaturation of proteins and, it also reduces greatly the time devoted to trypsin reaction with proteins. Stability of the enzyme trypsin in presence of organic solvents has been studied and, it has been found that it retains its activity under conditions that usually denature other proteins.[30-32] We took advantage of this fact and tested different conditions for tryptic digestion on protein standards in order to determine if an improvement on peptide assignment could be achieved. We observed an increment in the number of assignable masses for protein standards and also a decrease in the number of unmatched masses.

Experimental Section:

Protein standard solutions: Protein standards β-Lactoglobulin and apo-transferrin were obtained from Sigma-Aldrich (St. Louis, MO), bovine serum albumin from Calbiochem (San Diego, CA). Each protein standard mixture contained 100 μg of protein (unless otherwise stated) and was dissolved in the following solvent systems: A) 10 mM ammonium bicarbonate (Sigma-Aldrich, St. Louis, MO) and, B) 50 mM tris-HCl (Fisher Scientific, Fair Lawn, NJ) / 10 mM $CaCl_2$ (J.T. Baker, Phillipsburg, NJ) / variable concentration of acetonitrile (Fisher Scientific, Fair Lawn, NJ). Denaturation was achieved by heating at 95 ^{o}C. Disulfide bonds were reduced by addition of tris (2-carboxyethyl) phosphine, TCEP (Pierce Biotechnology, Rockford, IL). Prior to trypsin digestion, the denatured protein was subjected to centrifugal size exclusion chromatography using a 3 mL spin column packed with Sephadex G-25 (Sigma-Aldrich, St. Louis, MO) to remove any small molecule present in the mixture.

Rubredoxin from *Pyrococcus furiosus* (PF1282): This protein was purified from *P. furiosus* as described by Jenney and Adams.[33]

Protein trypsin digestion: Trypsin digestions were performed both under standard conditions (Promega, Madison WI), i.e. at 37 ^{o}C, pH 7, for 12 hours and, over either a 60 or 30 minutes period in presence of a variable amount of acetonitrile. The same amount of trypsin was added to each sample (300 ng) and, all of the digestions

were performed in duplicate. Trypsin activity was inhibited after digestion by addition of 5 μL 10% formic acid (J.T. Baker, Phillipsburg, NJ).

SDS-PAGE preparation: The gel solution, NEXT GEL™ PAGE 12.5%, and running buffer, NEXT GEL™ Running Buffer, used for 1D - SDS-PAGE were obtained from AMRESCO Inc (Solon, OH). Stains were obtained from Sigma-Aldrich (St. Louis, MO). Broad range pre-stained protein marker (NEB, Frankfurt, Germany) was used as a molecular weight standard. Analysis of protein samples on 1D SDS-PAGE were done following standard procedures.[34]

Mass Spectrometry: Samples were analyzed by matrix assisted laser desorption/ionization (MALDI) Fourier transform ion cyclotron resonance (FTICR) mass spectrometry using a 9.4 Tesla magnet (Bruker Daltonics Inc, Billerica, MA). This instrument is equipped with a SCOUT 100 MALDI source which desorbs ions at elevated pressure (~ 1 mTorr) to suppress metastable decomposition. Conditions for operation of the FTICR MS were similar to those reported previously,[35] and external mass calibration was established using a peptide mixture generated by tryptic digestion of chicken egg albumin (Sigma, St. Louis, MO). The MALDI matrix was 2,5-dihydroxybenzoic acid (DHB) (Lancaster, Pelham, NH). Attempts to MS/MS experiments were performed by an Apex-Qe-FTMS from Bruker Daltonics Inc. (Billerica, MA). This instrument is equipped with an Apollo II electrospray ionization source and a quadrupole mass analyzer collision cell which allows precursor selection and collisional activated dissociation of ions.

Peptide assignment: For assignment of the peptide masses from the mass spectra, monoisotopic peptide mass lists were prepared from DataAnalysis Version 3,.4 (Bruker Daltonics Inc, Billerica, MA) and compare against a list of the predicted tryptic peptides obtained by MS-Digest (http://prospector.ucsf.edu). Also, the web-based program, MS-nonspecific (http://prospector.ucsf.edu) was used to investigate the origin of unknown fragments.

LC separation and MS/MS analysis: A spray column (75 μm, 10cm; PicoTip EMITTER, New Objective, Woburn, MA) was prepared by packing silica C18 resins (Rainin Microsorb MV, 5 μm, 300 E pore size) with 50% isopropanol and 50% methanol. Prior to reversed-phase HPLC, the trypsinized sample was loaded onto the column using a pressurized stainless steel bomb and nitrogen gas at 1,000 psi for 30 min. The elution of peptides was initiated at a flow rate of approximately 400 nL/min with a 65-min linear gradient of 5 to 60% ACN in 0.1% formic acid/water after a 10 min rinse in 0.1% formic acid. The spectra were acquired by nano elecrospray ionization on a Finnigan LTQ Linear Ion Trap Mass Spectrometer (Thermo Finnigan, San Jose, CA) which was directly coupled to the reversed-phase HPLC system. The spray voltage was 1.92 kV, and the capillary temperature was 2200C. The instrument was operated in a data-dependent mode, i.e. the 9 most abundant ions detected in MS mode were independently selected for MS/MS analysis. MS/MS spectra obtained in this manner were searched against the theoretical MS/MS spectra of all peptides in the database to provide the best match and thus the most probable peptide sequence for each precursor ion.

Results and Discussion:

Digestion of protein standards: The objective of the experiments described in this chapter is to optimize our digestion protocol for specific protelysis by systematically testing different amount of the organic solvent acetonitrile and various digestion times. A second objective is to characterize the source of unpredicted peptides that appear in a batch digest. As mentioned before, proteomic analysis relies heavily upon the identification of peptides and subsequent determination of the parent protein.[36, 37] However, it is also known that not all of the peptides detected can be assigned to predicted fragments and, in addition, not all of the proteins present in a sample will generate peptides in the range of detectability of the instrument and some do not generate any peptides.[19, 36, 38, 39] There are many possibilities for the origin of unexpected peptides. These include non-specific proteolysis, contamination with proteins not present in the database being searched, post-translational modification, disulfide linkages, errors in the sequence database, mis-translation, etc. Some proteins are resistant to proteolysis under conditions known to effectively digest other proteins, therefore, digestion of complex protein mixtures does not usually contain peptides for every single protein present in the original undigested solution.[40-42] All these factors, in addition to the large dynamic range of concentrations of proteins in biological systems, lead to observation of only about 25% of all the proteins expected for any given proteome analysis.[38] Our purpose is to improve the effectiveness of the proteolysis and decrease the amount of non-specific cleavages by modifying the conditions for tryptic digestion.

First, we performed different digestions of the protein bovine serum albumin under the conditions listed in table 4.1. Our lab routinely uses ammonium bicarbonate (ABC) as the buffer of choice for dissolving both, proteins and, the enzyme trypsin. In this experiment, we used this buffer for some of the digestions and, we also used the buffer tris (hydroxymethyl) aminomethane hydrochloride (Tris-HCl) in presence of calcium chloride ($CaCl_2$) for the remaining samples. The organic solvent acetonitrile (ACN) was added to the samples containing tris-HCl, in order to aid with denaturation of the protein and speed the time required to obtain tryptic fragments.[23]

Solvent system and digestion time	**Protein Coverage (%)**	**Matched Masses (%) (# of matched masses / # of masses submitted)**
ABC . Overnight	46	23 (24/103)
20% ABC / 80% ACN . 1 h	42	32 (25/77)
20% Tris-HCl / 80% ACN . 1 h	53	44 (32/72)

Table 4.1 Peptide identifications for bovine serum albumin for different conditions of tryptic digestion, using MALDI-FTMS of the unseparated digest

Results showed that the digestion performed following the “standard” methodology (ABC, overnight digest) gave the poorest results regarding the amount of peptides masses that could not be assigned as predicted tryptic fragments (Table 4.1).

The highest protein coverage and percentage of matched masses was achieved for the sample containing the mixture tris-HCl / acetonitrile. The high percentage of non-assignable tryptic fragments found in the overnight digest (77%) could be rationalized, in part, by the fact that commercially available trypsin might contain small amounts of active chymotrypsin and that longer digestion times could allowed some of this enzyme activity to take place generating fragments that are not strictly tryptic cleavages.[24, 31] Based on the results obtained, we proceeded to perform different digestions in aqueous-organic solvent systems, containing variable concentrations of acetonitrile and, using tris-HCl as the aqueous solvent. Digestions of bovine serum albumin were performed using 80%, 65% and 50% of acetonitrile and results are shown in table 4.2

Solvent system and digestion time	**Protein Coverage (%)**	**Matched Masses (%)**	**Peptides with one missed cleavage (%)**	*** Peptides with more than one missed cleavages (%)**
Tris-HCl / 80% ACN . 1 h	42	40	39	5
Tris-HCl / 65% ACN . 30 min	58	42	38	4
Tris-HCl / 65% ACN . 1 h	55	43	32	4
Tris-HCl / 50% ACN . 30 min	69	49	37	6
Tris-HCl / 50% ACN . 1 h	54	52	38	4

Table 4.2 Peptide identifications for bovine serum albumin for different concentrations of acetonitrile in a tryptic digestion. MALDI-FTMS of unseparated digest.

* All these peptides exhibit missed cleavage other than RR, KK, RK or KR

We observed an even higher improvement of both protein coverage and, percentage of predicted masses observed, when decreasing the amount of acetonitrile present in the solvent system down to 50%. At the same time, we tested two different digestion times, 1 hour, which was previously found to be very effective, and 30 minutes. For samples containing 50% acetonitrile, the amount of non-predicted fragments seems to be the same for both 30 minutes and 1 hour of digestion, protein coverage seems to be higher when trypsin is allowed to cleave the protein for only 30 minutes. After determining the optimum concentration of acetonitrile for tryptic digestion of bovine serum albumin, different protein standards were tested by varying the digestion time and maintaining the concentration of acetonitrile at 50%. The protein standards tested were β-lactoglobulin (MW ~ 19 KDa), apo-transferrin (MW ~ 77 KDa) and bovine serum albumin (MW ~ 66 KDa). We would expect the smaller protein to digest faster than the heavier ones and, that is indeed what we observe for the data presented in Table 4.3. Times shorter than 30 minutes do not seem to work well for bovine serum albumin, even though it appears sufficient for the heavier protein apo-transferrin and the lighter β-lactoglobulin. Overall, results suggest that 30 minutes of trypsin digestion in presence of tris-HCl / acetonitrile (50/50) are optimum for achieving good protein coverage and the least amount of unassignable fragments.

Figure 4.1 shows a MALDI-FTICR mass spectrum for a bovine serum albumin digest performed under these conditions, the expected tryptic fragments are identified by a circle on their m/z peaks. The majority of the intense peaks are identified as fragments from bovine serum albumin. However, a significant number of unassignable

masses of low intensity are still observed (Figure 4.2). In order to determine the origin of these fragments, we proceeded to search the various peptide masses lists with the tool MS-nonspecific (http://prospector.ucsf.edu), which is provided by the UCSF mass spectrometry facility for finding peptides with non-specific cleavages. A non-specific digest of a small protein containing about 150 residues (i.e. β-lactoglobulin) could generate up to 3000 possible non-specific peptides in the mass range from 700 to 4000 Da. Therefore, a number of hits would be expected for any mass list submitted to this search engine. Since our analyses are performed by FTCIR technology which provides high mass accuracy, we could determine, based on the mass error, if a non-specific mass hit could indeed be considered as a matched mass. For the analysis of all protein standards the highest mass error found for matched tryptic peptides was 2.9 ppm which corresponded to a high m/z fragment, > 3000 Da (mass accuracy input for MS-Fit was 10 ppm). When searching non-specific cleavage peptides, MS-nonspecific returned several hits, as expected, but the majority of them showed high mass error (> 6 ppm) compared with the results obtained for the matched tryptic fragments. Therefore, we determined that these fragments more likely are false identification and, trypsin might indeed be cleaving specifically after lysine and arginine as expected.[13]

The presence of a number of unmatched masses could be explained as resulting from protein impurities. This was tested by analyzing the commercially available bovine serum albumin used for these experiments. Figure 4.3 shows an image of a 1D - sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of a sample of the intact bovine serum albumin used for all the experiments. Other components are

observed in this sample and they exhibit high intensity bands suggesting that the contaminant concentration is above trace level unlike claimed by Calbiochem®.[43] Some of the extra components are of higher mass, probable BSA oligomers. However, the most abundant extra peak occurs at lower mass, and is probably a fragment of the BSA. The unassignable fragments most likely are originated from proteolysis of these unknown-sequence components.

β-Lactoglobulin		
Digest duration	Matched Masses (%) (# of matched masses / # of masses submitted)	Coverage (%)
15 min	9 (11/116)	76
30 min	7 (7/93)	63
1 hour	6 (7/104)	63
Bovine serum albumin		
15 min	35 (26/75)	39
30 min *	29 (31/105)	53
1 hour	26 (28/107)	52
Apotransferrin		
15 min	42 (33/78)	51
30 min	37 (30/82)	54
1 hour	34 (31/89)	54

Table 4.3 Peptide identifications for different protein standards for different times of tryptic digestion containing 50% of acetonitrile

* Results differ from the ones obtained previously for the same condition digest

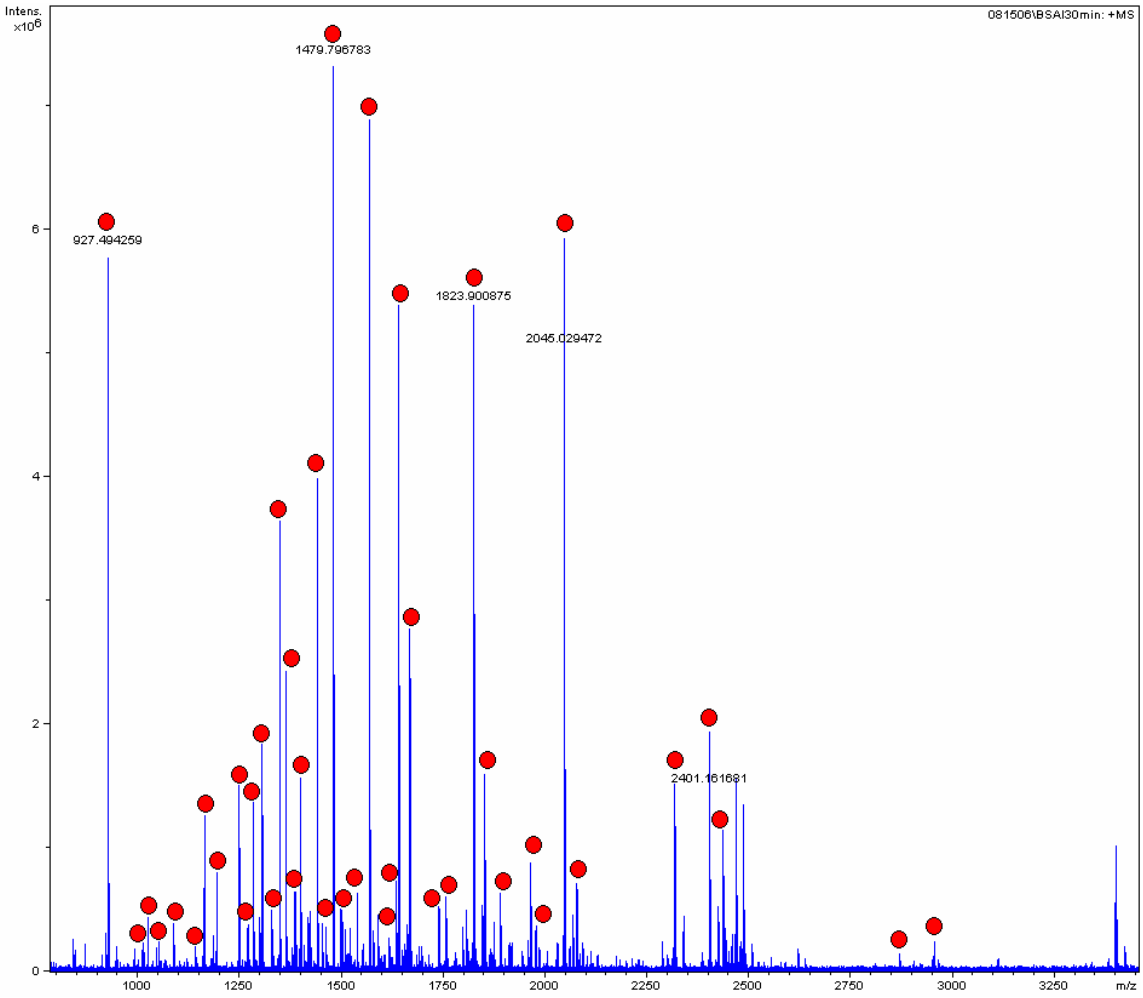

Figure 4.1 . MALDI-FTICR mass spectrum obtained of a bovine serum albumin digest; tryptic fragments are denoted with a circle.

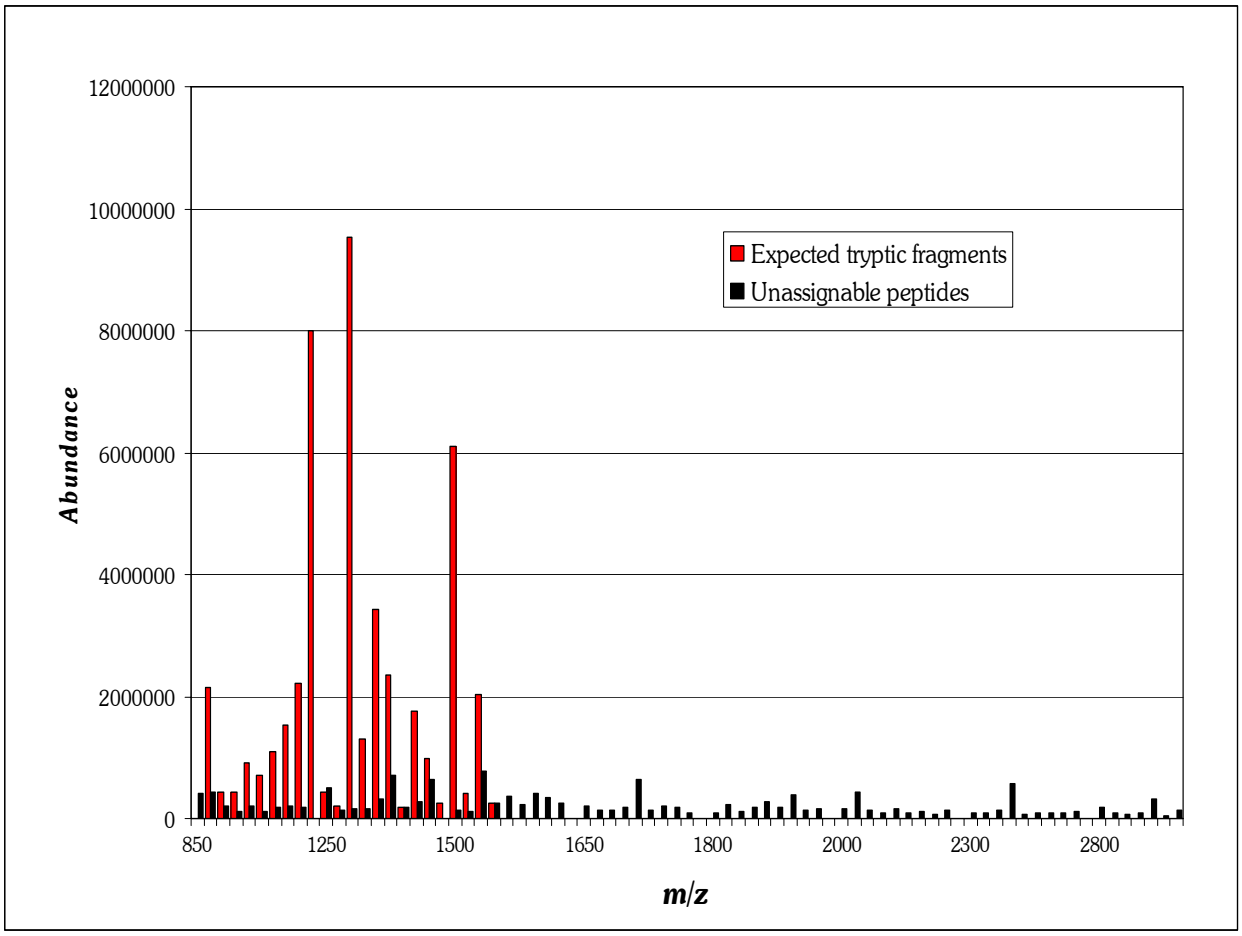

Figure 4.2 Histogram for peptides obtained from the bovine serum albumin digest shown above against abundance of the peptides. Red bars represent expected tryptic fragments and black bars the unassignable peptides.

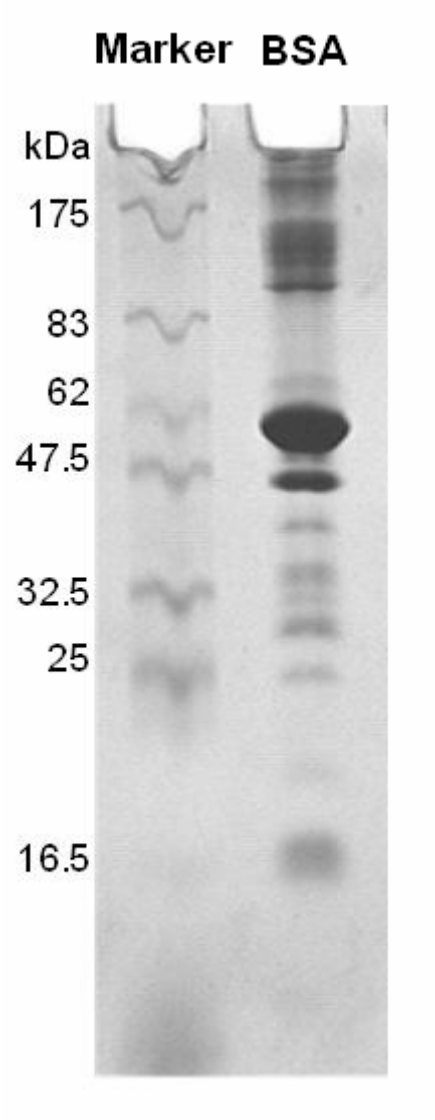

Figure 4.3 . 1-D SDS-PAGE Coomassie blue stained gel for bovine serum albumin from Calbiochem®.

Nevertheless, still question remained as to how effectively trypsin can actually react with the totality of the protein present in the sample. In order to investigate this issue, different digestions of bovine serum albumin were performed. 300 μg of protein were digested for each sample; conditions and results regarding protein coverage and percentage of matched masses obtained for these digestions are listed in table 4.4.

After analysis by MALDI-FTICR of these peptide mixture solutions, SDS-PAGE was performed in order to visualize how much of the protein was left over undigested, if any.

Solvent system and digestion time	**Protein Coverage (%)**	**Matched Masses (%)**
ABC . Overnight digest	29	48
Tris-HCl / ACN (50/50) . 60 m	37	54
Tris-HCl / ACN (50/50). 30 min	37	63
Tris-HCl, No $CaCl_2$ / ACN (50/50). 60 min	51	51

Table 4.4 Peptide identifications for bovine serum albumin for different digestion conditions

We purposely included a set of samples that did not contain $CaCl_2$ in order to determine if calcium reduces the rate of trypsin autolysis as reported.[44-46] Also, if autolysis of trypsin is actually diminished by the presence of $CaCl_2$, by performing gel electrophoresis on the resulting peptide mixture solution we could determine if their is some remaining undigested or partially digested protein. Visualization of the components of the sample on the 1-D polyacrylamide gel was performed by utilizing both coomassie blue (Figure 4.4) and, silver staining (Figure 4.4). Coomassie blue tends to be less sensitive than silver, meaning that usually more components of the sample are observed in silver stained gels. Figure 4.3 shows the coomasie blue stained gel containing the different digestion solutions; we also included a solution containing

300 μg of intact protein, in order to compare the intensity of this spot with that of any sample showing some undigested protein. There are not spot visible for any of the digested solutions suggesting that digestion has been complete, and the spot corresponding to the intact bovine serum albumin can be readily visualized. We proceeded to use silver staining to further investigate if indeed all of the protein was digested for all the cases.

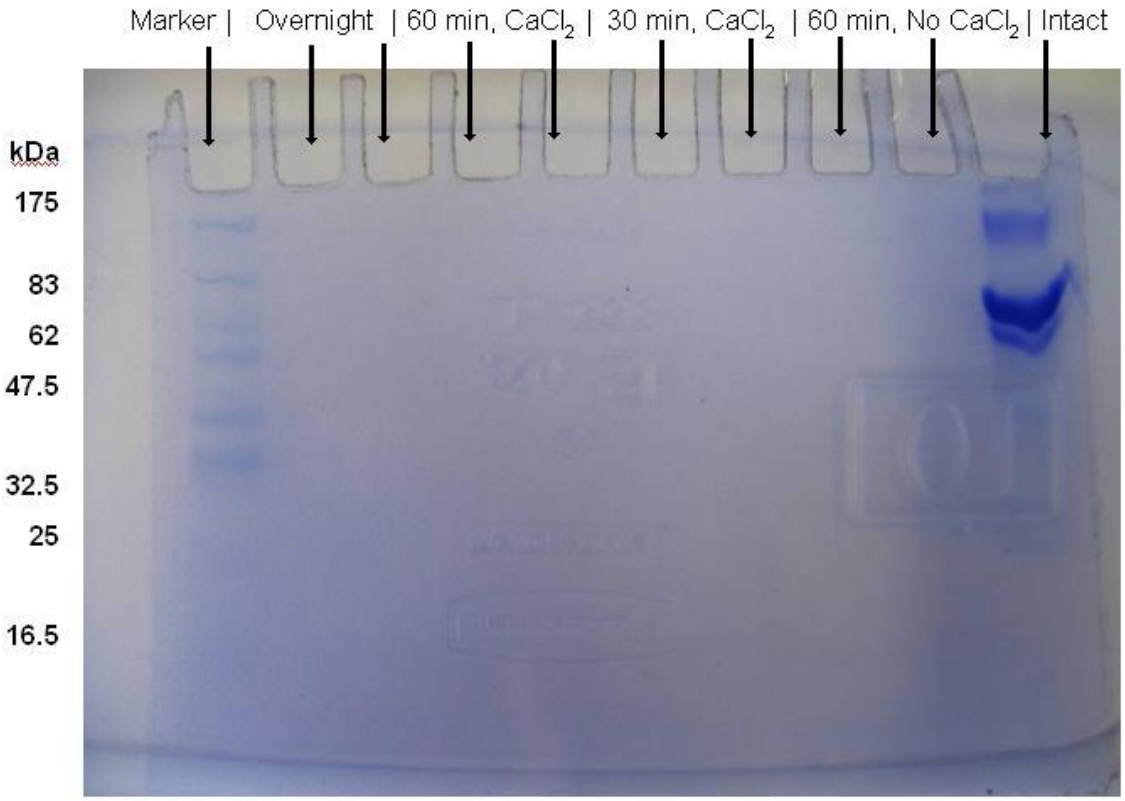

Figure 4.4 . 1-D SDS-PAGE Coomassie blue stained gel for different peptide mixture solutions

Figure 4.5 shows the silver stained gel which provides more information about the different samples. We still see that the majority of the protein was digested for each case and we can only observe spots around 27 KDa which should correspond to the enzyme trypsin. A more detailed look at this gel shows that for the samples digested overnight, the spot corresponding to trypsin seems a little less intense than that of the other samples. This suggests that trypsin undergoes autolysis when digestion is allowed to go overnight, producing extra peaks in the mass spectrum. For the samples allowed to react for only 1 hour with no $CaCl_2$, the band corresponding to trypsin also seems a little less intense compared to those of the samples with $CaCl_2$, suggesting that the short time of digestion does not allow trypsin to react with itself significantly, and that $CaCl_2$ suppresses autolysis, in order to further investigate the activity of trypsin in presence of $CaCl_2$, we performed an experiment in which we allowed trypsin to undergo autolysis for 60 minutes in two different solvent systems, A) ammonium bicarbonate / acetonitrile (50/50) and B) tris-HCl, $CaCl_2$ / acetonitrile (50/50). Figure 4.6 shows the spectra obtained by MALDI-FTICR for both samples and, it can be observed the lack of major autolysis peaks in the system containing $CaCl_2$ against the ABC/ACN mixture. The peaks at m/z 842.5094, 1045.5639 and 2211.1012 are well known autolysis peaks for trypsin, but we also observed some others masses that are not reported in the literature.[47] These results suggest then, the possible advantageous consequences of adding calcium chloride to protein digestion solutions.

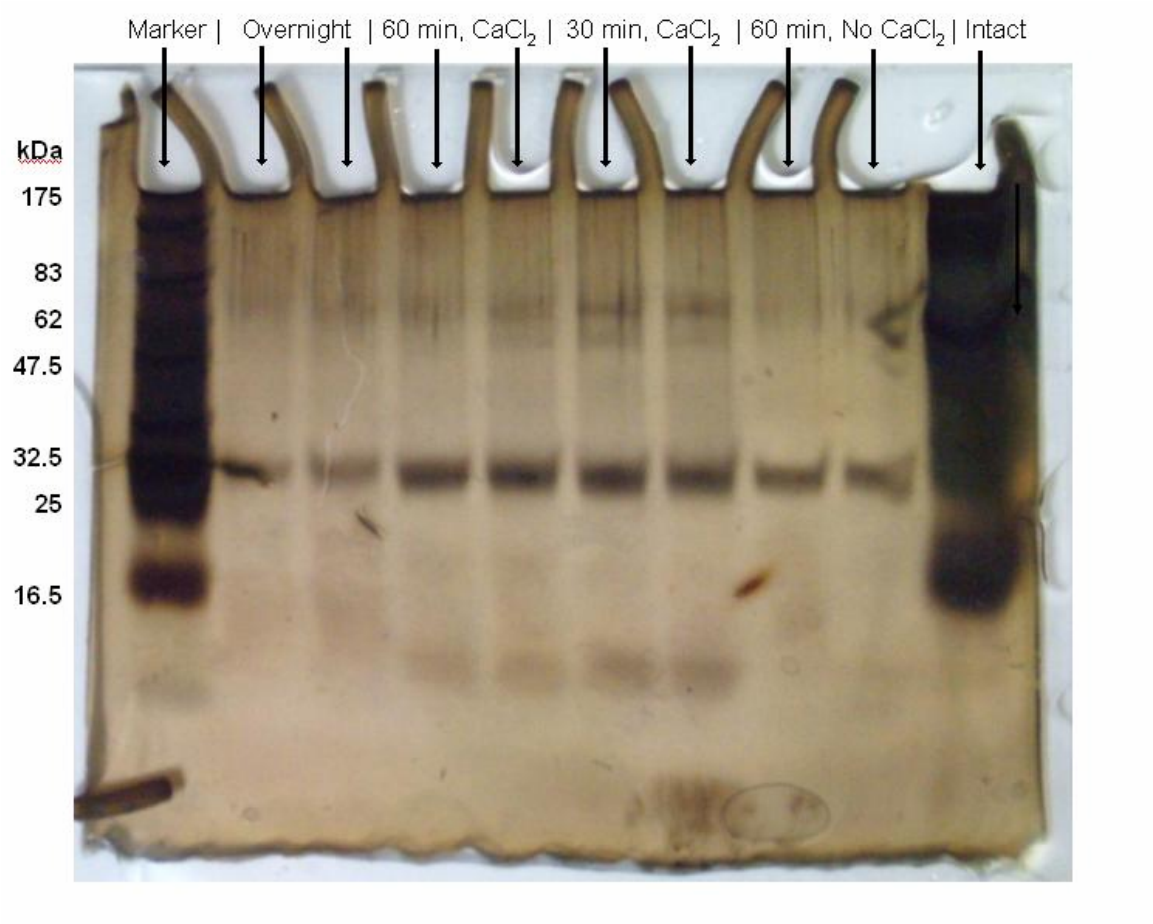

Figure 4.5 . 1-D SDS-PAGE Silver stained gel for different peptide mixture solutions.

The protein band at ~ 30 KDa corresponds to trypsin

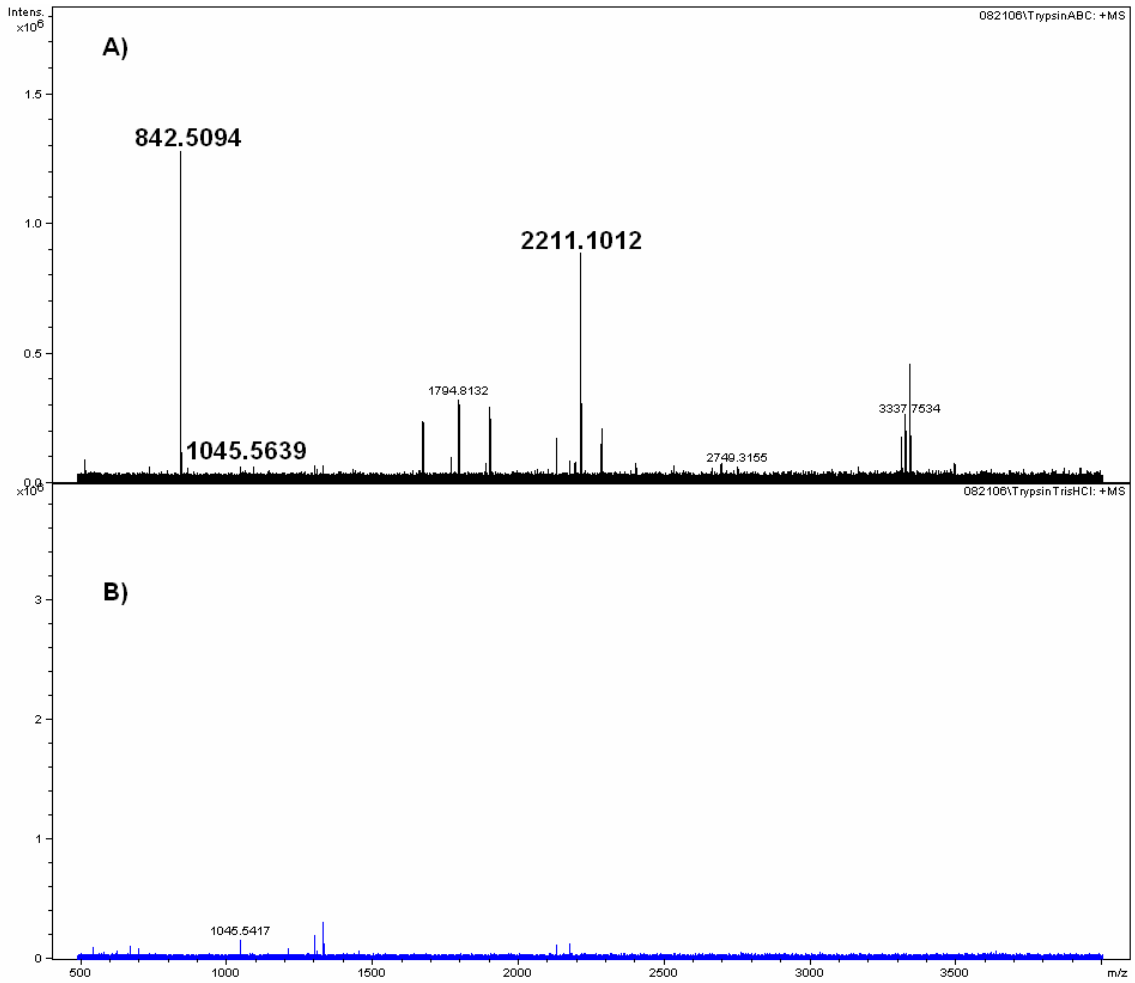

Figure 4.6 . MALDI-FTICR mass spectra obtained of trypsin autolysis in A) ABC / ACN (50/50) and B) Tris-HCl, $CaCl_2$ / ACN (50/50)

Digestion of Rubredoxin protein (PF1282) from *Pyrococcus furiosus:* As stated before, the protein standards used are not completely free from contaminants which generate unassignable fragments that cannot be explained in the analysis; therefore, we tested two different solvent systems and conditions for the digestion of the protein PF1282 (MW: 5896 Da) purified from E. coli which is believed to have

significantly less amount of contaminants than the protein standards used. Figure 4.7 shows a 1-D SDS-PAGE gel of the intact protein PF1282 from *P. furiosus* used for the experiments described here.

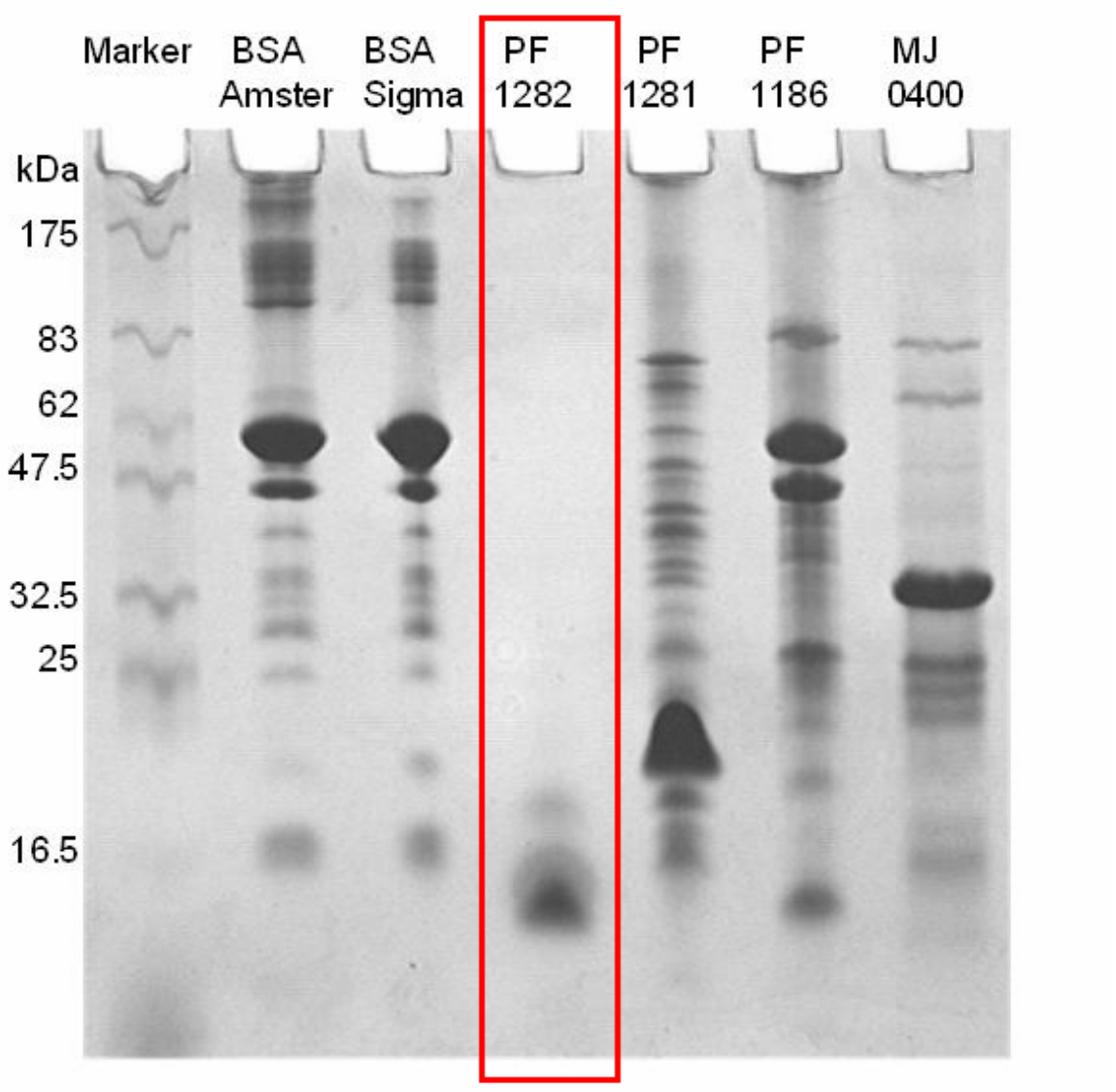

Figure 4.7 . 1-D SDS-PAGE Coomassie blue stained gel for protein PF1282 from *P. furiosus.*

First, the “standard” overnight digestion protocol was used to obtain tryptic fragments from this protein, along with the aqueous-organic solvent system, tris-HCl, $CaCl_2$ / acetonitrile (50/50), 30 minutes reaction. The amino acid sequence of this protein is shown below and, *in silico* digest of this protein generates the peptides shown in table 4.5 (masses of tryptic fragments obtained by using MS-Digest)

PF1282 Sequence:

AK ∫ WVCK ∫ ICGYIYDEDAGDPDNGISPGTK ∫ FEELPDDWVCPICGAPK ∫ SEFEK ∫ LED

Monoisotopic m/z	Start	End	Missed Cleavages	Sequence
734.4023	1	6	1	(-)AKWVCK(I)
996.4526	46	53	1	(K)SEFEKLED(-)
1918.8719	29	45	0	(K)FEELPDDWVCPICGAPK(S)
2300.0029	7	28	0	(K)ICGYIYDEDAGD PDNGISPGTK(F)
2539.1525	29	50	1	(K)FEELPDDWVCPICGAPKSEFEK(L)
2816.2548	3	28	1	(K)WVCKICGYIYDEDAGDPDNGISPGTK(F)

Table 4.5 Predicted tryptic fragments for PF1282 with mass above 700 amu

Figure 4.8 shows the results obtained for both samples when analyzed by MALDI-FTICR. Overnight digestion did not effectively generate tryptic fragments for

PF1282; only one expected peak shows at m/z 2300.0007 and it exhibits low signal intensity when compared with the same m/z peak obtained in 30 minutes by the use of the aqueous-organic solvent system. The peptide mixture solution in 50% acetonitrile shows four tryptic peptides that fully covered the amino acid sequence of this protein (Table 4.6). This methodology seems to be very effective in the case of the protein PF1282. The archaeal strain *P. furiosus* exhibits optimal growth at 100 °C which could explain why the protein PF1282 was not properly cleaved by the standard overnight digest.[48] Thermal denaturation is the only mean to unfold the protein in this protocol and, most likely, this protein is resistant to denaturation at the temperature used, therefore trypsin was not able to effectively cleave at the expected sites.[33] Fortunately, in the case of the mixture tris-HCl and acetonitrile, the organic solvent aids in the denaturation of the protein and allows it to be fragmented as expected.[32]

m/z obtained	Error (ppm)	Start	End	Missed Cleavages	Sequence
734.4023	0.0	1	6	1	(-)AKWVCK(I)
996.4524	-0.2	46	53	1	(K)SEFEKLED(-)
1918.8702	-0.9	29	45	0	(K)FEELPDDWVCPICGAPK(S)
2300.0013	-0.7	7	28	0	(K)ICGYIYDEDAGDPDNGISPGTK(F)

Table 4.7 Observed tryptic fragments for PF1282

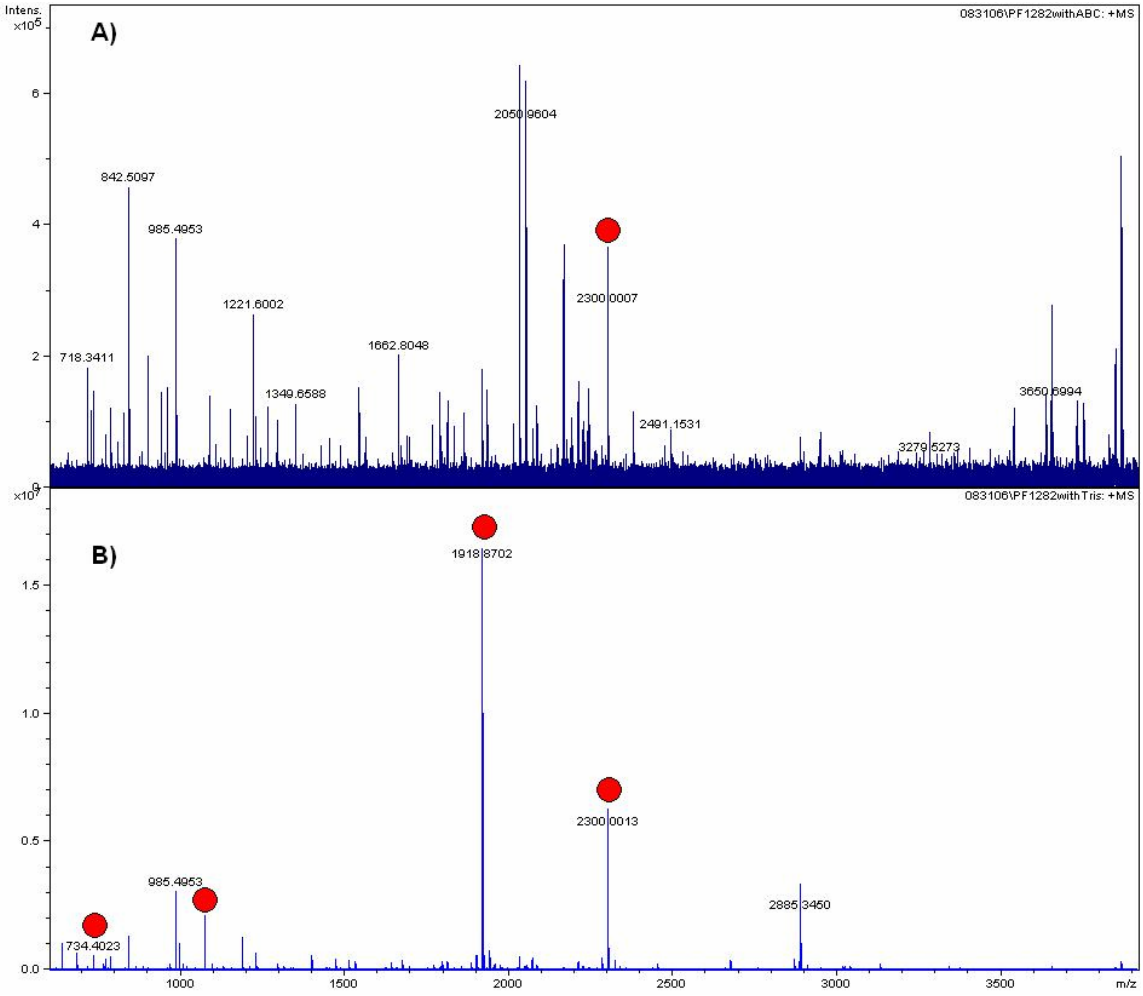

Figure 4.8 . MALDI-FTICR mass spectra obtained of tryptic digestion of PF1282 in A) ABC overnight digest and, B) Tris-HCl, $CaCl_2$ / ACN (50/50) 30 minutes digest (expected tryptic fragments are denoted with a circle)

As in the case of protein standards, some unexpected peaks are observed. Investigation of the origin of these peptides is required in order to determine if any non-

specific cleavage is taking place during our approach to trypsinolysis or if the peptides observed are solely originated by digestion of impurities contained in the original sample. To qualitatively determine the purity of the sample, we first analyzed the intact protein (original sample) by MALDI-FTICR and, the spectrum is shown in Figure 4.9. A number of unexpected sample components are observed in the range from 700 to 4000 m/z. These unknown fragments could undergo proteolysis along with the protein of interest and produce tryptic fragments of unknown sequence that would explain the presence of unassignable peptides in the analysis of the protein PF1282. However, in order to further investigate the origin of the unassignable peptides and determine if non-specific digestion is also responsible for some of the peptides, MS/MS data is required. We were facilitated MS/MS data for a digested solution of the protein PF1282, obtained by a Finnigan LTQ Linear Ion Trap Mass Spectrometer (Thermo Finnigan, San Jose, CA). Data confirmed the presence of one of the peptides found by MALDI-FTICR (2300.0029 m/z, ICGYIYDEDAGDPDNGISPGTK) and, it suggests the presence of a few peptides that could originate from non-specific cleavages (YIYDEDAGDPDNGISPGTK, GYIYDEDAGDPDNGISPGTK, NGISPGTK, and AGDPDNGISPGTK). These “non-specific” peptides do not exhibit high intensity and they could be generated from fragments of the protein already present in the original sample. A more complete approach to determine how specific trypsin is, would include MS/MS data of the original sample in order to fully determine the identity of the fragments observed by MALDI-FTICR and then, we could conclude about the origin of the resulting peptides. Of course, an even better way to study the specificity of the trypsin digestion by our protocol would

involve the use of a highly purified protein, in which no contaminants would complicate the analysis.

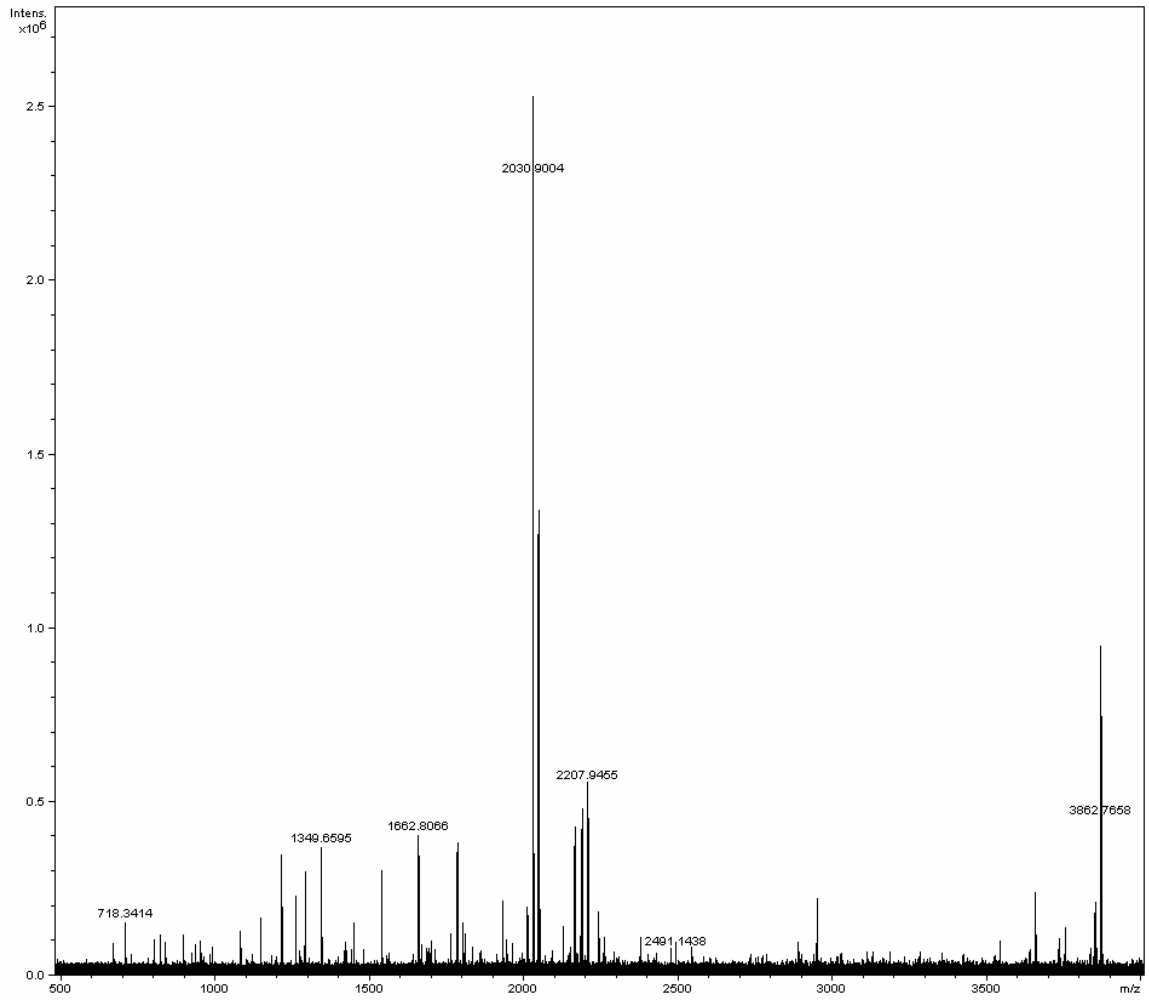

Figure 4.9 . MALDI-FTICR mass spectrum obtained of intact protein PF1282 dissolved in Tris-HCl, $CaCl_2$ / ACN (50/50)

Conclusions: The approach for trypsin digestion presented in this chapter could greatly improve the analysis of protein mixtures by decreasing the amount of unassignable masses generated during proteolysis and, by decreasing the digestion time from over 10 h to only 30 minutes which can also improve the high-throughput capabilities of the experiment. The methodology described here makes use of an aqueous-organic solvent system (50/50 v/v) which aids denaturation of proteins that can be difficult to unfold by application of heat only and, it also allows the digestion to occur at a much faster rate.[23, 32] Protein coverage was higher for all the proteins digested by the proposed approach than that obtained for overnight digested solutions. Also a significant decrease of unassignable masses was obtained by the shorter time digestion, suggesting that overnight reaction of trypsin could provide enough time for significant non-specific cleavages and autolysis to occur. As demonstrated for the protein PF1282, which seems to be proteolysis resistant under standard trypsin digest conditions, this approach might also contribute to obtain peptides from proteins that normally would be protease-resistant. This could enhance protein identification in proteomic analysis, since it is known that some proteins in a complex protein mixture do not generate peptides in the conditions that the majority of the components of the samples do.[38] Therefore, presence of peptides corresponding to proteins usually underrepresented in the peptide solution would allow identification of more components of the proteome.

References

(1) Aebersold, R.; Goodlett, D. R. Mass spectrometry in proteomics. *Chemical Reviews* **2001**, *101*, 269-295.

(2) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.

(3) Guerrera, I. C.; Kleiner, O. Application of mass spectrometry in proteomics. *Bioscience Reports* **2005**, *25*, 71-93.

(4) Link, A. J. Multidimensional peptide separations in proteomics. *Trends in Biotechnology* **2002**, *20*, S8-S13.

(5) Lill, J. Proteomic tools for quantitation by mass spectrometry. *Mass Spectrometry Reviews* **2003**, *22*, 182-194.

(6) Wu, C. C.; MacCoss, M. J. Shotgun proteomics: Tools for the analysis of complex biological systems. *Current Opinion in Molecular Therapeutics* **2002**, *4*, 242-250.

(7) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology* **1999**, *17*, 676-682.

(8) Yates, J. R. Mass spectral analysis in proteomics. *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33*, 297-316.

(9) Bienvenut, W. V.; Deon, C.; Pasquarello, C.; Campbell, J. M.; Sanchez, J. C.; Vestal, M. L.; Hochstrasser, D. F. Matrix-assisted laser desorption/ionization-

tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* **2002**, *2*, 868-876.

(10) Cagney, G.; Amiri, S.; Premawaradena, T.; Lindo, M.; Emili, A. In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Science* **2003**, *1*, 1-15.

(11) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *Journal of Proteome Research* **2002**, *1*, 47-54.

(12) Perham, R. N.; Jones, G. M. T. Determination of Order of Lysine-Containing Tryptic Peptides of Proteins by Diagonal Paper Electrophoresis - a Carboxyl-Terminal Sequence for Pepsin. *European Journal of Biochemistry* **1967**, *2*, 84-&.

(13) Olsen, J. V.; Ong, S. E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & Cellular Proteomics* **2004**, *3*, 608-614.

(14) Baldwin, M. A. Protein identification by mass spectrometry - Issues to be considered. *Molecular & Cellular Proteomics* **2004**, *3*, 1-9.

(15) Noda, Y.; Fujiwara, K.; Yamamoto, K.; Fukuno, T.; Segawa, S. Specificity of Trypsin Digestion and Conformational Flexibility at Different Sites of Unfolded Lysozyme. *Biopolymers* **1994**, *34*, 217-226.

(16) Stead, D. A.; Preece, A.; Brown, A. J. P. Universal metrics for quality assessment of protein identifications by mass spectrometry. *Molecular & Cellular Proteomics* **2006**, *5*, 1205-1211.

(17) Tang, Y.; Chen, Y. F.; Lichti, C. F.; Hall, R. A.; Raney, K. D.; Jennings, S. F. CLPM: A cross-linked peptide mapping algorithm for mass spectrometric analysis. *Bmc Bioinformatics* **2005**, *6*.

(18) Levander, F.; Rognvaldsson, T.; Samuelsson, J.; James, P. Automated methods for improved protein identification by peptide mass fingerprinting. *Proteomics* **2004**, *4*, 2594-2601.

(19) Ding, Q. X.; Xiao, L.; Xiong, S. X.; Jia, Y. F.; Que, H. P.; Guo, Y. J.; Liu, S. J. Unmatched masses in peptide mass fingerprints caused by cross-contamination: An updated statistical result. *Proteomics* **2003**, *3*, 1313-1317.

(20) Ekstrom, S.; Onnerfjord, P.; Nilsson, J.; Bengtsson, M.; Laurell, T.; Marko-Varga, G. Integrated microanalytical technology enabling rapid and automated protein identification. *Analytical Chemistry* **2000**, *72*, 286-293.

(21) Freije, J. R.; Mulder, P.; Werkman, W.; Rieux, L.; Niederlander, H. A. G.; Verpoorte, E.; Bischoff, R. Chemically modified, immobilized trypsin reactor with improved digestion efficiency. *Journal of Proteome Research* **2005**, *4*, 1805-1813.

(22) Slysz, G. W.; Schriemer, D. C. On-column digestion of proteins in aqueous-organic solvents. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 1044-1050.

(23) Russell, W. K.; Park, Z. Y.; Russell, D. H. Proteolysis in mixed organic-aqueous solvent systems: Applications for peptide mass mapping using mass spectrometry. *Analytical Chemistry* **2001**, *73*, 2682-2685.

(24) Strader, M. B.; Tabb, D. L.; Hervey, W. J.; Pan, C. L.; Hurst, G. B. Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems. *Analytical Chemistry* **2006**, *78*, 125-134.

(25) Li, A. Q.; Sowder, R. C.; Henderson, L. E.; Moore, S. P.; Garfinkel, D. J.; Fisher, R. J. Chemical cleavage at aspartyl residues for protein identification. *Analytical Chemistry* **2001**, *73*, 5395-5402.

(26) Bark, S. J.; Muster, N.; Yates, J. R.; Siuzdak, G. High-temperature protein mass mapping using a thermophilic protease. *Journal of the American Chemical Society* **2001**, *123*, 1774-1775.

(27) Pramanik, B. N.; Mirza, U. A.; Ing, Y. H.; Liu, Y. H.; Bartner, P. L.; Weber, P. C.; Bose, M. K. Microwave-enhanced enzyme reaction for protein mapping by mass spectrometry: A new approach to protein digestion in minutes. *Protein Science* **2002**, *11*, 2676-2687.

(28) Lopez-Ferrer, D.; Capelo, J. L.; Vazquez, J. Ultra fast trypsin digestion of proteins by high intensity focused ultrasound. *Journal of Proteome Research* **2005**, *4*, 1569-1574.

(29) Havlis, J.; Thomas, H.; Sebela, M.; Shevchenko, A. Fast-response proteomics by accelerated in-gel digestion of proteins. *Analytical Chemistry* **2003**, *75*, 1300-1306.

(30) Simon, L. M.; Laszlo, K.; Vertesi, A.; Bagi, K.; Szajani, B. Stability of hydrolytic enzymes in water-organic solvent systems. *Journal of Molecular Catalysis B-Enzymatic* **1998**, *4*, 41-45.

(31) Simon, L. M.; Kotorman, M.; Garab, G.; Laczko, I. Structure and activity of alpha-chymotrypsin and trypsin in aqueous organic media. *Biochemical and Biophysical Research Communications* **2001**, *280*, 1367-1371.

(32) Griebenow, K.; Klibanov, A. M. On protein denaturation in aqueous-organic mixtures but not in pure organic solvents. *Journal of the American Chemical Society* **1996**, *118*, 11695-11700.

(33) Jenney, F. E.; Adams, M. W. W. In *Hyperthermophilic Enzymes, Pt C*, 2001; Vol. 334, pp 45-55.

(34) Sambrook, E. F.; Maniatis, T. *Molecular cloning: A laboratory manual*: Cold Spring Harbor, NY, (1989).

(35) Wong, R. L.; Amster, I. J. Combining low and high mass ion accumulation for enhancing shotgun proteome analysis. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 205-212.

(36) Patterson, S. D.; Aebersold, R. H. Proteomics: the first decade and beyond. *Nature Genetics* **2003**, *33*, 311-323.

(37) Kolker, E.; Higdon, R.; Hogan, J. M. Protein identification and expression analysis using mass spectrometry. *Trends in Microbiology* **2006**, *14*, 229-235.

(38) Patterson, S. D. How much of the proteome do we see with discovery-based proteomics methods and how much do we need to see? *Current Proteomics* **2004**, *1*, 3-12.

(39) Chalkley, R. J.; Baker, P. R.; Hansen, K. C.; Medzihradszky, K. F.; Allen, N. P.; Rexach, M.; Burlingame, A. L. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole

selecting, quadrupole collision cell, time-of-flight mass spectrometer - I. How much of the data is theoretically interpretable by search engines? *Molecular & Cellular Proteomics* **2005**, *4*, 1189-1193.

(40) Fontana, A.; Zambonin, M.; deLaureto, P. P.; DeFilippis, V.; Clementi, A.; Scaramella, E. Probing the conformational state of apomyoglobin by limited proteolysis. *Journal of Molecular Biology* **1997**, *266*, 223-230.

(41) Park, Z. Y.; Russell, D. H. Thermal denaturation: A useful technique in peptide mass mapping. *Analytical Chemistry* **2000**, *72*, 2667-2670.

(42) Park, Z. Y.; Russell, D. H. Identification of individual proteins in complex protein mixtures by high-resolution, high-mass-accuracy MALDI TOF-mass spectrometry analysis of in-solution thermal denaturation/enzymatic digestion. *Analytical Chemistry* **2001**, *73*, 2558-2564.

(43) Calbiochem. *Retrieved from* *http://www.emdbiosciences.com/Products/ProductDisplay.asp?catno=12657& on September 12th, 2006*.

(44) Fraser, D.; Powell, R. E. The Kinetics of Trypsin Digestion. *Journal of Biological Chemistry* **1950**, *187*, 803-820.

(45) Gabel, D.; Kasche, V. Cooperative Transitions between Active Alpha-Trypsin and Beta-Trypsin Conformations. *Biochemical and Biophysical Research Communications* **1972**, *48*, 1011-&.

(46) Johnson, P.; Whateley, T. L. The Effects of Surface and Macromolecular Interactions on the Kinetics of Inactivation of Trypsin and Alpha-Chymotrypsin. *Biochemical Journal* **1981**, *193*, 285-294.

(47) Matrix Science. *Retrieved from Thttp://www.matrixscience.com/help/autolysis_help.html on September 12th, 2006*.

(48) Maeder, D. L.; Weiss, R. B.; Dunn, D. M.; Cherry, J. L.; Gonzalez, J. M.; DiRuggiero, J.; Robb, F. T. Divergence of the hyperthermophilic archaea Pyrococcus furiosus and P. horikoshii inferred from complete genomic sequences. *Genetics* **1999**, *152*, 1299-1305.

CHAPTER 5

CONCLUSIONS

Proteomics provides a way to qualitatively and quantitatively analyze the full complement or subset of the proteins present in an organism, tissue or cell under a given set of physiological or environmental conditions.[1, 2] MS provides a tool to generate the required data for correlation against proteins sequence database, and steers proteomics away from *de novo* sequencing methodology.[1, 3-6] Identification of proteins can be achieved by accurate mass measurement of the peptides generated by proteolysis.[7, 8] Since different amino acid compositions can result in isobaric peptides, confident identification of the amino acid sequence can be challenging at times.[2] This thesis describes two different approaches to facilitate proteomics analysis by accurate mass measurements using FTICR-MS. The first approach discussed in this thesis seeks to improve peptide assignment in proteomic analyses and, the second approach is geared toward enhancing the effectiveness of the trypsinolysis of protein mixture solutions.

The first experimental chapter presented in this thesis (chapter 3) describes a method for improving the identification of peptides in a shotgun proteome analysis using accurate mass measurement. The improvement is based upon the derivatization of cysteine residues with a novel reagent, 2,4–dibromo-(2'-iodo) acetanilide. The derivatization changes the mass defect of cysteine-containing proteolytic peptides in a manner that increases their identification specificity. Peptide masses were measured using MALDI mass spectrometry. Reactions with protein standards showed that the derivatization of cysteine is rapid and quantitative, and the data suggests that the derivatized peptides are more easily ionized and/or detected than unlabeled cysteine-containing peptides. The reagent was also tested on a ^{15}N-metabolically-labeled

proteome from *M. maripaludis*. Proteins were identified by their accurate mass values and from their nitrogen stoichiometry. Results showed that 47% of cysteine-labeled peptides were identified versus 27% identification for non-cysteine containing peptides. Also, this procedure permits the identification of proteins from the *M. maripaludis* proteome that are not usually observed by the standard protocol, and shows that better protein coverage is obtained with this methodology.[9]

Despite the peptide assignment improvement observed by the mass defect labeling approach, there was still a fraction of peptides that could not be assigned to any of the predicted fragments. This is a known fact for proteomic analysis and, most of the time these fragments are attributed to peptides resulting from nonspecific cleavages or peptides containing residues carrying post-translational or chemical modifications.[10] Since this portion of unassignable peptides can make up for a significant percentage of the total masses obtained for an specific study, we decided to approach this issue by modifying the conditions of the trypsinolysis.[11] Therefore, chapter 4 describes a methodology to digest proteins that makes use of acetonitrile in order to aid protein denaturation and accelerate the digestion process. A standard trypsin digest requires about 12 hours to complete proteolysis of the sample components.[12] Our approach decreases the digestion time to only 30 minutes and, more importantly, we observed a significant decrease of unassignable masses generated during proteolysis. The presence of an aqueous-organic solvent system also aids denaturation of proteins that can be difficult to unfold by application of heat only and, it also allows the digestion to occur at a much faster rate.[13, 14] Analysis of standard proteins showed higher protein coverage for peptide solution obtained by the proposed method as compared to the

overnight standard protocol. Another advantage of the approach utilized here is the digestion of protease-resistant proteins. As demonstrated for the rubredoxin protein (PF1282) from *P. furiosus* which seems to be proteolysis resistant in standard trypsin digest conditions, full sequence coverage of the protein was obtained by the aqueos-organic type of digestion. This finding could enhance protein identification in proteomic analysis, since some proteins in a complex protein mixture do not easily generate peptides under the standard conditions that the majority of the components of the samples do.[15] Therefore, the ability to generate peptides from proteins usually underrepresented in the peptide solution would increase the number of components identified. Applying the proposed digestion protocol to protein mixtures could potentially lead to significant improvement of the shotgun proteomic analysis.

References

(1) Bogdanov, B.; Smith, R. D. Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrometry Reviews* **2005**, *24*, 168-200.

(2) Aebersold, R.; Goodlett, D. R. Mass spectrometry in proteomics. *Chemical Reviews* **2001**, *101*, 269-295.

(3) Domon, B.; Aebersold, R. Review - Mass spectrometry and protein analysis. *Science* **2006**, *312*, 212-217.

(4) Griffin, T. J.; Goodlett, D. R.; Aebersold, R. Advances in proteome analysis by mass spectrometry. *Current Opinion in Biotechnology* **2001**, *12*, 607-612.

(5) Kolker, E.; Higdon, R.; Hogan, J. M. Protein identification and expression analysis using mass spectrometry. *Trends in Microbiology* **2006**, *14*, 229-235.

(6) Baldwin, M. A. Protein identification by mass spectrometry - Issues to be considered. *Molecular & Cellular Proteomics* **2004**, *3*, 1-9.

(7) Chan, T. W. D.; Duan, L.; Sze, T. P. E. Accurate mass measurements for peptide and protein mixtures by using matrix-assisted laser desorption/ionization Fourier transform mass spectrometry. *Analytical Chemistry* **2002**, *74*, 5282-5289.

(8) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.

(9) Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. Mass defect labeling of cysteine for improving peptide assignment in shotgun proteomic analyses. *Analytical Chemistry* **2006**, *78*, 3417-3423.

(10) Levander, F.; Rognvaldsson, T.; Samuelsson, J.; James, P. Automated methods for improved protein identification by peptide mass fingerprinting. *Proteomics* **2004**, *4*, 2594-2601.

(11) Ding, Q. X.; Xiao, L.; Xiong, S. X.; Jia, Y. F.; Que, H. P.; Guo, Y. J.; Liu, S. J. Unmatched masses in peptide mass fingerprints caused by cross-contamination: An updated statistical result. *Proteomics* **2003**, *3*, 1313-1317.

(12) Retrieved from http://www.promega.com/tbs/tb309/tb309.pdf on October 19th, 2006. *Technical bulletin for trypsin digest of proteins* **Promega**.

(13) Russell, W. K.; Park, Z. Y.; Russell, D. H. Proteolysis in mixed organic-aqueous solvent systems: Applications for peptide mass mapping using mass spectrometry. *Analytical Chemistry* **2001**, *73*, 2682-2685.

(14) Griebenow, K.; Klibanov, A. M. On protein denaturation in aqueous-organic mixtures but not in pure organic solvents. *Journal of the American Chemical Society* **1996**, *118*, 11695-11700.

(15) Patterson, S. D. How much of the proteome do we see with discovery-based proteomics methods and how much do we need to see? *Current Proteomics* **2004**, *1*, 3-12.