ENHANCING SHOTGUN PROTEOMICS BY MASS DEFECT LABELING OF TRYPTOPHAN, IMPROVING MASS MEASUREMENT ACCURACY AND ON-TARGET DIGESTION

by

CHUNYAN LI

(Under the Direction of I. Jonathan Amster)

ABSTRACT

Proteomics is the large-scale study of the proteins expressed in a cell, tissue and organism under a given set of physiological or environmental conditions. Mass spectrometry (MS) in combination with a variety of separation methods has become a primary tool for proteomic studies. Of all proteomic approaches, shotgun proteomics has gained tremendous popularity in recent years. This method involves batch proteolysis of a protein mixture, separation of peptides by high performance liquid chromatography (HPLC), and protein identification from mass spectrometric analysis of peptides. This dissertation describes three approaches to facilitate shotgun proteomic analysis by accurate mass measurement using Fourier transform ion cyclotron resonance mass spectrometry (FTICR/MS). The first approach, referred to as mass defect labeling (MDL), is based on the derivatization of tryptophan residues in protein with a reagent that alters their mass defects. We describe here the development and application of a tryptophan MDL reagent that is compatible to matrix-assisted laser desorption/ionization (MALDI). The second approach employs the stored waveform inverse Fourier transform (SWIFT) excitation to improve the mass measurement accuracy for higher mass peptides. We demonstrate that sub

part-per-million (ppm) can be achieved by combining SWIFT with stepwise-external calibration. The third approach couples accurate mass measurement with nitrogen stoichiometry to improve protein identification in shotgun proteomics using HPLC-MALDI-FTICR/MS. We present here the results from identification and quantitation of the proteins in a $^{15}N$-metabolically labeled proteome sample that was obtained from *Methanococcus maripaludis.* The forth approach combines on-target digestion with MALDI-FTICR/MS analysis for rapid protein identification. On-target digestion of a HPLC separated proteome with MALDI-FTICR/MS analysis offers an alternative to the standard method of shotgun analysis.

INDEX WORDS: Shotgun proteomics, Mass spectrometry, FTICR, MALDI, Accurate mass measurement, $^{15}N$ metabolic labeling, Mass defect labeling, SWIFT, Mass accuracy, Digestion

# ENHANCING SHOTGUN PROTEOMICS BY MASS DEFECT LABELING OF TRYPTOPHAN, IMPROVING MASS MEASUREMENT ACCURACY AND ON-TARGET DIGESTION

by

CHUNYAN LI

B.S., China Textile University, 1998 and M.S., Donghua University, 2001

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

# ENHANCING SHOTGUN PROTEOMICS BY MASS DEFECT LABELING OF TRYPTOPHAN, IMPROVING MASS MEASUREMENT ACCURACY AND ON-TARGET DIGESTION

by

CHUNYAN LI

Major Professor: I. Jonathan Amster

Committee: Lance Wells
Ron Orlando

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

# DEDICATION

This dissertation is dedicated firstly to my parents, Ze Li and Xuanying Du. Thank them for their unconditional love and support throughout my life. Wang Lu, my husband has been standing by my side, as always, fulfilling me with passion and love. And this dedication also goes to my beloved baby son, Caleb J. Lu, who has already brought and will keep bringing me much of joy, love and a lot more.

ACKNOWLEDGEMENTS

My deepest and first acknowledgement goes to my major advisor, Dr. I. Jonathan Amster, for his support and guidance throughout my study and research at UGA. This research and my dissertation would not be existent without him. Great and sincere thanks have to go to my committee members, Dr. Lance Wells, Dr. Robert S. Phillips and Dr. Ron Orlando. They have been consistently giving me their valuable time and advice on writing my dissertation.

There are also many people that have helped on providing me with their lab facility and their brain-storming discussions and many other unforgettable and greatly appreciated things. Among them, Dr. William B. Whitman and Yuchen Liu helped me on preparing the $^{15}N$-metabolically labeled proteome sample, Vijay Gawandi and Adam Protos supported me on synthesizing the tryptophan labeling reagents, and Peng Zhao helped me on performing the protein separation. My group members, Richard Wong, Bryan Parks, Jeremy Wolff, Hilda Barry, Dhanashri Bagal, Jing Li, Lisabeth Hoffman, Melissa Warren and Franklin Leach, besides all inspiring suggestions and talks, have also made my journey in the group full of fun and joy.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

# CHAPTER 1

# INTRODUCTION AND LITERATURE REVIEW

## 1. PROTEOMICS

The term “proteomics” was coined in 1994 to refer to a systematic study of the proteins present in a cell, tissue or organism [1]. Mass spectrometry (MS) has become a primary analytical tool for proteomic analysis since the introduction of two “soft” ionization techniques, electrospray ionization (ESI) [2] and matrix-assisted laser desorption/ionization (MALDI) [3, 4], which allow for mass analysis of large, thermally labile biomolecules such as proteins with little or no fragmentation. With the rapid development of MS instrumentation and the emergence of many powerful bioinformatics tools, MS-based proteomics has made significant advances in large-scale identification and quantification of proteins, characterization of post-translational modifications (PTMs), and investigation of protein-protein interactions. Currently, MS-based proteomics can be classified into two categories, bottom-up proteomics and top-down proteomics.

### 1.1 Classic Proteomics

Classic proteomics generally involves protein separation by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), protein detection by mass spectrometry, and protein identification via genome sequence database searching [5]. Over the past decades, 2D-PAGE has been the most widely used analytical technique to separate complex protein mixtures. In this method, proteins are separated based on their different isoelectric points (pI) in the first dimension, and further separated according to their relative molecular masses in the second dimension. Although this method offers high resolution for protein separation, it has a number of

limitations [6-8]. First, 2D-PAGE has a preference for certain proteins over others, such as the proteins that are very small or large, highly acidic or basic, and very hydrophobic (membrane proteins). Second, this method is biased toward high-abundance proteins, excluding the more physiologically significant low-abundance proteins. Third, 2D-PAGE is limited by its labor-intensive protocol and the difficulty in automation. To overcome these problems, researchers have moved to the use of liquid chromatography (LC) as an alternative method to 2D-PAGE in proteomic studies.

### 1.2 Bottom-Up Proteomics

Bottom-up proteomics refers to proteolytically digesting one or many proteins into smaller peptide fragments prior to mass spectrometric analysis [9-11]. The proteins can be first separated by gel electrophoresis, and then digested and analyzed by mass spectrometry (the peptide mass fingerprinting (PMF) approach). Alternatively, a complex protein mixture can be digested first, and then separated by liquid chromatography followed by mass spectrometric analysis (the shotgun proteomics approach). In the latter case, the digest can contain thousands to hundreds of thousands of peptides and thus require separation in two or more chromatographic dimensions before mass spectrometric analysis. Protein identification can be made by the comparison of the observed masses of the peptides or their tandem mass spectra with those predicted from a sequence database. Bottom-up proteomics has been considered the most mature and widely used approach for protein identification and characterization. However, this approach has some practical limitations. The complete protein sequencing is difficult to achieve as only a fraction of the total peptide population for a given protein can be identified. Moreover, the time-consuming

digestion and multidimensional LC-MS/MS (liquid chromatography - tandem mass spectrometry) analyses limit the throughput of the proteomic analysis using this approach.

### 1.2.1 Peptide Mass Fingerprinting (PMF)

Peptide mass fingerprinting (PMF) is an analytical technique for protein identification that was developed in 1993 by several groups independently [12-14]. In this method, the unknown protein of interest is first cleaved with a proteolytic enzyme (e.g. trypsin), and then the resulting peptides are analyzed by a mass spectrometer such as MALDI-TOF. The accurately measured peptide masses are compared to theoretical masses generated from the genomic database to determine the identity of the unknown protein. One disadvantage of this technique is the requirement for pure proteins. For complex mixtures of proteins, a purification step typically using two-dimensional electrophoresis (2-DE) is needed to isolate the protein of interest prior to proteolysis and mass spectrometric analysis. Another disadvantage is the requirement for several peptides to uniquely identify a protein.

### 1.2.2 Shotgun Proteomics

Of all bottom-up approaches, shotgun proteomics has gained the widest acceptance. This approach involves a batch digestion of an unseparated protein mixture, separation of the resulting peptides by single or multidimensional liquid chromatography, and protein identification from mass spectrometric analysis of peptides [15-17]. There are two mass spectrometric methods used in shotgun proteomics for protein identification. One method is accurate mass measurement of a set of peptides derived from the parent protein using high resolution mass spectrometer (*e.g.* Fourier transform ion cyclotron resonance mass spectrometer, FTICR/MS), but more commonly,

protein identification relies on the fragmentation of one or more of these peptides using tandem mass spectrometry (MS/MS). Protein identification is accomplished by matching the measured peptide masses or tandem mass spectra with the ones predicted from a genomic database. Over the past few years, a variety of shotgun proteomic approaches have been developed for improving the specificity of protein identification and the throughput of analysis. In the following sections, two widely used techniques, multidimensional protein identification technology (MudPIT) and accurate mass and time tags (AMT's), and a novel method developed in our laboratory called mass defect labeling (MDL), will be described.

#### 1.2.2.1 Multidimensional Protein Identification Technology (MudPIT)

Multidimensional protein identification technology (MudPIT), which was pioneered by Yates and coworkers [15, 16], is a technique for protein identification from a complex biological sample using multidimensional liquid chromatography coupled with tandem mass spectrometry. In this approach, the peptides generated from batch digestion of the sample are first separated by two-dimensional liquid chromatography which integrates strong cation exchange (SCX) resin and reversed-phase (RP) resin in a biphasic column, and then the eluate is directly introduced into an ESI source for MS and MS/MS analyses.

#### 1.2.2.2 Accurate Mass and Time Tags (AMT's)

The concept of "accurate mass and time (AMT)" tags was introduced by Smith and coworkers [18-21], and utilizes accurate mass measurement by FTICR mass spectrometry in combination with accurate elution time measurement by high resolution liquid chromatography. In this approach, tandem mass spectrometry is first used to establish 'potential peptide AMT

tags', and the mass and elution time for each of the identified peptides is recorded. These validated peptide AMT tags constitute a reference "look-up" for subsequent peptide identification based on retention time and accurate mass measurement. Peptides are considered to be identified once their measured mass and retention time match with the values in the "look-up" table. In this approach, LC-MS/MS analysis needs to be performed only once for creating peptide AMT tags for a particular biological system, thus the throughput of proteomic analysis is significantly improved.

#### 1.2.2.3 Mass Defect Labeling (MDL)

"Mass defect" is defined as the difference between the exact mass of a compound and its nominal molecular weight calculated using integer atomic masses. Peptides are primarily composed of the elements from the first two rows of the periodic table, most of whose mass defects fall into the range of +/-0.008 amu (C, 0.0 mmu; H, 7.8 mmu; O, -5.1 mmu; N, 3.1 mmu; S, -28.0 mmu). In contrast to small molecules, peptides have significant mass defects due to the numerous atoms present; however, the distribution of the peptide mass defects is generally narrow (Figure 1.1a). The peptide molecular weights occupy only one-third of a unit mass at any given nominal mass, which causes the peptide masses to overlap and therefore reduces the specificity of peptide assignment based on accurate mass measurement. Greater specificity would be possible if the peptide masses are distributed more evenly across the mass scale. This can be achieved by a method called mass defect labeling (MDL), which is based on the derivatization of selected amino acids with a reagent that alters their mass defects (Figure 1.1b). Cysteine and tryptophan are two attractive targets for MDL, as they are less frequently occurring amino acids. Cysteine and tryptophan occur in the amino acid composition of a typical protein at

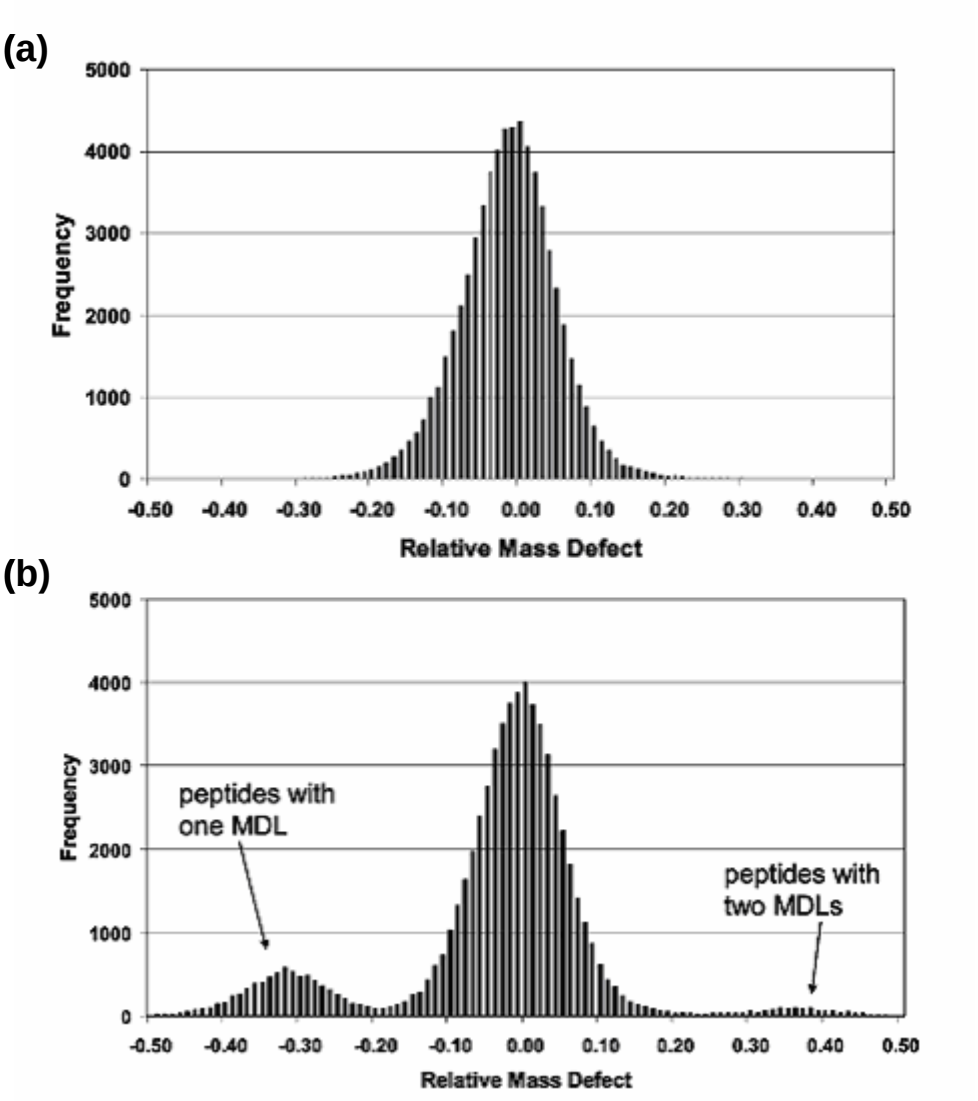



**Figure 1.1** Composite distribution of mass defects for all tryptic peptides of *M. maripaludis* with molecular weights between 700 and 3500 amu (a) before and (b) after MDL-labeling all the cysteine-containing peptides. In (a), the horizontal axis is the mass difference (amu) between a peptide's mass defect and the average mass defect for all peptides of the same nominal mass. In (b), the central distribution corresponds to all peptides that do not contain cysteine; the singly and doubly labeled cysteine-containing peptides appear in the smaller distribution centered at -0.30 amu and +0.40 amu, respectively. Adapted from Hernandez *et al.* [22]

a rate of 1.7 % and 1.3 % respectively [23], and appear in about only 15-20 % of tryptic peptides. Therefore, mass defect labeling of these two residues can significantly reduce the complexity of mass spectrum by shifting a small portion of peptides (labeled peptides) to a region of mass scale that is not occupied by unlabeled peptides. In this approach, the elements with large mass defect, such as bromine (-82.0 mmu), are incorporated into a tagging reagent to serve as mass defect label. Hernandez *et al*. recently reported a novel cysteine mass defect label, 2,4-dibromo-(2'-iodo)acetanilide, and its applications on the shotgun proteomic analysis of *Methanococcus maripaludis* using HPLC-MALDI-FTICR mass spectrometry, demonstrating the effectiveness of this approach on improving the specificity of protein identification [22].

### 1.3 Top-Down Proteomics

Top-down proteomics is based on the ionization of intact proteins or large protein fragments followed by tandem mass spectrometric analysis [10, 11, 24, 25]. Generally, the intact protein molecular ions produced by electrospray ionization (ESI) are introduced into a Fourier transform ion cyclotron resonance (FTICR) or quadrupole ion trap (QIT) mass spectrometer for mass measurement and tandem mass spectrometric analysis. The fragmentation methods commonly used for top-down proteomics are electron capture dissociation (ECD) and electron transfer dissociation (ETD). The two major advantages of top-down proteomics are the potential access to the complete protein sequence and the ability to locate and characterize post-translational modifications (PTMs). In addition, the time-consuming protein digestion required for the bottom-up strategy is eliminated. As relatively young in proteomics, however, the top-down strategy currently suffers from several limitations. Firstly, the highly complex mass spectra generated from multiply charged proteins are difficult to interpret. Secondly, the top-down

strategy only works well for the relatively small proteins (< 50 kDa) [26] and simple protein mixtures. Nonetheless, the top-down strategy provides an alternative to the bottom-up strategy and shows its promise in proteomic analysis.

### 1.4 Quantitative Proteomics

Quantitative proteomics has traditionally been carried out using gel-based methods such as 2D-PAGE followed by identification using mass spectrometry. To overcome the problems inherent with gel-based methods, gel-free methods, mainly based on reverse-phase liquid chromatography coupled with stable isotopic labeling and mass spectrometry, and more recently label-free quantitation, have become more widely used for quantitative proteomics. In contrast to label-free quantitation, which is more bioinformatics-based [27, 28], stable isotopic labeling is easily implemented and more frequently used in current quantitative proteomics. In this approach, two or more samples representing different cell states are differently labeled, equally combined, processed, and analyzed by mass spectrometry [29-34]. The labeling is performed by introducing a tag that is chemical equivalent but has different stable isotope composition, allowing for the relative quantitation of the labeled samples without changing their biochemical properties. The ratio of intensities of the differentially labeled peptides can be used to determine the relative abundance of the peptides and therefore the proteins in the original cell states. Stable heavy isotopes, such as $^{2}H$, $^{13}C$, $^{15}N$ and $^{18}O$, can be incorporated either metabolically, enzymatically or via chemical reactions into proteins and peptides. Accordingly, three strategies of stable isotopic labeling are presented here: (1) metabolic labeling (e.g. SILAC); (2) enzymatic labeling (e.g. proteolytic $^{18}O$ labeling); and (3) chemical labeling using isotopic tags (e.g. ICAT and iTRAQ$^{TM}$).

### 1.4.1 Metabolic Labeling

Metabolic labeling is based on growing two populations of cells in isotopically-labeled media of different enrichment. After isotope incorporation, the light- and heavy-labeled samples are combined, digested, and analyzed by mass spectrometry. The relative quantitation is made by determining the ratio of intensities of light/heavy peptide pairs from mass spectra. $^{15}$N-enriched media were originally used to metabolically label the proteins in cell lines and small organisms [35, 36]. Recently, a more efficient technique, SILAC (stable isotope labeling by amino acids in cell culture), was developed by Mann's group [37, 38], involving the growth of cells in media containing isotopically labeled amino acid. In this approach, leucine [37], lysine [39], methionine [40], and arginine [41] are most frequently used for isotope incorporation. The main advantage of SILAC, which also holds true for $^{15}$N metabolic labeling, is minimizing the variation between two samples as isotope incorporation is performed in the early stage of the preparative procedures. Since labeling takes place on the protein level, full protein information, such as post translational modifications, can be preserved. In contrast to other labeling strategies, where a defined mass shift is introduced into every labeled peptide, the mass shift produced using SILAC or $^{15}$N metabolic labeling may vary from peptide to peptide as it depends on the number of heavy isotopes incorporated into the peptides. This feature complicates the quantitation; however, it provides additional sequence information that can be used to facilitate peptide identification. The disadvantages of SILAC include high cost, incomplete incorporation when applied to plant cell lines, and inability to label tissue samples and body fluids. Nonetheless, SILAC is still one of the most effective labeling strategies for quantitative proteomics.

### 1.4.2 Enzymatic Labeling

Enzymatic labeling, mainly referred to as proteolytic $^{18}O$ labeling, is a method that relies on incorporation of $^{18}O$ atoms into the C-terminal carboxylic group of peptides during proteolysis in the presence of $H_2{}^{18}O$ [42, 43]. The $^{18}O$-labeled peptides are equally combined with the peptides that are digested in unlabeled water, and then subjected to MS analysis for quantitation. This technique typically incorporates two $^{18}O$ atoms into a peptide, resulting in the mass shift of 4 Da, although a minor component with one labeled oxygen is often observed. The commonly used enzymes include trypsin [44], chymotrypsin [45] and Glu-C [46]. This technique is advantageous due to the simplicity of the experiment and the ability to label all types of samples including body fluids. However, the following drawbacks render this technique less popularly used than the metabolic and chemical labeling methods: First, proteolytic $^{18}O$ labeling is performed downstream of procedures, introducing the opportunity for technical variation between two samples; Second, the small mass difference complicates the quantitation, especially when low resolution mass spectrometers are used; Third, trypsin-mediated $^{16}O/^{18}O$ back-exchange can occur after mixing two samples, even under the conditions of low pH and temperature where the trypsin is assumed to be inactivated [47], making the determination of $^{16}O/^{18}O$ ratio and therefore quantitation inaccurate.

### 1.4.3 Chemical Labeling Using Isotopic Tags

Chemical labeling using stable isotopes is the most widely used method in current quantitative proteomics. This technique is based on selectively labeling a specific site on a peptide or protein with an isotopic tag. Two popular chemical reagents, isotope-coded affinity tags (ICAT) and isobaric tag for relative and absolute quantitation (iTRAQ$^{TM}$), are introduced.

#### 1.4.3.1 Isotope-Coded Affinity Tags (ICAT)

As one of the earliest chemical reagents introduced for quantitative proteomics, the ICAT (isotope-coded affinity tags) was first reported in 1999 by Aebersold's group [48]. This approach is used to modify cysteine-containing proteins that are derived from two different cell states. As shown in Figure 1.2a, the original ICAT reagent consists of a reactive group with specificity towards thiol groups in cysteine residues, an isotope coded linker region containing eight $^{1}H$ (light version) or eight $^{2}H$ (heavy version) atoms, and a biotin group that allows for isolation of the ICAT-labeled peptides via biotin-avidin affinity chromatography. In this approach, two samples are individually labeled with light and heavy ICAT reagents, equally combined and digested by trypsin. The ICAT-labeled cysteine-containing peptides are enriched by biotin-avidin affinity chromatography and subjected to MS and MS/MS analyses (Figure 1.2b). The original ICAT reagent was improved using $^{12}C/^{13}C$ [49] instead of $^{1}H/^{2}H$ because the use of $^{2}H$ results in a differential elution between the light- and heavy-labeled peptides during liquid chromatographic separation [50]. Furthermore, acid-cleavable ICAT reagents were developed to improve the MS and MS/MS performance by removing the bulky biotin group in MS analysis [51]. Cysteine is a rare amino acid (1.7 % in the amino acid composition of a typical protein [23]), thus capture of the labeled cysteine-containing peptides significantly reduces the sample complexity. However, this feature gives ICAT a drawback in that the proteins that do not contain cysteine residues can not be quantified using this approach. Nonetheless, the ICAT remains one of the most widely used chemical approaches for quantitative proteomics.

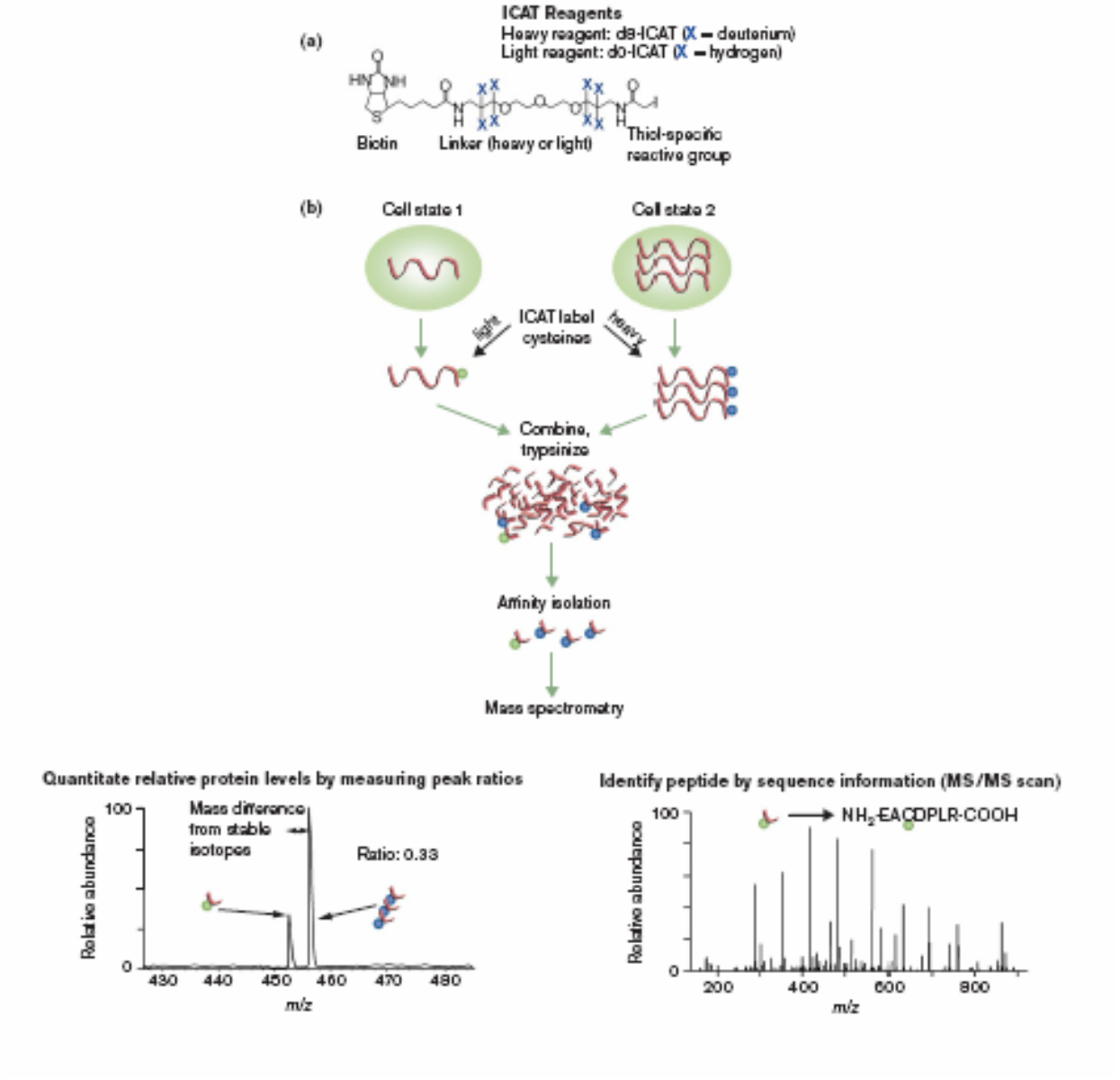



**Figure 1.2** The ICAT strategy for quantifying differential protein expression. (a) Structure of the original ICAT reagent; (b) Schematic of the ICAT strategy. Adapted from Gygi *et al.* [52]

#### 1.4.3.2 Isobaric Tag for Relative and Absolute Quantitation (iTRAQ™)

Isobaric tag for relative and absolute quantitation (iTRAQ™) is a amine-specific chemical tagging method and was first reported in 2004 by Applied Biosystems [53]. This method involves labeling of N-termini and lysine side chains of the peptides in a digest mixture with tags of varying mass. The iTRAQ reagent consists of three distinct groups: a reporter group, a balancer group, and an amine-reactive group (Figure 1.3). Both the reporter group and the balancer group contain a certain number of differentially tagged sites, and they are assembled in a manner that the sum of their masses are kept constant (145 Da) by differential isotopic enrichment. Four peptides of same sequence are individually labeled with four sets of iTRAQ reagents and then analyzed by MS/MS. While the labeled peptides are mass indistinguishable, the reporter ions generated from MS/MS fragmenting those peptides vary in mass from 114 to 117 Da, which can be used as signature to quantify the protein levels in each sample state. In contrast to most other labeling methods which compare only two samples, iTRAQ allows the multiplex analysis up to four samples, and more recently eight samples [54], at the same time, thus significantly increasing the throughput of quantitative proteomics. The limitation of this approach is that the information on the protein level (e.g. PTMs) can not be obtained as the labeling takes place after enzymatic digestion of samples.

## 2. MASS SPECTROMETRY IN PROTEOMICS

Mass spectrometry has proved to be a powerful analytical tool for the characterization of large biomolecules in proteomics. The principle of mass spectrometry involves the ionization of analyte to generate charged molecules and the measurement of their mass-to-charge ratios (*m/z*). A mass spectrometer is typically composed of three components: ion source, mass analyzer and

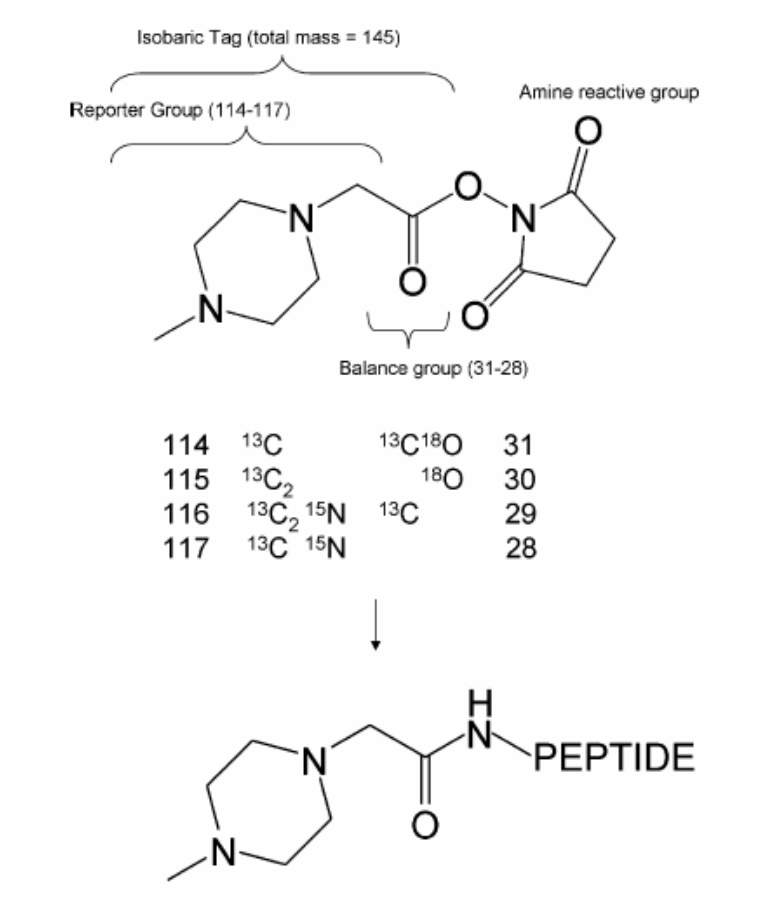



**Figure 1.3** Structure of the iTRAQ reagent. Adapted from Honda *et al.* [32]

detector. In the following sections, two widely used ionization methods and several types of mass spectrometers will be described in details.

### 2.1 Electrospray Ionization (ESI)

Electrospray ionization (ESI) is a continuous "soft" ionization method developed in 1989 by Fenn and coworkers for analysis of large biomolecules using mass spectrometry [2]. The illustration of ESI process is shown in Figure 1.4. In this method, the analyte is first dissolved in the solvent containing water and volatile organic compounds (e.g. methanol, acetonitrile). Then the analyte solution flows through a capillary tube to which a high voltage (1-6 kV) is applied. The electric field generated between the capillary and a counter electrode causes the solution to form a fine mist of charged droplets at the tip of the capillary. Either dry gas, heat, or both are applied to the droplets at atmospheric pressure in order to evaporate the solvent from the droplets. During evaporation, the charge density on the droplet increases as its size decreases. Once the Coulombic repulsion exceeds the forces of surface tension, the droplets break to release the ions by forming a "Taylor cone" [55]. The ions formed in the ESI source are then directed into a mass spectrometer through a series of electrostatic or radio-frequency (RF) ion lenses. Since analyte is introduced in solution, ESI can be easily coupled to liquid chromatography (LC) or capillary electrophoresis (CE) for enhanced MS analysis. The ions generated by ESI are generally multiply charged, allowing for analysis of high mass ions with a relatively low *m/z* range instrument. In addition, multiple charging benefits the tandem MS analysis.

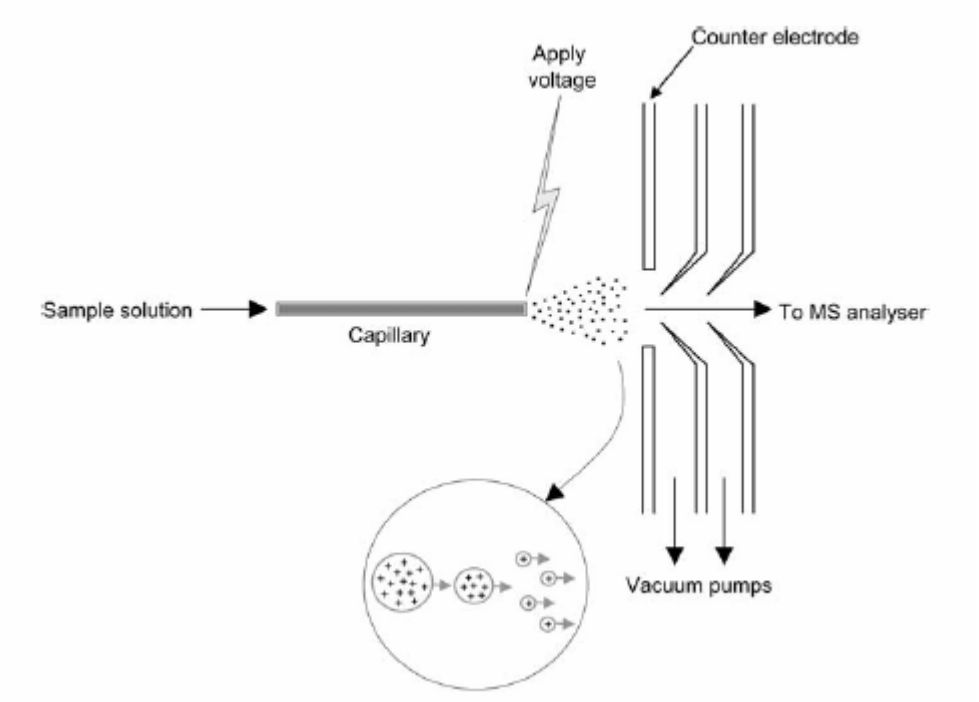



**Figure 1.4** Schematic diagram of an ESI interface and the ESI process. Adapted from Lane *et al.* [56]

## 2.2 Matrix-Assisted Laser Desorption/Ionization (MALDI)

Matrix-assisted laser desorption/ionization (MALDI) is a pulsed “soft” ionization method developed in 1988 by Tanaka, Karas and Hillenkamp [3, 4]. In MALDI, the analyte is first co-crystallized with a matrix compound, and then the analyte-matrix mixture is exposed to a pulsed laser. During laser irradiation, the matrix vaporizes and carries the analyte molecules from the condensed phase into the gas phase (Figure 1.5). Once desorped, the charged analyte molecules are electrostatically transferred from the MALDI source into the mass spectrometer. Generally, the matrix is a UV-absorbing weak organic acid, thus it can protect the analyte by preventing it from direct laser irradiation. The commonly used matrices in MALDI are 2,5-dihydroxybenzoic acid **(**DHB), 3,5-dimethoxy-4-hydroxycinnamic acid **(**sinapinic acid), and α–cyano-4-hydroxycinnamic acid (CHCA). While the exact mechanism for MALDI is not known, it is generally believed that ionization occurs through proton transfer from the protonated matrix to the analyte molecules during the desorption process [57]. The advantages of MALDI over ESI are better tolerance of salts and relatively simple mass spectra to interpret as MALDI primarily produces singly charged ions. The limitation of MALDI is that matrix background in low mass range makes it difficult to analyze the compounds below 700 Da.

## 2.3 Mass Spectrometers

Commonly used mass analyzers include the quadrupole [58], quadrupole ion trap (QIT) [59, 60], linear ion trap (LIT) [61, 62], time-of-flight (TOF) [63-65], Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) [66-68], and orbitrap [69, 70].

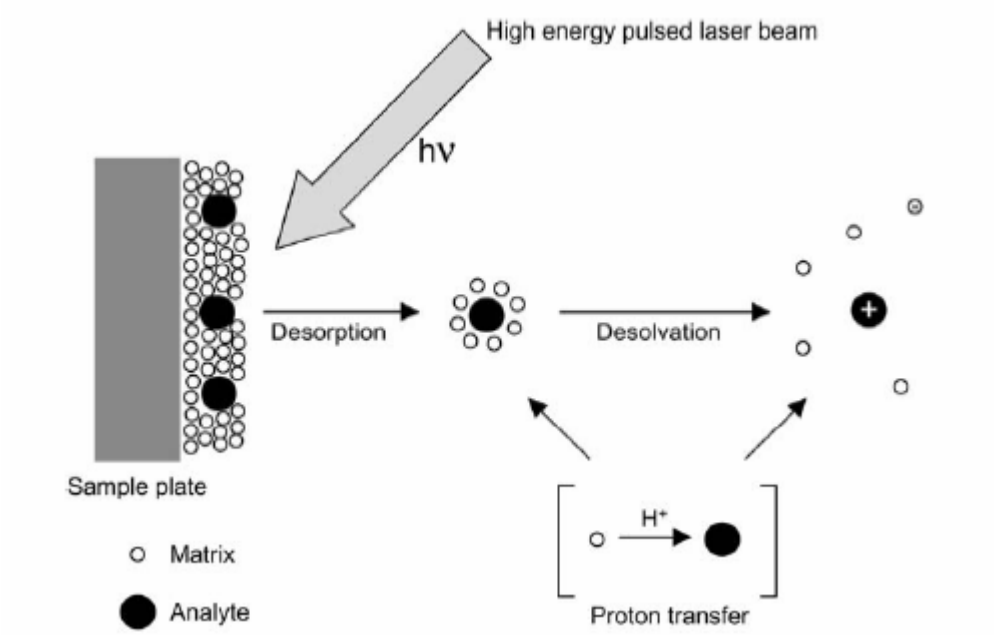



**Figure 1.5** Schematic diagram of the MALDI process. Adapted from Hoffmann *et al.* [57]

### 2.3.1 Quadrupole Mass Spectrometer

The quadrupole mass analyzer [58] has been used with ESI sources since the 1950's and is still one of the most commonly used mass analyzers. The quadrupole consists of an assembly of four precisely-mounted parallel rods equally spaced around a central axis, to which radio-frequency (RF) and direct current (DC) voltages are applied. It resolves the mass-to-charge ratio (*m/z*) based on the motion of ions in modulated electric fields. In the x-y plane, which is perpendicular to the rods, ions undergo transverse oscillations at frequencies that depend on their *m/z* values. Along the z-axis parallel to the center of the assembly, ions move at a constant velocity, as there is no acceleration along this axis. With properly selected RF and DC voltages, the ions with a certain *m/z* range pass along the z-axis and reach the detector positioned at the end of the assembly; the ions outside of this *m/z* range are filtered out by being neutralized when striking the rods or the walls of the device. Quadrupoles may have a mass range up to *m/*z 4,000. Their advantages include the compact size, ease of operation, fast scan speed and low construction costs; and their disadvantage is a relatively low mass resolution.

### 2.3.2 Quadrupole Ion Trap Mass Spectrometer

The quadrupole ion trap (QIT) [59, 60] mass analyzer operates by first storing ions in a three-dimensional quadrupole RF field and then detecting them by mass selective ejection into a detector. This device consists of a doughnut-shaped ring electrode and two hyperbolic end-capped electrodes. Ions created in external ion sources are first injected and trapped. Trapped ions oscillate in the electric filed generated by an RF voltage applied to the ring electrode with the end-capped electrodes grounded. To stabilize the motion of the ions, helium gas is introduced into the ion trap typically at the pressure of ~1 mTorr to dampen the kinetic energy of ions,

focusing the ions toward the center of the trap. To acquire a mass spectrum, the RF voltage on the ring electrode is ramped so that the ions are sequentially ejected from the ion trap into the detector one *m/z* value at a time. The primary advantage of the QIT mass spectrometer is that multiple stages of ion dissociation ($MS^n$) can be performed within a single analyzer. Because of its ability to trap and accumulate ions, the QIT mass spectrometer exhibits better sensitivity than quadrupoles. The disadvantage of QIT mass analyzer is its relatively low mass accuracy and low mass resolution.

### 2.3.3 Linear Ion Trap Mass Spectrometer

The linear ion trap (LIT) mass spectrometer [61, 62] operates by the same mechanism as the QIT. However, ions are trapped by a combination of a two-dimensional quadrupole RF field and an axial static (DC) electric field. The LIT has a much larger entrance aperture and a larger storage volume, which gives the LIT a higher ion injection efficiency and greater ion storage capability than the QIT. The linear ion trap is composed of set of quadrupole rods and two end electrodes which are placed at either end of the rods. Ions are radially trapped by applying RF voltage to quadrupole and a trapping DC voltage to the end electrodes in the presence of helium gas. Ions can be ejected either axially (along the z-axis), by modulating the DC voltage on the end electrodes, or radially (in the x-y plane), by ramping the RF voltage on quadrupole to increase the magnitude of the ion oscillations. Compared to the QIT, the LIT mass spectrometer has improved sensitivity, resolution and mass accuracy. This device can be used stand alone or in combination with other mass analyzers for tandem mass analysis.

### 2.3.4 Time-of-Flight Mass Spectrometer

Time-of-flight (TOF) [63-65] mass spectrometers are among the simplest of mass analyzers. Mass-to-charge ratios (*m/z*) are determined by measuring the time that ions take to travel through a field-free region between the source and the detector. The flight time of ions depends on their *m/z* values, the number of charges, and the acceleration potential. If all ions are accelerated with the same potential, lighter ions will reach the detector earlier than heavier ones of the same charge state. Two techniques have been developed for improving the mass resolution of the TOF mass analyzer. The first technique is referred to as "delayed extraction" [71], which compensates for the temporal and spatial distributions of the ions by introducing a time delay (nanoseconds to microseconds) between ion formation and extraction of ions from the source. The second improvement is to minimize the effect of kinetic energy distribution of the ions by the use of ion mirrors or reflectrons [72] at the end of the flight tube, where more energetic ions penetrate the reflectron deeper before being deflected and arrive at the detector as the same time as less energetic ions of the same *m/z* value. The TOF mass analyzer has virtually unlimited mass range, while the TOF reflectron has mass range up to *m/z* ~10,000 and typically achieves the resolution of 10,000 or greater.

### 2.3.5 Fourier Transform Ion Cyclotron Resonance Mass Spectrometer

Fourier transform ion cyclotron resonance mass spectrometry (FTICR/MS), also known as Fourier transform mass spectrometry (FTMS), is a magnetic field-based ion trap [66-68]. This basis of operation is ion cyclotron motion in a plane perpendicular to a superimposed uniform magnetic field at a frequency dependant on an ion's *m/z* value. The basic principles of ion cyclotron motion were described in the 1930s by Lawrence *et al.* [73], and it was first

incorporated into a mass spectrometer, known as omegatron, in the 1950s by Sommer *et al.* [74]. The first FTMS instrument was built after Comisarow and Marshall adapted Fourier transform methods to ICR spectrometry in 1974 [66], but the FTMS instruments were not widely used until several advances have been made, which include the introduction of external ion sources, such as MALDI and ESI, the development of robust pumping systems that can achieve and maintain ultra-high vacuum, and the availability of the computers that have adequate data storage capacity and processing speed. The main advantage of the FTMS instruments over the other mass spectrometers is the unprecedented mass accuracy, resolution, and sensitivity. Despite the relatively expensive costs and the requirement of expert skills for operation, the FTMS instruments have been proved a powerful platform for proteomic studies.

#### 2.3.5.1 Basic Components

Although several types of FTMS instruments have been built over past years for various application purposes, all FTMS instruments have in common four main components: a magnet, an analyzer cell, an ultra-high vacuum system and a sophisticated data system [75]. The magnet is an essential part for the FTMS instruments, as the mass measurement is accomplished based on the ion motion in a uniform magnetic field. Most of modern FTMS instruments employ a superconducting magnet with the strength of magnetic field ranging from 3 to 12 tesla. The analyzer cell is the central component of the FTMS instruments because it is the place where the ions are trapped, mass analyzed and detected. The analyzer cell resides in the homogeneous region of the magnet. It can take on different geometries, such as a cube and an open-ended cylinder, but generally consists of three pairs of electrodes: a front and back trapping electrode, two opposite excitation electrodes, and two opposite detection electrodes (Figure 1.6). In order to

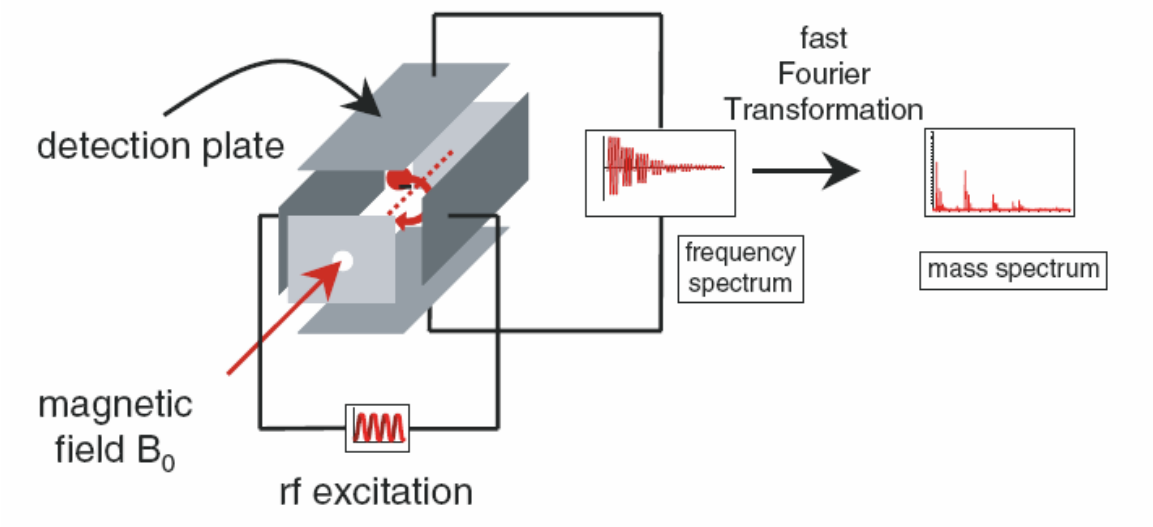



**Figure 1.6** Schematic diagram of a cubic ICR analyzer cell, consisting of two opposite excitation plates (right and left), two detection plates (top and bottom) and two trapping plates (front and back). The direction of the magnetic field ($B_0$) is also indicated. Adapted from Schrader *et al.* [76]

obtain ultra-high resolution, the FTMS instruments are required to operate in an ultra-high vacuum system ($10^{-9}$ to $10^{-10}$ torr) during ion detection which can be achieved by using cryogenic pumps or turbomolecular pumps. A sophisticated data system is also needed to enable coordinating all the electronic devices during the acquisition of data and to process and analyze the data.

#### 2.3.5.2 Cyclotron Ion Motion

In FTMS, the ions entering the analyzer cell undergo cyclotron motion in the presence of homogenous magnetic field. Figure 1.7 illustrates the principle of ion cyclotron motion, which basically results from the balance between the Lorentz force $F_1$ and the centrifugal force $F_2$ operating on an ion in opposite directions. Since the magnitude of the Lorentz force $F_1$ is equal to that of the centrifugal force $F_2$ (i.e. $F_1 = F_2$), the ion cyclotron frequency can be calculated from the equation below:

$$f_c = \frac{q\mathrm{B}}{2\pi m} \qquad (1)$$

where $f_c$ is the unperturbed cyclotron frequency, $B$ the strength of the magnetic field, $q$ the charge of the ion, and $m$ the mass of the ion. As the magnetic field is spatially uniform in the analyzer cell and only drifts a few ppm per year, the factor B in eq 1 keeps constant, thus the $m/q$ ratio of an ion can be accurately determined from its cyclotron frequency. Unlike the most other mass spectrometers, where the kinetic energy has a profound effect on the resolution, cyclotron frequency of an ion is independent of its kinetic energy, thus enabling the FTMS instruments to produce an extraordinary high resolution. It is also noticed that the radius of the ion's cyclotron orbit is proportional to the square root of its kinetic energy. Therefore, increasing the kinetic energies of the ions can excite them to a larger cyclotron orbit.

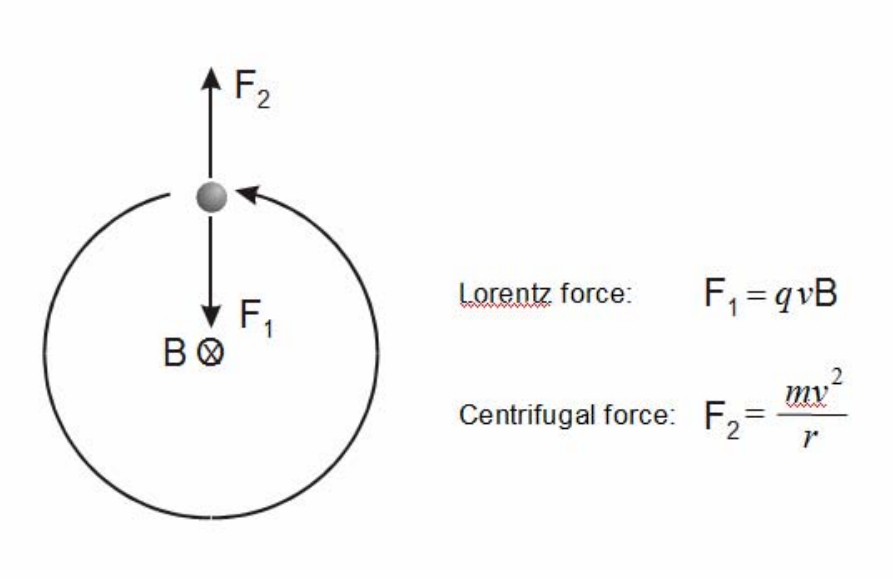



**Figure 1.7** Cyclotron motion results from the balance between the Lorentz force $F_1$ and the centrifugal force $F_2$. In the equations, $m$ is the mass of the ion, $v$ the velocity of the ion, $r$ the radius of cyclotron motion, $q$ the charge of the ion, and $B$ the strength of magnetic field.

#### 2.3.5.3 Ion Excitation and Detection

During FTMS analysis, a series of events called experimental sequence takes place in the order of time. A simple FTMS experimental sequence consists of ion formation, ion excitation and ion detection [75]. The ions are first created either in the external ion sources, such as MALDI or ESI, and then sent down to the cell through a series of electrostatic or RF ion lenses, or inside the cell by electron-impact ionization (EI). When the ions are trapped in the analyzer cell, they often undergo cyclotron motion at a very small radius. In order to detect them, a RF pulse is applied on the excitation electrodes to excite the ions to a larger cyclotron radius. The ions can be excited only when the ion's cyclotron frequency is in resonance with the frequency of the applied RF electric field. All ions with same *m/z* values are excited and move coherently. When they move close to the cell electrodes, they can induce an oscillating differential image current in two opposite detection electrodes, which can be amplified, digitized and stored for further processing. When all ions with different *m/z* values are simultaneously excited, the image current signal is a composite of different frequencies. The frequency components of the signal can be obtained by applying fast Fourier transform (FFT) to the time domain signal. The frequency spectrum is then converted into a mass spectrum by applying a calibration formula derived from eq 1. Unlike all other mass spectrometers, where the ions are detected by destructive collisions with an electron multiplier, the image current detection employed in FTMS is non-destructive. The maximum mass resolution that can be obtained by FTMS is proportional to the length of the image current signal, or transient, as shown in eq 2,

$$R = \frac{f_c T}{2} \qquad (2)$$

where R is mass resolution, $f_c$ is the cyclotron frequency, and T is the length of a transient. Thus, a longer transient should be recorded in order to achieve a high mass resolution.

#### 2.3.5.4 Modes of Excitation

Ions can be excited using many types of excitation waveforms. To ensure excitation of all ions with different *m/z* values (broadband detection), an excitation waveform comprising multiple frequencies is often employed. The most commonly used waveform for broadband detection is a rapid frequency sweep called "chirp", which provides the excitation with a relatively flat magnitude over a broad frequency range (from several kilohertz to megahertz) in a short period of time (1-2 ms) (Figure 1.8c). However, the chirp excitation has the disadvantages of non-uniform excitation amplitude across the spectrum and limited mass selectivity at the start and end frequencies of the sweep. To address these issues, a method that produces an optimal excitation waveform was proposed by Marshall *et al.* in the mid 1980s [77]. This method, called stored waveform inverse Fourier transform (SWIFT), is accomplished by specifying the desired excitation profile in the mass domain, converting it to a frequency domain excitation spectrum, and performing an inverse Fourier transform to produce the time domain excitation waveform (Figure 1.8d).

### 2.3.6 Orbitrap Mass Spectrometer

Orbitrap mass spectrometry is a new type of mass analysis recently invented by Alexander Makarov [69] and is a modification of the Kingdon trap [78]. As its name suggests, the orbitrap is an ion trap, but unlike the FTMS or QIT, it uses a static electric field. This device consists of an outer barrel-shaped electrode and a coaxial inner spindle-shaped electrode that form a trapping electrostatic field. Ions generated from external ion sources are accumulated and cooled in a RF-only curved quadrupole called a "C" trap and then tangentially injected into the orbitrap. Ions can be trapped in the orbitrap because their electrostatic attraction to the inner electrode is

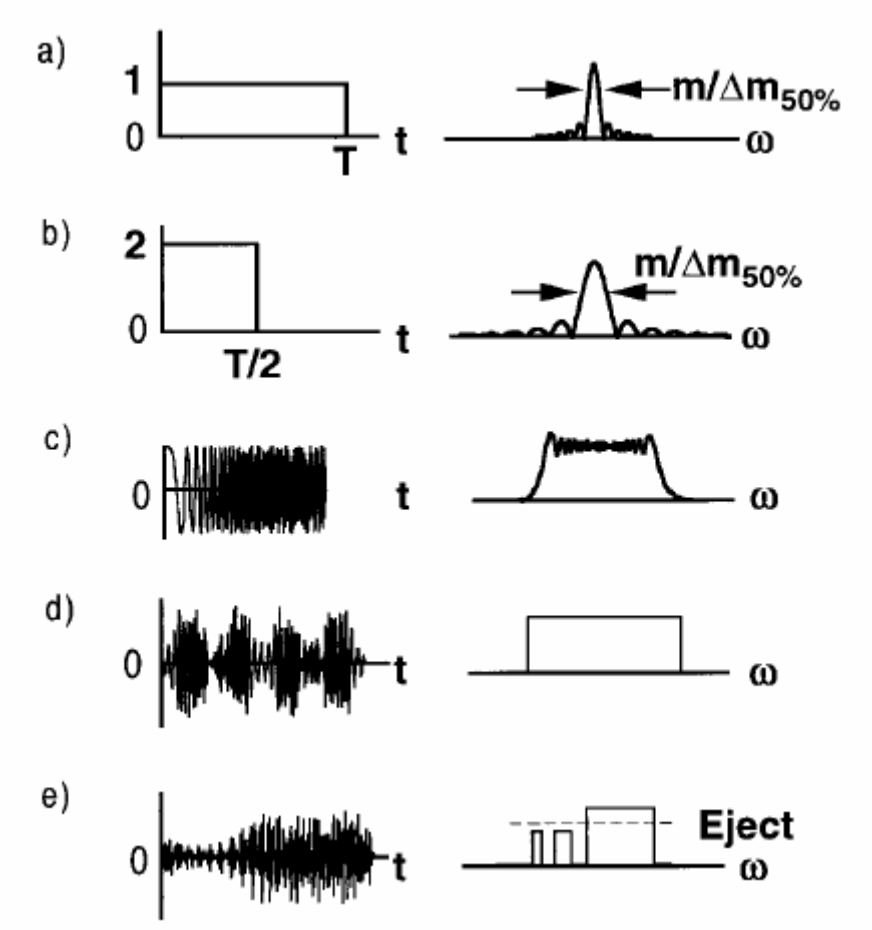



**Figure 1.8** Time-domain (left) and frequency-domain (right) excitation waveforms. (a), (b) Rectangular pulses. (c) Frequency-sweep (‘‘chirp’’). (d), (e) Stored waveform inverse Fourier transform (‘‘SWIFT’’) waveforms. Adapted from Marshall *et al.* [67]

compensated by centrifugal forces as ions rotate around the central electrode in rings. In addition, the ions also move back and forth along the axis of the central electrode. The axial frequency of harmonic oscillations is independent of the energy and position of the ions, and is inversely proportional to the square root of the *m/z* ratio. Similar to the FTMS instruments, the image current is induced and recorded on split outer electrodes in the time domain and then converted into *m/z* ratio by fast Fourier transformation (FFT). Orbitrap provides high mass accuracy and resolution comparable to FTMS, but it is more cost-effective as no magnet is needed. Since the first commercial instrument (LTQ-Orbitrap) was released in 2005, orbitrap has been gaining increasing attention in the proteomic field.

### 2.4 Tandem Mass Spectrometry

Tandem mass spectrometry, also known as MS/MS or $MS^n$, has been extensively used in proteomic studies for protein identification, peptide sequencing and characterization of post-translational modifications (PTMs). This technique involves the multiple steps of MS analysis which can be performed either with individual mass spectrometer elements separated “in space” or in a single mass spectrometer with the MS steps separated “in time” [79, 80]. For “in space” configurations such as TOF-TOF and LIT-TOF, the ions of interest (precursor ions) that are isolated from the first mass spectrometer undergo fragmentation in a collision cell which is placed in between the two mass spectrometers, and the fragment ions are detected by the second mass spectrometer. For “in-time” configurations such as FTMS and LIT, mass selection of precursor ions, fragmentation of precursor ions, and detection of fragment ions occur consecutively within the same mass spectrometer. There are many methods used for fragmentation, such as collision-induced dissociation (CID) [81-83], infrared multiphoton

dissociation (IRMPD) [83, 84], electron capture dissociation (ECD) [85, 86] and electron transfer dissociation (ETD) [87, 88]. Different types of methods can result in different fragmentation pattern and thus different information about the structure and composition of the analyte. For proteins and peptides, there are three points of possible cleavage along the amide backbone, resulting in six types of fragment ions referred to as a, b, and c series, when the N-terminus retains the charge, and x, y, and z series, when the C-terminus retains the charge [89].

Collision-induced dissociation (CID) [81-83] is the most commonly used fragmentation method in tandem mass spectrometry. This method is based on the collisions of the precursor ions with a neutral gas (e.g. helium or argon). CID fragments proteins and peptides depending upon the collision energy, amino acid sequence, and the number of charges, and it generally produces b and y type ions. Infrared multiphoton dissociation (IRMPD) [83, 84] is a fragmentation method based on shooting the precursor ions with a relatively high-powered IR laser (typically a $CO_2$ laser) through the center of the mass analyzer. For peptides, IRMPD yields similar a fragmentation pattern to CID, but provides more extensive cleavage than is obtained with CID. Both CID and IRMPD tend to break the weak bonds in proteins or peptides, thus the labile groups including some PTMs will be lost during fragmentation. To address this problem, two fragmentation methods, electron capture dissociation (ECD) [85, 86] and electron transfer dissociation (ETD) [87, 88], were recently developed. ECD involves the capture of low-energy electrons (<1 ev) by the multiply charged precursor ions. This method generally produces c and z type ions, thus can be used as a complementary method to CID to enhance peptide sequencing. Electron-transfer dissociation (ETD) is a technique similar to ECD that employs the gas-phase ion-ion chemistry. In ETD, fragmentation occurs when electrons are transferred from the single-charged anions (e.g. anthracene) to the multiply charged precursor ions, yielding the fragment

ions that are analogous to those of ECD. Both ECD and ETD fragment proteins or peptides more randomly along the backbone, thus the PTMs can be preserved during the fragmentation, allowing for site-specific PTM analysis using these two techniques. In contrast to ECD, which is primarily used with FTICR mass spectrometer, ETD can be implemented on a stand alone ion trap mass spectrometer, providing a high promise for its more widespread use in proteomics.

## 3. SCOPE OF DISSERTATION

This dissertation describes three approaches for facilitating shotgun proteomic analysis by accurate mass measurement using MALDI-FTICR mass spectrometry. The first approach described in Chapter 2 is based on mass defect labeling of tryptophan residues in proteins with a reagent that is MALDI compatible. Chapter 3 reports a method that employs the SWIFT excitation to improve the mass accuracy of higher mass peptides. Chapter 4 describes a HPLC-MALDI-FTICR/MS approach for shotgun proteomic analysis of a $^{15}$N-metabolically labeled proteome sample obtained from *Methanococcus maripaludis*. In Chapter 5, an approach that combines on-target digestion with MALDI-FTICR/MS analysis for rapid protein identification is described. In the final chapter of this dissertation, conclusions and future work are addressed.

## REFERENCES

1 Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; HumpherySmith, I.; Williams, K. L.; Hochstrasser, D. F. From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. *Bio-Technology* **1996,** *14*, 61-65.

2 Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **1989,** *246*, 64-71.

3 Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, 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.

4 Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Anal. Chem.* **1988,** *60*, 2299-2301.

5 Peng, J. M.; Gygi, S. P. Proteomics: The Move to Mixtures. *Journal of Mass Spectrometry* **2001,** *36*, 1083-1091.

6 Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of Two-Dimensional Gel Electrophoresis-Based Proteome Analysis Technology. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97*, 9390-9395.

7 Yates, J. R. Mass Spectrometry - from Genomics to Proteomics. *Trends in Genetics* **2000,** *16*, 5-8.

8 Lane, C. S. Mass Spectrometry-Based Proteomics in the Life Sciences. *Cmls-Cellular and Molecular Life Sciences* **2005,** *62*, 848-869.

9 Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003,** *422*, 198-207.

10 Bogdanov, B.; Smith, R. D. Proteomics by FTICR Mass Spectrometry: Top Down and Bottom Up. *Mass Spectrom. Rev.* **2005,** *24*, 168-200.

11 Wehr, T. Top-Down versus Bottom-up Approaches in Proteomics. *Lc Gc North America* **2006,** *24*, 1004.

12 Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. Rapid Identification of Proteins by Peptide-Mass Fingerprinting. *Current Biology* **1993,** *3*, 327-332.

13 Mann, M.; Hojrup, P.; Roepstorff, P. Use of Mass-Spectrometric Molecular-Weight Information to Identify Proteins in Sequence Databases. *Biol. Mass Spectrom.* **1993,** *22*, 338-345.

14 Yates, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. Peptide Mass Maps - a Highly Informative Approach to Protein Identification. *Anal. Biochem.* **1993,** *214*, 397-408.

15 Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001,** *73*, 5683-5690.

16 Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001,** *19*, 242-247.

17 Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P. Multidimensional Separations-Based Shotgun Proteomics. *Chem. Rev.* **2007,** *107*, 3654-3686.

18 Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y. F.; 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.

19 Qian, W. J.; Camp, D. G.; Smith, R. D. High-Throughput Proteomics Using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Expert Review of Proteomics* **2004,** *1*, 87-95.

20 Jacobs, J. M.; Monroe, M. E.; Qin, W. J.; Shen, Y. F.; Anderson, G. A.; Smith, R. D. Ultra-Sensitive, High Throughput and Quantitative Proteomics Measurements. *Int. J. Mass spectrom.* **2005,** *240*, 195-212.

21 Zimmer, J. S. D.; Monroe, M. E.; Qian, W. J.; Smith, R. D. Advances in Proteomics Data Analysis and Display Using an Accurate Mass and Time Tag Approach. *Mass Spectrom. Rev.* **2006,** *25*, 450-482.

22 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. *Anal. Chem.* **2006,** *78*, 3417-3423.

23 Creighton, T. E. Proteins: Structures and Molecular Properties. 2nd ed. *W. H. Freeman and Company*: New York, **1993**.

24 Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top-Down Mass Spectrometry of a 29-KDa Protein for Characterization of Any Posttranslational Modification to within One Residue. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99*, 1774-1779.

25 Kelleher, N. L. Top-Down Proteomics. *Anal. Chem.* **2004,** *76*, 196A-203A.

26 Han, X. M.; Jin, M.; Breuker, K.; McLafferty, F. W. Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science* **2006,** *314*, 109-112.

27 Wang, W. X.; Zhou, H. H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards. *Anal. Chem.* **2003,** *75*, 4818-4826.

28 Liu, H. B.; Sadygov, R. G.; Yates, J. R. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **2004,** *76*, 4193-4201.

29 Tao, W. A.; Aebersold, R. Advances in Quantitative Proteomics via Stable Isotope Tagging and Mass Spectrometry. *Curr. Opin. Biotechnol.* **2003,** *14*, 110-118.

30 Leitner, A.; Lindner, W. Current Chemical Tagging Strategies for Proteome Analysis by Mass Spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2004,** *813*, 1-26.

31 Prange, A.; Proefrock, D. Chemical Labels and Natural Element Tags for the Quantitative Analysis of Bio-Molecules. *J. Anal. At. Spectrom.* **2008,** *23*, 432-459.

32 Honda, A.; Suzuki, Y.; Suzuki, K. Review of Molecular Modification Techniques for Improved Detection of Biomolecules by Mass Spectrometry. *Anal. Chim. Acta* **2008,** *623*, 1-10.

33 Gevaert, K.; Impens, F.; Ghesquiere, B.; Van Damme, P.; Lambrechts, A.; Vandekerckhove, J. Stable Isotopic Labeling in Proteomics. *Proteomics* **2008,** *8*, 4873-4885.

34 Iliuk, A.; Galan, J.; Tao, W. A. Playing Tag with Quantitative Proteomics. *Analytical and Bioanalytical Chemistry* **2009,** *393*, 503-513.

35 Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Pasa-Tolic, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. Quantitative Analysis of Bacterial and Mammalian Proteomes Using a Combination of Cysteine Affinity Tags and 15N-Metabolic Labeling. *Anal. Chem.* **2001,** *73*, 2132-2139.

36 Snijders, A. P. L.; de Vos, M. G. J.; Wright, P. C. Novel Approach for Peptide Quantitation and Sequencing Based on N-15 and C-13 Metabolic Labeling. *J. Proteome Res.* **2005,** *4*, 578-585.

37 Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, Silac, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* **2002,** *1*, 376-386.

38 Mann, M. Functional and Quantitative Proteomics Using Silac. *Nature Reviews Molecular Cell Biology* **2006,** *7*, 952-958.

39 Gu, S.; Pan, S. Q.; Bradbury, E. M.; Chen, X. A. Precise Peptide Sequencing and Protein Quantification in the Human Proteome through in Vivo Lysine-Specific Mass Tagging. *J. Am. Soc. Mass Spectrom.* **2003,** *14*, 1-7.

40 Ong, S. E.; Mittler, G.; Mann, M. Identifying and Quantifying in Vivo Methylation Sites by Heavy Methyl Silac. *Nature Methods* **2004,** *1*, 119-126.

41 Ong, S. E.; Kratchmarova, I.; Mann, M. Properties of C-13-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (Silac). *Journal of Proteome Research* **2003,** *2*, 173-181.

42 Miyagi, M.; Rao, K. C. S. Proteolytic O-18-Labeling Strategies for Quantitative Proteomics. *Mass Spectrom. Rev.* **2007,** *26*, 121-136.

43 Fenselau, C.; Yao, X. D. O-18(2)-Labeling in Quantitative Proteomic Strategies: A Status Report. *Journal of Proteome Research* **2009,** *8*, 2140-2143.

44 Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. Protease-Catalyzed Incorporation of O-18 into Peptide Fragments and Its Application for Protein Sequencing by Electrospray and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Electrophoresis* **1996,** *17*, 945-953.

45 Yao, X. D.; Afonso, C.; Fenselau, C. Dissection of Proteolytic O-18 Labeling: Endoprotease-Catalyzed O-16-to-O-18 Exchange of Truncated Peptide Substrates. *Journal of Proteome Research* **2003,** *2*, 147-152.

46 Reynolds, K. J.; Yao, X. D.; Fenselau, C. Proteolytic O-18 Labeling for Comparative Proteomics: Evaluation of Endoprotease Glu-C as the Catalytic Agent. *Journal of Proteome Research* **2002,** *1*, 27-33.

47 Angel, P. M.; Orlando, R. Trypsin Is the Primary Mechanism by Which the O-18 Isotopic Label is Lost in Quantitative Proteomic Studies. *Anal. Biochem.* **2006,** *359*, 26-34.

48 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.

49 Zhang, R. J.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. Controlling Deuterium Isotope Effects in Comparative Proteomics. *Anal. Chem.* **2002,** *74*, 3662-3669.

50 Zhang, R. J.; Sioma, C. S.; Wang, S. H.; Regnier, F. E. Fractionation of Isotopically Labeled Peptides in Quantitative Proteomics. *Anal. Chem.* **2001,** *73*, 5142-5149.

51 Li, J. X.; Steen, H.; Gygi, S. P. Protein Profiling with Cleavable Isotope-Coded Affinity Tag (cICAT) Reagents - the Yeast Salinity Stress Response. *Molecular & Cellular Proteomics* **2003,** *2*, 1198-1204.

52 Gygi, S. P.; Rist, B.; Aebersold, R. Measuring Gene Expression by Quantitative Proteome Analysis. *Curr. Opin. Biotechnol.* **2000,** *11*, 396-401.

53 Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-Reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics* **2004,** *3*, 1154-1169.

54 Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H. 8-Plex Quantitation of Changes in Cerebrospinal Fluid Protein Expression in Subjects Undergoing Intravenous Immunoglobulin Treatment for Alzheimer's Disease. *Proteomics* **2007,** *7*, 3651-3660.

55 de la Mora, J. F.; Van Berkel, G. J.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. Electrochemical Processes in Electrospray Ionization Mass Spectrometry - Discussion. *J. Mass Spectrom.* **2000,** *35*, 939-952.

56 Lane, C. S. Mass Spectrometry-Based Proteomics in the Life Sciences. *Cellular and Molecular Life Sciences* **2005,** *62*, 848-869.

57 Hoffmann, E.; Stroobant, V. Mass Spectrometry: Principles and Applications. ed. *Wiley*: Chichester, **2003**.

58 Yost, R. A.; Enke, C. G. Selected Ion Fragmentation with a Tandem Quadrupole Mass-Spectrometer. *J. Am. Chem. Soc.* **1978,** *100*, 2274-2275.

59 Stafford, G. C.; Kelley, P. E.; Syka, J. E. P.; Reynolds, W. E.; Todd, J. F. J. Recent Improvements in and Analytical Applications of Advanced Ion Trap Technology. *Int. J. Mass Spectrom. Ion Processes* **1984,** *60*, 85-98.

60 Todd, J. F. J.; March, R. E. A Retrospective Review of the Development and Application of the Quadrupole Ion Trap Prior to the Appearance of Commercial Instruments. *Int. J. Mass spectrom.* **1999,** *191*, 9-35.

61 Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2002,** *13*, 659-669.

62 Douglas, D. J.; Frank, A. J.; Mao, D. M. Linear Ion Traps in Mass Spectrometry. *Mass Spectrom. Rev.* **2005,** *24*, 1-29.

63 Mamyrin, B. A. Time-of-Flight Mass Spectrometry (Concepts, Achievements, and Prospects). *Int. J. Mass spectrom.* **2001,** *206*, 251-266.

64 Grotemeyer, J. Special Issue: Time-of-Flight Mass Spectrometry - Foreword. *Int. J. Mass spectrom.* **2001,** *206*, 179-179.

65 Mirsaleh-Kohan, N.; Robertson, W. D.; Compton, R. N. Electron Ionization Time-of-Flight Mass Spectrometry: Historical Review and Current Applications. *Mass Spectrom. Rev.* **2008,** *27*, 237-285.

66 Comisaro, M. B.; Marshall, A. G. Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974,** *25*, 282-283.

67 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.

68 Marshall, A. G.; Hendrickson, C. L.; Ernmetta, M. R.; Rodgers, R. P.; Blakney, G. T.; Nilsson, C. L. Fourier Transform Ion Cyclotron Resonance: State of the Art. *European Journal of Mass Spectrometry* **2007,** *13*, 57-59.

69 Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000,** *72*, 1156-1162.

70 Perry, R. H.; Cooks, R. G.; Noll, R. J. Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications. *Mass Spectrom. Rev.* **2008,** *27*, 661-699.

71 Brown, R. S.; Lennon, J. J. Mass Resolution Improvement by Incorporation of Pulsed Ion Extraction in a Matrix-Assisted Laser-Desorption Ionization Linear Time-of-Flight Mass-Spectrometer. *Anal. Chem.* **1995,** *67*, 1998-2003.

72 Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A. Mass-Reflectron a New Nonmagnetic Time-of-Flight High-Resolution Mass-Spectrometer. *Zhurnal Eksperimentalnoi I Teoreticheskoi Fiziki* **1973,** *64*, 82-89.

73 Lawrence, E. O.; Edlefsen, N. E. On the Production of High Speed Protons. *Science* **1930,** *72*, 376-377.

74 Hipple, J. A.; Sommer, H.; Thomas, H. A. A Precise Method of Determining the Faraday by Magnetic Resonance. *Physical Review* **1949,** *76*, 1877-1878.

75 Amster, I. J. Fourier Transform Mass Spectrometry. *Journal of Mass Spectrometry* **1996,** *31*, 1325-1337.

76 Schrader, W.; Klein, H. W. Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (LC-FTICR MS): An Early Overview. *Analytical and Bioanalytical Chemistry* **2004,** *379*, 1013-1024.

77 Marshall, A. G.; Wang, T. C. L.; Ricca, T. L. Tailored Excitation for Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry. *J. Am. Chem. Soc.* **1985,** *107*, 7893-7897.

78 Kingdon, K. H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.* **1923,** *21*, 408-418.

79 Hernandez, P.; Muller, M.; Appel, R. D. Automated Protein Identification by Tandem Mass Spectrometry: Issues and Strategies. *Mass Spectrom. Rev.* **2006,** *25*, 235-254.

80 Feng, X. J.; Liu, X.; Luo, Q. M.; Liu, B. F. Mass Spectrometry in Systems Biology: An Overview. *Mass Spectrom. Rev.* **2008,** *27*, 635-660.

81 Sleno, L.; Volmer, D. A. Ion Activation Methods for Tandem Mass Spectrometry. *J. Mass Spectrom.* **2004,** *39*, 1091-1112.

82 Wells, J. M.; McLuckey, S. A., Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Biol. Mass Spectrom.* **2005,** 402, 148-185.

83 Laskin, J.; Futrell, J. H. Activation of Large Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2005,** *24*, 135-167.

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

85 Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998,** *120*, 3265-3266.

86 Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation of Gaseous Multiply-Charged Proteins is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity. *J. Am. Chem. Soc.* **1999,** *121*, 2857-2862.

87 Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 9528-9533.

88 Coon, J. J.; Ueberheide, B.; Syka, J. E. P.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. Protein Identification Using Sequential Ion/Ion Reactions and Tandem Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 9463-9468.

89 Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass-Spectra of Peptides. *Biomedical Mass Spectrometry* **1984,** *11*, 601-601.

# CHAPTER 2

## A MATRIX-ASSISTED LASER DESORPTION/IONIZATION COMPATIBLE REAGENT FOR TAGGING TRYPTOPHAN RESIDUES

---

Chunyan Li, Vijay Gawandi, Adam Protos, Robert S. Phillips and I. Jonathan Amster, *Eur. J. Mass Spectrom*. **2006**, 12, 213-221

Reprinted here with the permission of the publisher

**ABSTRACT**

Chemical tagging of amino acids is an important tool in proteomics analysis, and has been used to introduce isotope labels and mass defect labels into proteolytic peptides by derivatization of cysteine or lysine residues. Here we present a new reagent with chemical specificity for tryptophan residues. Previously, 2-nitrobenzenesulfenyl chloride has been used as a highly specific reagent for labeling tryptophan residues. We show that this tag undergoes UV dissociation during MALDI. The multiplicity of photofragments increases the difficulty of characterizing the derivatization products. To overcome this problem, we have synthesized a new reagent, 2-(trifluoromethyl)benzenesulfenyl chloride, which is shown to react quantitatively with tryptophan in peptides and proteins. Most significantly, it exhibits high photostability in MALDI-FTMS analyses.

**INTRODUCTION**

Proteomic analysis has become a key technology for studying biological processes in cells or tissues. Mass spectrometry (MS) is one of the central analytical techniques in proteomics research, and is often facilitated by prior sample preparation steps that can reduce the enormous complexity of the protein mixtures. A number of so-called tagging strategies have been developed which target specific amino acid residues [1, 2]. In addition, the attachment of stable-isotope tags allows the relative quantitation of protein levels of two samples, e.g. those representing different cell states. One of the well-known tagging approaches uses isotope-coded affinity tags (ICAT), a method that was introduced by Gygi *et al*. in 1999 [3]. The ICAT (light/heavy) reagent can be used to modify cysteine-containing proteins that are derived from two different states. More recently, Hernandez *et al*. reported a new method called mass defect

labeling (MDL) which is based on the derivatization of cysteine-containing peptides with a reagent that changes their mass defects [4]. Derivatization of lysine to homoarginine was proposed by several groups to enhance their intensities in the investigation of proteins or peptides by MALDI-TOF-MS [5-8]. The same derivatization procedure was later adapted by Cagney and Emili [9], who termed the approach "mass-coded abundance tagging" (MCAT), where C-terminal lysine residues of tryptic peptides were modified through differential guanidination. The extension of these and other tagging approaches to other amino acids requires the development of new reagents with chemical specificity for the selected amino acid.

Tryptophan is an attractive target for chemical tagging strategies, as it is the least abundant of the commonly occurring amino acids. Tryptophan occurs at a rate of 1.3% in the amino acid composition of a typical protein [10], and less than 10% of tryptic peptides will contain this residue. Thus, proteomic approaches which target tryptophan-containing peptides will exhibit a considerable reduction in sample complexity. Derivatization of tryptophan in proteins has been achieved through photooxidation [11], iodination [12], and ozonization [13]. The use of N-bromosuccinimide represents a convenient spectrophotometric method for determining the tryptophan content in proteins [14], but it is only rarely used today for the quantitative analysis of tryptophan in proteins. Koshland *et al*. have reported the use of 2-hydroxy-5-nitrobenzyl bromide as a selective reagent for tryptophan at acidic pH [15]. However, this reagent is extremely sensitive to hydrolysis and not very soluble under aqueous conditions. Sulfenyl halides have been reported to have specific reactivity with tryptophan and cysteine residues of proteins in acidic media [16-20]. The nitro-substituted benzenesulfenyl chlorides are commonly used for modifying tryptophan residues. Such compounds include 2-nitrobenzenesulfenyl chloride (2-NBSCl), 4-nitrobenzenesulfenyl chloride (4-NBSCl), 2,4-dinitrobenzenesulfenyl

chloride (DNBSCl) and 2-nitro-4-carboxybenzenesulfenyl chloride (NCBSCl). Sulfenyl chlorides react with tryptophan to form a thioether at 2-position of the indole ring, as shown in Figure 2.1 for the reaction of 2-NBSCl.

Previous applications of nitrobenzenesulfenyl chlorides were mainly focused on the spectrophotometric determination of the tryptophan content in proteins. The reaction of tryptophan with nitrobenzenesulfenyl chloride can be easily quantitated by spectrophotometry since the nitrobenzenesulfenyl group absorbs radiation in the visible region of the spectrum. For example, the tryptophan derivative obtained by treating a tryptophan-containing protein with 2-NBSCl absorbs at 365 nm. However, mass spectrometry can also be used to monitor derivatized tryptophan. Kuyama *et al*. have reported a method for relative quantitation of proteins based on mass spectrometric analysis of tryptophan residues that has been labeled with light and heavy 2-nitrobenzenesulfenyl chloride (2-NBSCl-$^{12}C_6$ or 2-NBSCl-$^{13}C_6$) [21]. Both MALDI and electrospray ionization (ESI) mass spectrometry were used to make the quantitative analysis of the labeled tryptic peptides.

For peptides that have been labeled with 2-nitrobenzenesulfenyl chloride (2-NBSCl), we have observed considerable fragmentation of the molecular ion when examined by MALDI Fourier transform mass spectrometry (FTMS). As with most nitroaromatic compounds, this reagent has a strong absorption band at the wavelength of the nitrogen laser, 337 nm. To overcome this limitation, we have synthesized and tested a new reagent, 2-(trifluoromethyl)benzenesulfenyl chloride (2-TFBSCl). Here we report the results of the application of this reagent which has chemical selectivity for tryptophan residues and which is stable to MALDI analysis.

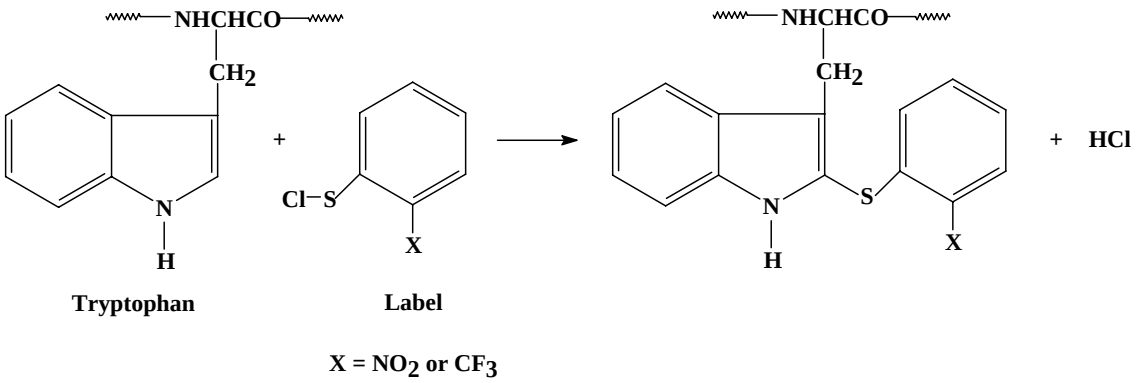



**Figure 2.1** The reaction of 2-X-benzenesulfenyl chloride with tryptophan residue (X = $NO_2$ or $CF_3$).

## MATERIALS AND METHODS

### Chemicals and Reagents

[Tyr$^4$-]-bombesin (M.W. = 1668.7942 Da, p-EQRYGNQWAVGHLM-$NH_2$), myoglobin (from horse skeletal muscle) and ammonium bicarbonate were purchased from Sigma Chemical Co. (St. Louis, MO, USA). 2,5-Dihydroxybenzoic acid (DHB) was obtained from Lancaster Synthesis (Pelham, NH, USA). Sinapinic acid and α-cyano-4-hydroxycinnamic acid (CHCA) were obtained from Acros Organics (Morris Plains, NJ, USA). Sequencing-grade modified trypsin was obtained from Promega (Madison, WI, USA). Sephadex LH-20 was obtained from Amersham Biosciences Corp. (Piscataway, NJ, USA). Guanidine HCl, acetonitrile (HPLC grade) and trifluoroacetic acid (TFA) were obtained from Fisher Scientific (Fair Lawn, NJ, USA). Acetic acid (glacial) was obtained from EM Science (Gibbstown, NJ, USA). Water used in all experiments was purified using a NanopureInfinity ultrapure water system (Barnstead/Thermolyne, Dubuque, IA, USA).

### Syntheses of Tryptophan Labeling Reagents

#### *Synthesis of 2-NBSCl*

*o*-Nitrothiophenol (1.0 mmol) was dissolved in 100 mL of nitromethane at 0 $^oC$. 1.1 equivalents of sulfuryl chloride were added slowly, and a white precipitate formed after 10-15 minutes. The crude o-nitrobenzenesulfenyl chloride was isolated by evaporation of nitromethane under reduced pressure, and the residue was recrystallized from $CCl_4$ to provide 2-nitrobenzene sulfenyl chloride in 77% yield.

***Synthesis of 2-TFBSCl***

2-(Trifluoromethyl)aniline (25 mmol) was dissolved in 15 mL of 6 N HCl in a 50 mL beaker, and cooled to 0-5 $^{o}$C in an ice-salt bath. 2-(Trifluoromethyl)aniline hydrochloride was converted to a diazonium salt by the addition of 26 mmol of $NaNO_2$ in 4 mL of water. Sodium fluoroborate (36 mmol) in 10 mL of water was added to this solution with vigorous stirring for 10 minutes. The solid diazonium fluoroborate salt was collected on a Büchner funnel, and washed rapidly with 10 mL of cold 5% $NaBF_4$ solution, then twice with 10 mL portions of absolute alcohol and twice with ether. The diazonium group was converted to the corresponding thiophenol by using polymer-supported hydrosulfide anion under mild conditions [22]. Thiophenol was also obtained by reaction of the tetrafluoroborate salt with $Na_2S$. The purity of the product was analyzed by NMR and GC-MS (supplementary data).

2-(Trifluoromethyl)benzenethiol (1.0 mmol) was dissolved in 100 mL of nitromethane at 0 $^{o}$C. Sulfuryl chloride (1.1 equivalents) was added slowly, and a white precipitate formed after 10-15 minutes. The resulting 2-(trifluoromethyl)benzenesulfenyl chloride was isolated by evaporation of nitromethane in 75% yield.

**Labeling Procedures**

The peptide or protein was dissolved in 30% acetic acid solution at the concentration of 1 mg/mL. Protein was denatured by 6 M guanidine HCl at 95 ºC for 10 minutes and then cooled to the room temperature. Ordinarily, cysteine is alkylated by iodoacetamide to prevent its reaction with sulfenyl chloride. However, the peptide and protein used in the following experiments contain no cysteine, therefore this step was omitted. The sample was allowed to react with a 25-fold molar excess of the labeling reagent dissolved in glacial acetic acid for 2 hours, pH 3, at

room temperature in the dark. The labeled protein was run through the gel Sephadex LH-20 to remove the excess labeling reagent and other small molecules. The protein was collected and digested with trypsin overnight at 37 ºC. The resulting peptides were analyzed by MALDI or ESI mass spectrometry.

**MALDI-FTMS and ESI-FTMS Analyses**

Matrix assisted laser desorption ionization (MALDI) mass spectra were collected on a 7 Tesla Bio-Apex Fourier Transform mass spectrometer (Bruker Daltonics, Billerica, MA) fitted with a Scout 100 source. A nitrogen laser (337 nm) was employed and the mass spectra were recorded in the positive mode with the sum of 120 laser shots. The matrices sinapinic acid and α-cyano-4-hydroxycinnamic acid (CHCA) were separately prepared as saturated solution in 50% acetonitrile with 0.1% trifluoroacetic acid. Recrystallized 2,5-dihydroxybenzoic acid (DHB) (6 mg) was dissolved in 50 μL of the solution mentioned above. Peptide solution (0.5 μL, ~1 mg/mL) was mixed with an equal volume of matrix solution on the target and allowed to dry in the dark. The target was kept away from light sources throughout the procedures.

Electrospray ionization (ESI) mass spectra were collected on a 7 Tesla Bio-Apex Fourier Transform mass spectrometer (Bruker Daltonics, Billerica, MA) using a heated metal capillary at a temperature of 135 $^{o}$C. The emitter was a 150 μm I.D. fused silica with a narrow (~20 μm) tip. A DC potential of 1150 V was applied to the solution via a liquid junction. The sample (~1 mg/mL) was diluted by 20-fold with the solution of methanol, $H_2O$ and concentrated acetic acid (49:49:2) (v/v/v) and then introduced into the mass spectrometer through the capillary at the flow rate of 1.0 μL/min.

**Tandem Mass Spectrometric Analysis**

Tandem mass spectrometry (MS/MS) of the ion of interest was performed with a 7 Tesla Bruker Apex IV QeFTMS fitted with an Apollo II ESI source, a $CO_2$ laser for infrared multiphoton dissociation (IRMPD). The sample was made to 0.1 mg/mL in a solution of methanol, $H_2O$ and formic acid (50:50:0.1) (v/v/v), and ionized by nanospray using a pulled fused silica tip (model# FS360-75-15-D-5, New Objective, Wobrun, MA). The sample solutions were infused at a rate of 10 μL/hour. For collisionally activated dissociation (CAD), the precursor ion was isolated and dissociated in the external quadrupole in the presence of argon gas. For the IRMPD experiment, the precursor ion was isolated in the external quadrupole and accumulated for 2 seconds before injection into the FTMS cell, and then irradiated with $CO_2$ laser for 12 milliseconds. 24 spectra were signal averaged per acquisition.

**RESULTS AND DISCUSSION**

**Derivatization of Tryptophan Residues with 2-NBSCl**

Although MALDI-MS analysis of 2-NBS-derivatized peptides has been reported by others [21], we find that such derivatives undergo photoinduced decomposition during MALDI analysis. Figure 2.2 shows the MALDI-FTMS mass spectrum of bombesin. A single isotopic distribution was observed for the underivatized peptide (Figure 2.2a), with a monoisotopic peak at *m/z* 1669.7817, corresponding to the protonated molecular ion of unlabeled bombesin. The derivatized peptide shows three isotopic clusters (Figure 2.2b). The highest of these clusters has a monoisotopic peak at *m/z* 1822.8032. The mass shift from underivatized bombesin is 153.0215 Da, which matches the predicted value (152.9884 Da) very well indicating the completion of the reaction. In addition to the signal from derivatized bombesin, additional peaks at *m/z* 1808.8139,

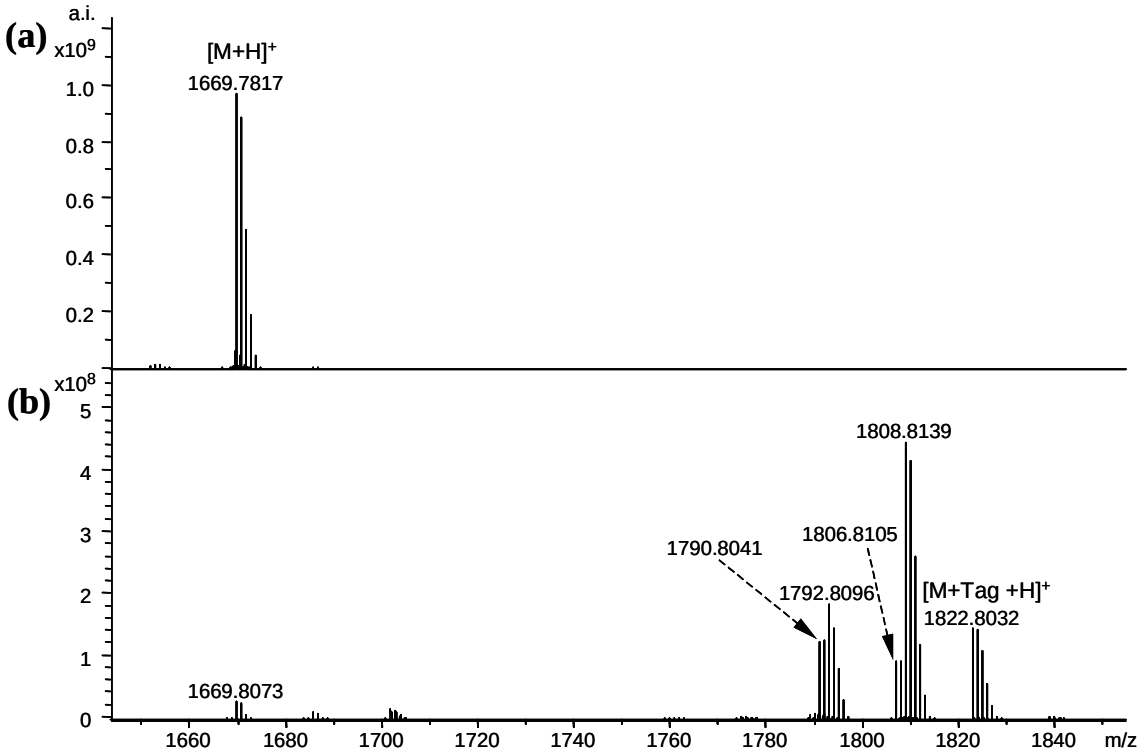



**Figure 2.2** MALDI-FTMS spectra of the bombesin (a) before and (b) after reaction with 2-NBSCl when using DHB as matrix. The monoisotopic peak of derivatized bombesin was observed at *m/z* 1822.8032 (predicted mass:1669.7817 + 152.9884 = 1822.7701 Da) and its dissociation products at *m/z* 1808.8139, 1806.8105, 1792.8096, 1790.8041.

1806.8105, 1792.8096 and 1790.8041 were observed. These may be explained as photofragments of the 2-NBSCl derivative, as nitrophenyl compounds exhibit a strong absorption band that overlaps with the radiation (337 nm) from the nitrogen laser. In order to verify this assumption, the same sample was analyzed by ESI mass spectrometry, Figure 2.3. The underivatized peptide (Figure 2.3a) shows a single isotopic cluster corresponding to the doubly charged molecular ion at *m/z* 835.4141. The derivatized peptide (Figure 2.3b) shows a single isotopic cluster corresponding to a doubly charged molecular ion at *m/z* 911.9163. The mass difference between the deconvolved masses is 153.0044 Da, in close agreement with the calculated value of 152.9884 Da. No other peaks were found in the ESI mass spectrum, strongly suggesting that the extra peaks in the MALDI spectrum are the result of photoinduced decomposition of the $[M+Tag+H]^+$ species.

From these data, we conclude that 2-NBSCl undergoes ultraviolet (UV)-induced photodissociation during MALDI-MS analysis. It is instructive to compare these results with the photodissociation products of nitrobenzene by UV irradiation, which has been previously reported by others [23, 24]. Kosmidis and coworkers have investigated the dissociation pathways of nitrobenzene using a single dye laser ranging from 225 to 275 nm in conjunction with a time-of-flight (TOF) mass spectrometer. In their studies, the parent ($C_6H_5NO_2^+$), nitrosobenzene ($C_6H_5NO^+$), phenoxy ($C_6H_5O^+$) and phenyl ($C_6H_5^+$) ions were all observed in addition to many other lighter daughter fragments. Similar to these results, we find that the derivatized bombesin undergoes loss of O (*m/z* 1806.8105) and NO (*m/z* 1792.8096), presumably from the nitrophenyl group. In addition, we observe a significant peak at *m/z* 1808.8139, corresponding to $[M+Tag+H-14]^+$, which is difficult to rationalize, as 14 amu is a highly unusual neutral loss in mass spectrometry. It is possible that this peak arises from elimination of NO from

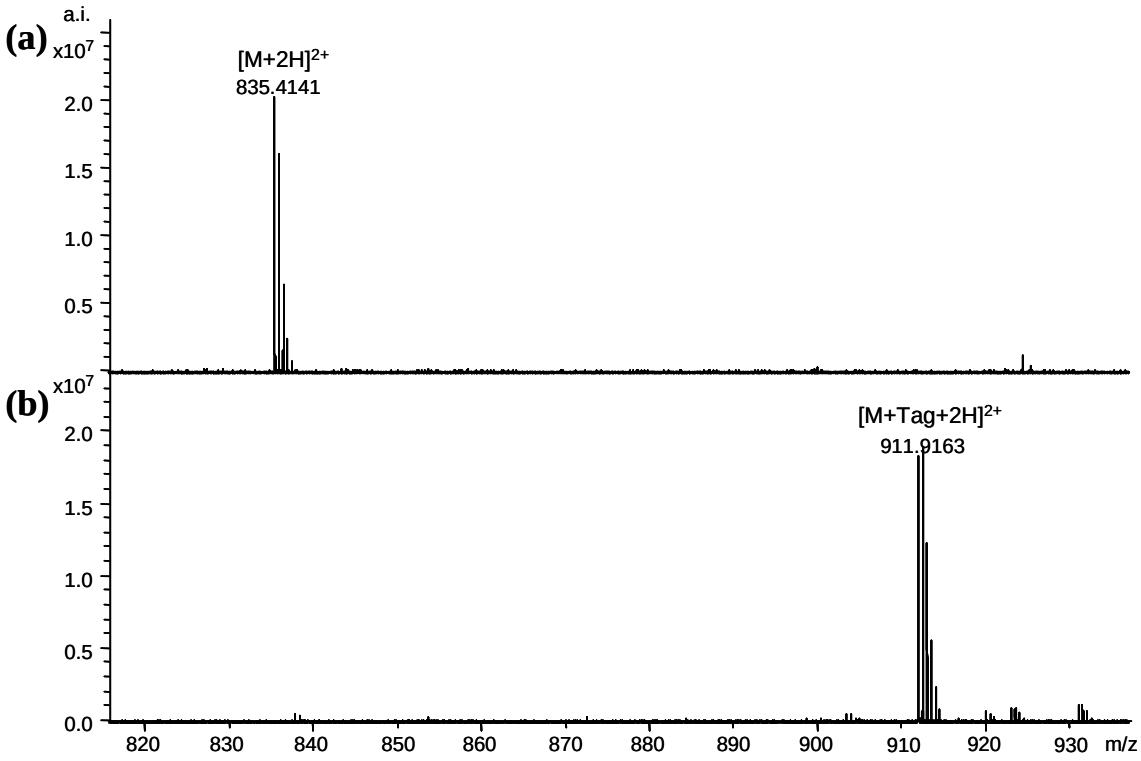



**Figure 2.3** ESI-FTMS spectra of the bombesin (a) before and (b) after reaction with 2-NBSCl. The number '+2' indicates the charge state of the ions produced.

$[M+Tag+H+16]^+$ at *m/z* 1838.8108 which corresponds to the methionine-oxidized, tryptophan-labeled peptide. However, the abundance of the *m/z* 1838.8108 ion seems too low to fully account for the abundance of this fragment ion. There is also a less abundant fragment ion at *m/z* 1790.8041, corresponding to $[M+Tag+H-32]^+$, which also does not appear in the UV photodissociation mass spectrum of nitrobenzene. Even though our knowledge of the origin of the ions at *m/z* 1808.8139 and 1790.8041 is not very clear, these observations, nevertheless, indicate that the 2-NBSCl derivative of tryptophan undergoes substantial photodissociation during MALDI through multiple fragmentation channels, particularly when using DHB as a matrix.

Although 2-NBSCl exhibits high specificity and reactivity toward tryptophan residues, its photosensitivity increases the difficulty of characterizing the derivatization products by MALDI mass spectrometry due to the multiplicity of photoproducts. In order to eliminate the dissociation of 2-NBSCl derivative, we have examined the effects of different levels of laser power and various matrices on these neutral losses. We found that dissociation was reduced with the decrease of laser power (data not shown); however, the dissociation still could not be completely avoided even at the threshold for observing ion formation. The influence of the matrices on the dissociation has been studied as well. When DHB was used as matrix, the dissociation products of derivatized bombesin were prevalent in MALDI mass spectrum. In contrast, when using sinapinic acid or α-cyano-4-hydroxycinnamic acid (CHCA) as matrix, the protonated molecular ion of derivatized bombesin was the dominant peak, as shown in Figure 2.4. Although DHB is a "cool" matrix compared to CHCA, and exhibits less metastable decomposition for peptides in FTMS analysis, surprisingly, it causes more dissociation of the NBS-labeled tryptophan. The ability of CHCA to prevent dissociation of photosensitive species

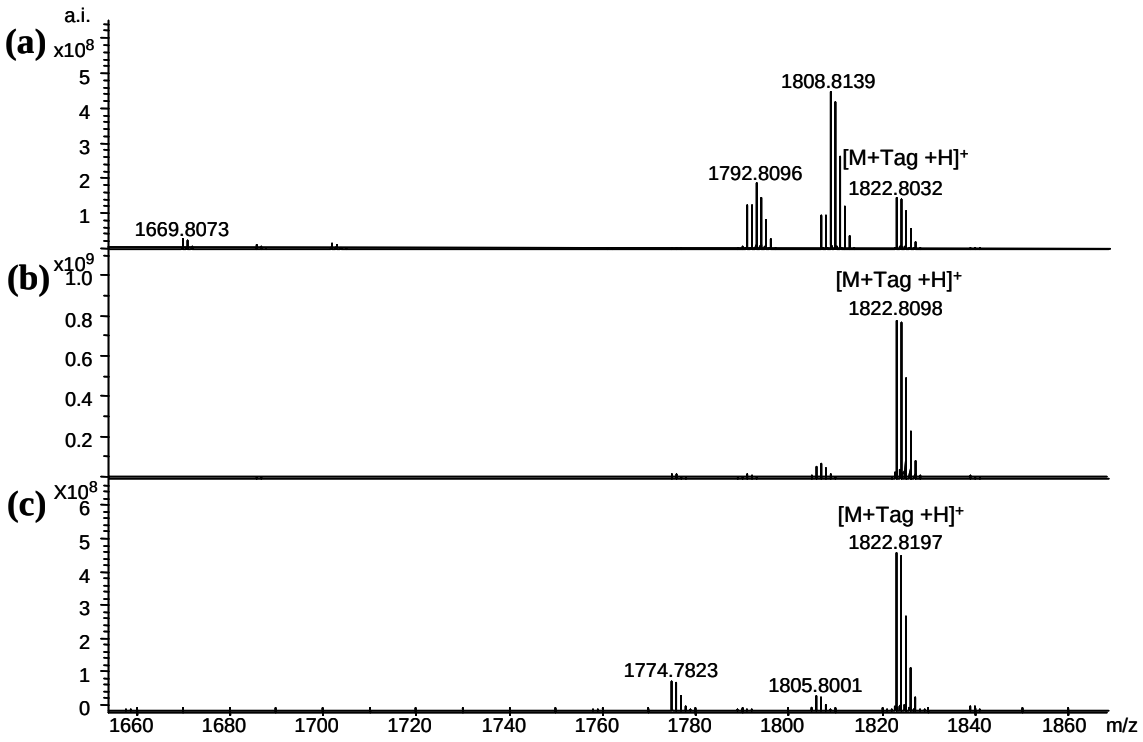



**Figure 2.4** MALDI-FTMS spectra of the derivatized bombesin when using (a) DHB, (b) sinapinic acid and (c) CHCA as matrix. The monoisotopic peaks of derivatized bombesin were observed at *m/z* 1822.8098 and 1822.8197 when using sinapinic acid and CHCA as matrix respectively.

has been reported by Low *et al* [25]. In our experiments, both sinapinic acid and CHCA showed a good ability for the detection of the 2-NBSCl derivative; however, a small amount of dissociation was still observed in their MALDI mass spectra. These observations indicate that the dissociation of 2-NBSCl derivatives can be reduced with lower laser power or with the choice of matrices; however, the complete elimination of such dissociation still could not be achieved under any set of conditions. Moreover, sinapinic acid and CHCA are not optimal matrices for underivatized peptides by FTMS, as they are prone to promoting the metastable loss of $H_2O$ or $NH_3$ from the molecular ion.

**Derivatization of Tryptophan Residues with 2-TFBSCl**

To eliminate the photodissociation of the NBS-labeled tryptophan in MALDI mass spectrometry using DHB as a matrix, we have developed a new tryptophan derivatizing reagent, 2-(trifluoromethyl)benzenesulfenyl chloride (2-TFBSCl). The choice of the 2-trifluoromethyl substituent is based on the requirement of a strong electron withdrawing group to stabilize the benzenesulfenyl chloride so that it makes a practical derivatization reagent. A second criterion is that this aromatic substituent does not produce an absorption band above 300 nm as does the nitro group in the conventional reagent for tryptophan derivatization. We have synthesized this reagent and tested its reactivity toward peptides and proteins that contain tryptophan residues.

2-TFBSCl was first used to derivatize bombesin, and the labeled bombesin was analyzed by MALDI-FTMS using DHB as matrix. As shown in Figure 2.5b, only a single isotopic cluster was observed for the derivatized peptide, with a monoisotopic peak at *m/z* 1845.7808, corresponding to the protonated molecular ion of labeled bombesin. This strongly suggests the photostability of 2-TFSCl in MALDI analysis. The mass shift is 175.9991 Da, which matches the

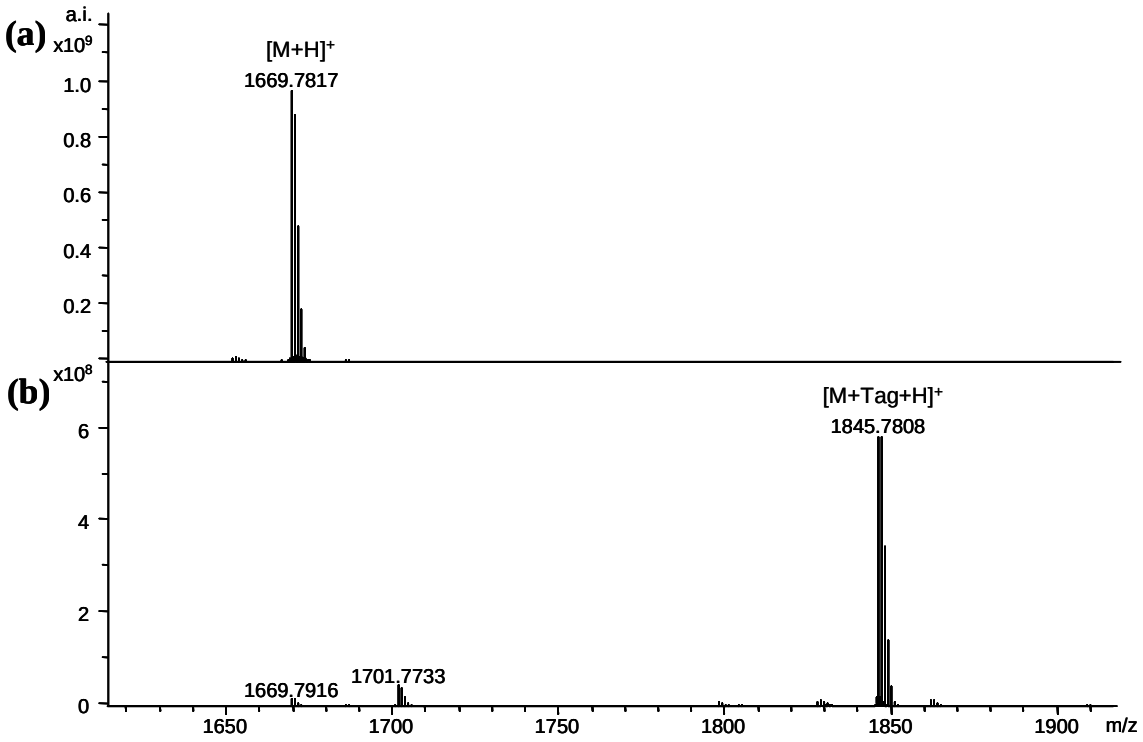



**Figure 2.5** MALDI-FTMS spectra of the bombesin (a) before and (b) after reaction with 2-TFBSCl when using DHB as matrix. The monoisotopic peak of derivatized bombesin was observed at *m/z* 1845.7808 (predicted mass: 1669.7817 + 175.9908 = 1845.7725 Da).

predicted value (175.9908 Da) very well indicating the completion of the reaction.

To test its specificity, 2-TFBSCl was then used to label the protein, myoglobin, which contains two tryptophan residues. The tryptic fragments of the derivatized myoglobin were analyzed by MALDI-FTMS using DHB as matrix. Both tryptophan residues occur in the same tryptic fragment, GLSDGEWQQVLNVWGK. Consistent with this, only one tryptic peptide was found to change mass between the underivatized and derivatized myoglobin tryptic digest mass spectra, shown in Figure 2.6. For the underivatized myoglobin, Figure 2.6a, the peak at *m/z* 1815.9041 corresponds to the protonated molecular ion of the peptide containing two tryptophan residues. It shifted to *m/z* 2167.9005, Figure 2.6b, after derivatization, indicating that both of the tryptophan residues were labeled. Although some singly labeled peptide was observed at *m/z* 1991.9163, its intensity is extremely low, indicating that the reaction has gone nearly to completion. Again, no photodissociation peaks were found for the derivatized peptide in myoglobin, demonstrating the stability of 2-TFBSCl labeled peptides during MALDI analysis.

2-TFBSCl showed high reactivity toward tryptophan-containing peptides in MALDI-FTMS analysis, and its specificity to the tryptophan residue was confirmed by MS/MS of the TFBS-labeled peptides. Both collisionally activated dissociation (CAD) and infrared multiphoton dissociation (IRMPD) experiments were performed on the labeled peptide in myoglobin. Results are shown in Figure 2.7, Table 2.1 and Table 2.2. A series of **y** ions were observed in both CAD and IRMPD spectrum. The **b6** and **b7** ions found in the CAD experiment at *m/z* 559.2375 and *m/z* 921.3099, respectively (Figure 2.7a) differ by 362.0724 Da, in close agreement with the calculated mass of the TFBS-labeled tryptophan (186.0793 + 175.9908 = 362.0701 Da), locating the site of modification to the tryptophan residue. This was also indirectly confirmed by the **y3** ion observed at *m/z* 566.2066 in Figure 2.7a, which includes the second tryptophan residue in

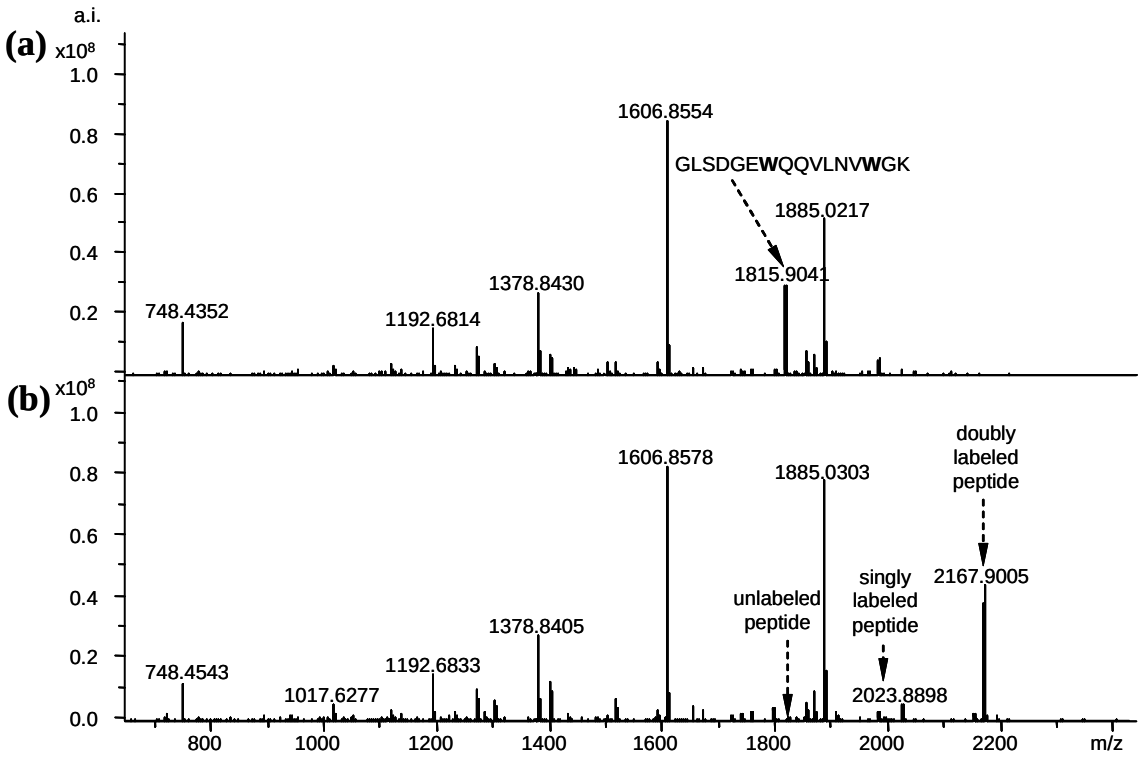



**Figure 2.6** MALDI-FTMS spectra of the tryptic fragments of (a) underivatized myoglobin and (b) 2-TFBS-derivatized myoglobin when using DHB as matrix. The peak at *m/z* 1815.9041 in (a) corresponding to a peptide containing two tryptophan residues shifted to *m/z* 2167.9005 in (b).

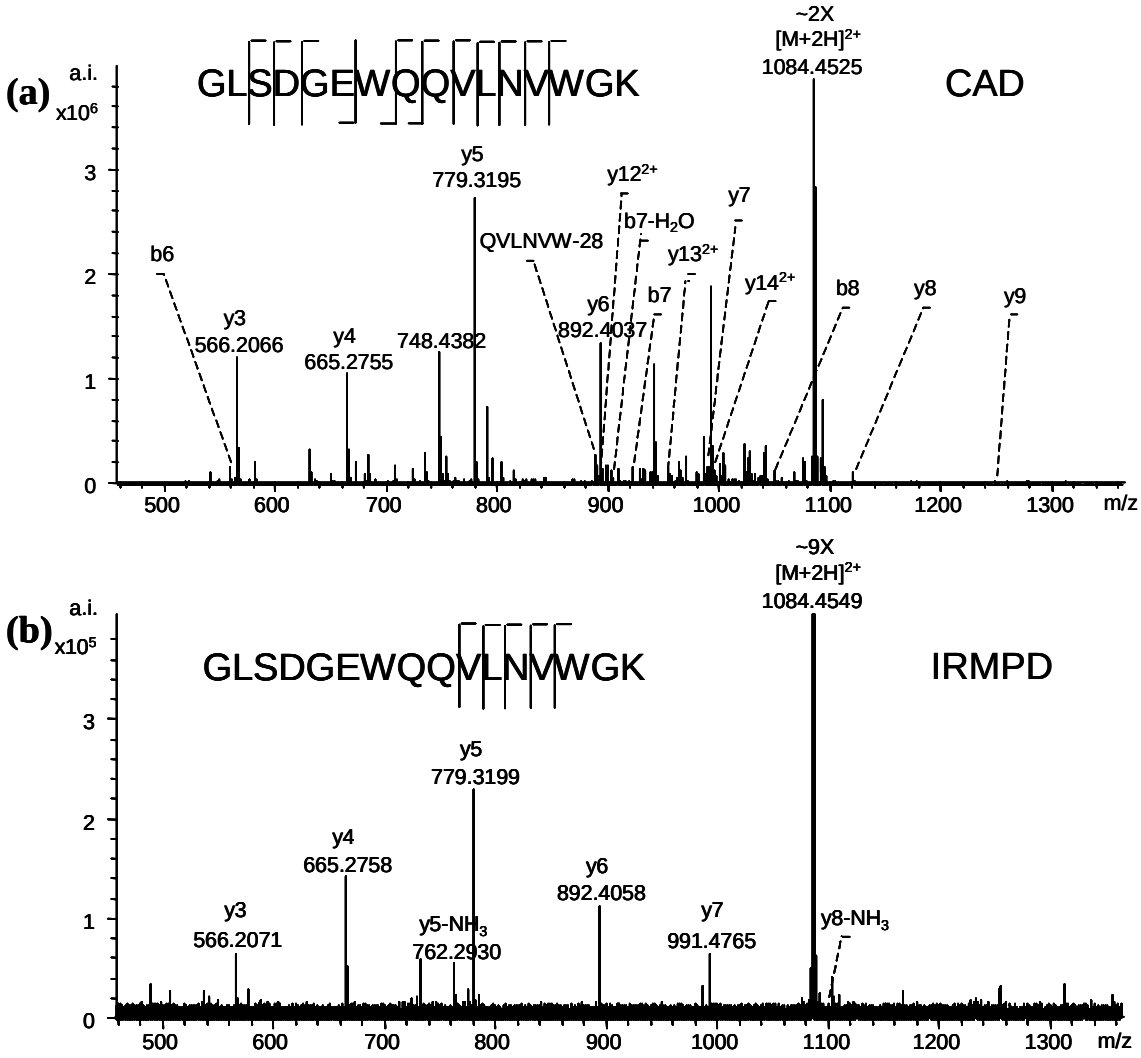



**Figure 2.7** (a) CAD spectrum and (b) IRMPD spectrum of the TFBS-labeled peptide in myoglobin (GLSDGEWQQVLNVWGK).

**Table 2.1** Assignment of peaks in CAD spectrum of GLSDGEWQQVLNVWGK labeled with TFBSCl.

| Fragment Ion | Theoretical Mass | Experimental Mass | R.I. (%) | Error in ppm |
|---|---|---|---|---|
| b6 | 559.2364 | 559.2375 | 2.4 | 2.0 |
| y3 | 566.2049 | 566.2066 | 18.3 | 3.0 |
| y4 | 665.2733 | 665.2755 | 15.7 | 3.3 |
| y5 | 779.3163 | 779.3195 | 41.7 | 4.1 |
| QVLNVW-28 | 888.4054 | 888.4032 | 4.3 | -2.5 |
| y6 | 892.4003 | 892.4037 | 20.3 | 3.8 |
| $y12^{+2}$ | 898.3639 | 898.3673 | 2.7 | 3.8 |
| b7-$H_2O$ | 903.2959 | 903.3008 | 2.0 | 5.4 |
| b7 | 921.3065 | 921.3099 | 2.6 | 3.7 |
| $y13^{+2}$ | 955.8774 | 955.8824 | 1.3 | 5.2 |
| y7 | 991.4687 | 991.4736 | 4.8 | 4.9 |
| $y14^{+2}$ | 999.3934 | 999.3971 | 2.9 | 3.7 |
| b8 | 1049.3651 | 1049.3723 | 2.1 | 6.9 |
| y8 | 1119.5273 | 1119.5341 | 1.7 | 6.1 |
| y9 | 1247.5859 | 1247.5936 | 0.4 | 6.2 |

**Table 2.2** Assignment of peaks in IRMPD spectrum of GLSDGEWQQVLNVWGK labeled with TFBSCl.

| Fragment Ion | Theoretical Mass | Experimental Mass | R.I. (%) | Error in ppm |
|---|---|---|---|---|
| y3 | 566.2049 | 566.2071 | 1.8 | 3.9 |
| y4 | 665.2733 | 665.2758 | 3.8 | 3.8 |
| y5-$NH_3$ | 762.2897 | 762.293 | 1.5 | 4.3 |
| y5 | 779.3163 | 779.3199 | 6.2 | 4.6 |
| y6 | 892.4003 | 892.4058 | 3.1 | 6.2 |
| y7 | 991.4687 | 991.4765 | 1.8 | 7.9 |
| y8-$NH_3$ | 1102.5008 | 1102.5075 | 1.2 | 6.1 |

myoglobin. These observations confirm that the reactivity of 2-TFBSCl is highly specific for tryptophan residues in proteins.

**Derivatization of Tryptophan Residues with 4,6-dibromo-2-TFBSCl**

Given the good photostability and high reaction specificity of 2-TFBSCl toward tryptophan residues, we have developed a tryptophan mass defect label (MDL) based on this compound. To function as a MDL reagent, we have incorporated two bromine atoms into 2-TFBSCl as bromine is an element with large mass defect (-82.00 mmu).We tested this MDL reagent (4,6-dibromo-2-TFBSCl) with bombesin, and the MDL-labeled bombesin was analyzed by MALDI-FTMS using DHB as matrix. As shown in Figure 2.8a, the peak with a monoisotopic mass at *m/z* 1669.7872 corresponds to the protonated molecular ion of unlabeled bombesin. After derivatization, it shifts to *m/z* 2001.5929 as shown in Figure 2.8b, corresponding to the protonated molecular ion of MDL-labeled bombesin. The mass shift is 331.8057 Da, which matches very well with the predicted value (331.8118 Da). Moreover, the peak at *m/z* 2001.5929 exhibits an isotopic distribution characteristic for the dibromo species (shown in the inset), further confirming the assignment of this peak to the MDL-labeled bombesin. Besides this major peak in Figure 2.8b, several less abundant peaks are also observed at *m/z* 1921.6713, 1953.5920 and 2017.5887, corresponding to $[M+Tag+H-80]^+$, $[M+Tag+H-48]^+$ and $[M+Tag+H+16]^+$, respectively. The peak at *m/z* 2017.5887 results from the oxidization of methionine in the MDL-labeled bombesin, but the origin of the peaks at *m/z* 1953.5920 and 1921.6713 are difficult to explain. They are unlikely the photodissociation products of $[M+Tag+H]^+$ at *m/z* 2001.5929 because of the photostability of 2-TFBSCl proven in previous studies, but possibly arise from reaction of bombesin with the side products present in the MDL reagent during synthesis. Nevertheless,

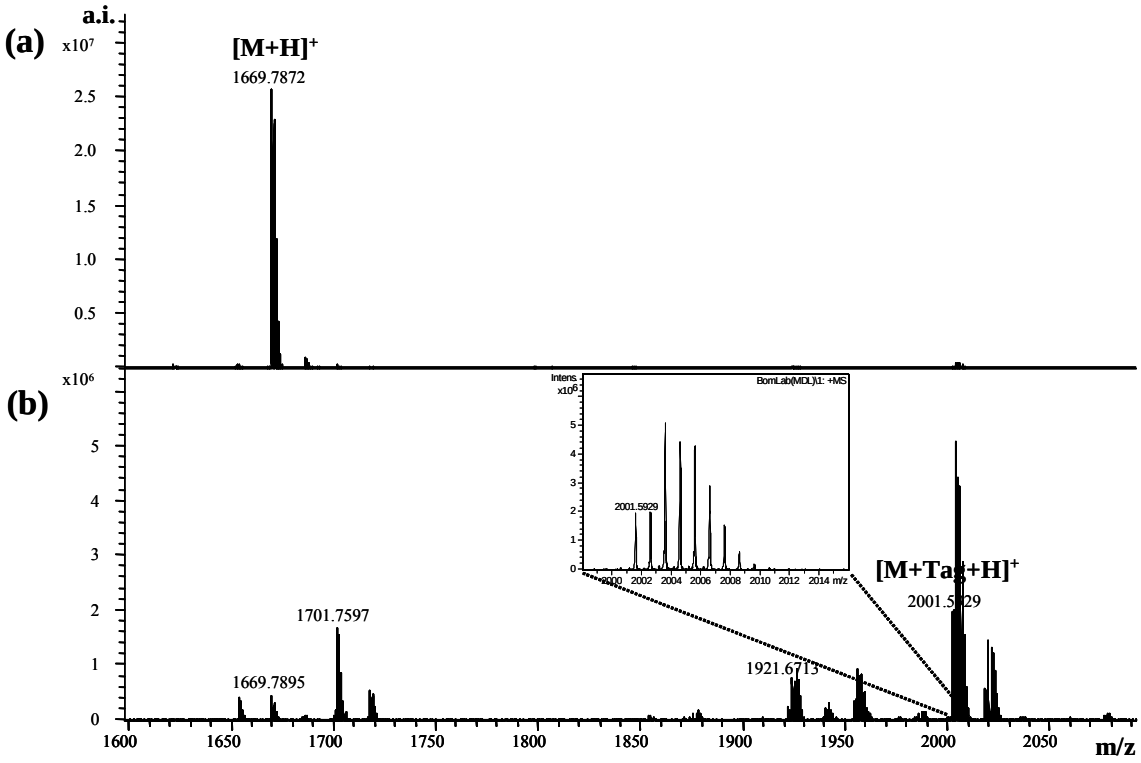



**Figure 2.8** (a) MALDI-FTMS spectra of (a) underivatized bombesin and (b) MDL-derivatized bombesin when using DHB as matrix. The peak at *m/z* 1669.7872 in (a) shifted to *m/z* 2001.5929 in (b) after derivatization.

bombesin is shown to be MDL-labeled with high conversion rate, demonstrating the effectiveness of the MDL reagent, 4,6-dibromo-2-TFBSCl, on tagging tryptophan residues.

## CONCLUSIONS

2-Nitrobenzenesulfenyl chloride has been widely used for many years as a tryptophan derivatizing reagent; however, its photosensitivity limits its applications in MALDI analysis. We reported here the development and application of a new tryptophan tag, 2-(trifluoromethyl)benzenesulfenyl chloride. This compound shows identical reaction specificity and efficiency toward tryptophan compared to 2-nitrobenzenesulfenyl chloride. Most significantly, it exhibits high photostability in MALDI analysis. Therefore, 2-(trifluoromethyl)benzenesulfenyl chloride would be a good choice of tryptophan derivatizing reagent for the studies of proteins based on MALDI-MS strategy. This compound can serve as the reactive end of a reagent for mass defect label, and shows promise as a tool for proteomics analyses.

## REFERENCES


1 Hamdan, M.; Righetti, P. G. Modern Strategies for Protein Quantification in Proteome Analysis: Advantages and Limitations. *Mass Spectrom. Rev.* **2002,** *21*, 287-302.

2 Leitner, A.; Lindner, W. Current Chemical Tagging Strategies for Proteome Analysis by Mass Spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2004,** *813*, 1-26.

3 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. *Nature Biotechnology* **1999,** *17*, 994-999.

4 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. *Anal. Chem.* **2006,** *78*, 3417-3423.

5 Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. Increased Sensitivity of Tryptic Peptide Detection by MALDI-TOF Mass Spectrometry Is Achieved by Conversion of Lysine to Homoarginine. *Anal. Biochem.* **2000,** *287*, 110-117.

6 Beardsley, R. L.; Karty, J. A.; Reilly, J. P. Enhancing the Intensities of Lysine-Terminated Tryptic Peptide Ions in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 2147-2153.

7 Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. Improved Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Tryptic Hydrolysates of Proteins Following Guanidination of Lysine-Containing Peptides. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 2070-2073.

8 Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate De Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 2348-2356.

9 Cagney, G.; Emili, A. De Novo Peptide Sequencing and Quantitative Profiling of Complex Protein Mixtures Using Mass-Coded Abundance Tagging. *Nature Biotechnology* **2002,** *20*, 163-170.

10 Creighton, T. E. Proteins: Structures and Molecular Properties. 2nd ed. *W. H. Freeman and Company*: New York, **1993**.

11 Weil, L.; James, S.; Buchert, A. R. Photooxidation of Crystalline Chymotrypsin in the Presence of Methylene Blue. *Arch. Biochem. Biophys.* **1953,** *46*, 266-278.

12 Ramachandran, L. K. Protein-Iodine Interaction. *Chemical Reviews* **1956,** *56*, 109-218.

13 Previero, A.; Coletti, M. A.; Galzigna, L. Modification of Tryptophan Residues in Trypsin Alpha-Chymotrypsin + Pepsinogen. *Biochem. Biophys. Res. Commun.* **1964,** *16*, 195-198.

14 Patchornik, A.; Lawson, W. B.; Witkop, B. Selective Cleavage of Peptide Bonds .2. The Tryptophyl Peptide Bond and the Cleavage of Glucagon. *J. Am. Chem. Soc.* **1958,** *80*, 4747-4748.

15 Koshland, D. E.; Karkhanis, Y. D.; Latham, H. G. Environmentally-Sensitive Reagent with Selectivity for Tryptophan Residue in Proteins. *J. Am. Chem. Soc.* **1964,** *86*, 1448-1450.

16 Scoffone, E.; Fontana, A.; Rocchi, R. Selective Modification of Tryptophan Residue in Peptides and Proteins Using Sulfenyl Halides. *Biochem. Biophys. Res. Commun.* **1966,** *25*, 170-174.

17 Scoffone, E.; Fontana, A.; Rocchi, R. Sulfenyl Halides as Modifying Reagents for Polypeptides and Proteins .I. Modification of Tryptophan Residues. *Biochemistry* **1968,** *7*, 971-979.

18 Fontana, A.; Scoffone, E. Sulfenyl Halides as Modifying Reagents for Polypeptides and Proteins. *Methods Enzymol.* **1972,** *25*, 482-494.

19 Fontana, A.; Veronese, F. M.; Boccu, E. Reaction of Sulfenyl Halides with Cytochrome - Novel Method for Heme Cleavage. *FEBS Lett.* **1973,** *32*, 135-138.

20 Tanaka, S.; Mohri, N.; Kihara, H.; Ohno, M. Selective Separation of Tryptophan-Containing Peptides Via Hydrophobic Modification with 2-Nitro-4-Carboxy-Phenylsulfenyl Chloride. *Journal of Biochemistry* **1985,** *97*, 1377-1384.

21 Kuyama, H.; Watanabe, M.; Toda, C.; Ando, E.; Tanaka, K.; Nishimura, O. An Approach to Quantitative Proteome Analysis by Labeling Tryptophan Residues. *Rapid Commun. Mass Spectrom.* **2003,** *17*, 1642-1650.

22 Bandgar, B. P.; Sadavarte, V. S.; Shinde, S. D.; Kamble, V. T. Rapid Synthesis of Thiophenols under Mild and Non-Aqueous Conditions. *Sulfur Letters* **2000,** *24*, 123-126.

23 Marshall, A.; Clark, A.; Jennings, R.; Ledingham, K. W. D.; Sander, J.; Singhal, R. P. Laser-Induced Dissociation, Ionization and Fragmentation Processes in Nitroaromatic Molecules. *Int. J. Mass Spectrom. Ion Processes* **1992,** *116*, 143-156.

24 Kosmidis, C.; Ledingham, K. W. D.; Clark, A.; Marshall, A.; Jennings, R.; Sander, J.; Singhal, R. P. On the Dissociation Pathways of Nitrobenzene. *Int. J. Mass Spectrom. Ion Processes* **1994,** *135*, 229-242.

25 Low, W.; Kang, J.; DiGruccio, M.; Kirby, D.; Perrin, M.; Fischer, W. H. MALDI-MS Analysis of Peptides Modified with Photolabile Arylazido Groups. *J. Am. Soc. Mass Spectrom.* **2004,** *15*, 1156-1160.

# CHAPTER 3

## IMPROVED MASS ACCURACY FOR HIGHER MASS PEPTIDES BY USING SWIFT EXCITATION FOR MALDI-FTICR MASS SPECTROMETRY

---

Chunyan Li*, Li Jing*, Richard L. Wong, Desmond A. Kaplan and I. Jonathan Amster, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 76-81

Reprinted here with the permission of the publisher

*These two authors contributed equally to this work

**ABSTRACT**

Stepwise-external calibration has previously been shown to produce sub part-per-million (ppm) mass accuracy for the MALDI-FTICR/MS analyses of peptides up to *m/z* 2500. The present work extends these results to ions up to *m/z* 4000. Mass measurement errors for ions of higher mass-to-charge are larger than for ions below *m/z* 2500 when using conventional chirp excitation to detect ions. Mass accuracy obtained by using stored waveform inverse Fourier transform (SWIFT) excitation was evaluated and compared to chirp excitation. Analysis of measurement errors reveals that SWIFT excitation provides smaller deviations from the calibration equation and better mass accuracy than chirp excitation for a wide mass range and for widely varying ion populations.

**INTRODUCTION**

Accurate mass measurement has long been recognized as a powerful tool in mass spectrometry, enabling the assignment of unique elemental compositions for small molecules (MW < 500Da) [1], and more recently, used for making higher confidence peptide identifications [2]. Accurate mass measurements are carried out using a variety of mass spectrometers. Time-of-flight (TOF) mass spectrometers now provide accuracy within 10 ppm [3, 4]. Orbitrap mass measurement accuracies have been reported to be 2 to 5 ppm [5, 6]. Fourier-transform ion cyclotron resonance (FTICR) mass spectrometry, developed by Comisarow and Marshall [7, 8], currently provides the best mass resolution and mass accuracy (< 1 ppm) of all types of mass analyzers [9-11] and has proven to be useful for protein identification by database searching [2, 12]. Mass measurement accuracy (MMA) at the sub part-per-million (ppm) level using internal calibration [13, 14] and several ppm using external calibration have been demonstrated [15, 16],

and these have led to much greater identification specificity, as described in recent reviews [17, 18].

For FTICR/MS, space-charge is the principal cause of mass measurement error [15, 19, 20]. The best MMA is obtained by using internal calibration, as this eliminates global space charge effects [16]. Conventionally, internal calibration is achieved by mixing a calibrant with the analyte. Internal calibration can be achieved without adding calibrant directly into the analyte by using a dual-spray source [14, 21] in ESI experiments or by using the internal calibration on adjacent samples (InCAS) calibration method [22, 23] in MALDI experiments. However, internal calibration requires having both calibrant and analyte ions present at the same time in the analyzer cell, which congests the mass spectrum and can lead to overlapping peaks. Such issues can be avoided with external calibration, but space-charge shifts of cyclotron frequencies can lead to systematic errors in mass measurement. The most accurate external calibration procedures rely on a calibration equation that accounts for ion intensities [15, 16, 24], or for matching the ion abundance between the analyte and calibrant spectra, e.g. by automatic gain control (AGC) [14, 25]. However, AGC is not applicable to MALDI-FTICR measurements due to the large shot-to-shot variation in ion intensity that is characteristic of MALDI.

We recently described a two-step external calibration procedure for MALDI-FTICR, stepwise-external calibration [26], in which a mass spectrum is first acquired at low trapping potential, with sub ppm mass accuracy by external calibration. This is then followed by reacquiring the spectrum at higher trapping potential for the same sample, which provides higher dynamic range. The peaks from the low trapping potential spectrum are used as “confidently-known masses” or pseudo-calibrants for internal calibration of the spectrum collected at higher trapping potential. Stepwise-external calibration provides many advantages of internal

calibration without its disadvantages. Mass accuracy has been improved 2-4 times for ions below mass-to-charge ratio (*m/z*) 2500, and a root-mean-square (RMS) error of 0.9 ppm has been demonstrated for 609 measured peptide ions, whose mass errors distribute in a Gaussian fashion. Although most tryptic peptides have molecular weights less than 2500 Da, higher mass peptides have greater information content, and the protein identification rate can be increased significantly by incorporating data from spectra acquired with tuning parameter to enhance ions of higher mass-to-charge [27]. When the stepwise-external calibration approach is applied to ions above *m/z* 2500, we find that the RMS error increases to approximately 3 ppm [27]. Work by Masselon *et al*. suggests that random error in FTICR mass measurement may be related to the type of excitation waveform used for ion detection [24]. In our previous work, all measurements on the Bruker FTICR mass spectrometer were made using frequency-sweep (chirp) excitation. Smith and coworkers have shown that data collected using stored-waveform inverse Fourier transform (SWIFT) excitation provides better MMA than using chirp excitation for ESI measurements of ions up to *m/z* 1800 [24]. Presumably, this is a result of a more uniform power applied across all frequencies leading to a more uniform distribution of radii of gyration for all ions by SWIFT compared to chirp [28]. These results have encouraged us to examine this approach to improving mass accuracy for higher mass singly-charged ions when using stepwise-external calibration.

Here we present results of a comparison of chirp and SWIFT excitation for accurate mass measurement in MALDI-FTICR/MS using stepwise-external calibration. First, we evaluate the standard deviation (SD) of internal calibration mass error as a function of ion excitation power using chirp and SWIFT excitation. We also examine the MMA that can be obtained for ions up to *m/z* 4000 by chirp and SWIFT excitation using two calibration procedures, conventional

external calibration [19] and stepwise-external calibration [26]. We show that SWIFT excitation provides significantly better mass accuracy, particularly at higher mass-to-charge, than can be achieved by chirp excitation when using stepwise-external calibration.

## EXPERIMENTAL

Bovine serum albumin (BSA), and chicken egg albumin (ovalbumin) were purchased from Sigma (St. Louis, MO). Each protein sample was dissolved in alkaline solution (10 mM ammonium bicarbonate) to make a 1mg/mL solution and denatured by heating at 90 °C for 10 min. Disulfide bonds were reduced with tris (2-carboxythyl) phosphine (Pierce Biotechnology, Rockford, IL). Denatured proteins were digested overnight at 37 °C using trypsin (Promega, Madison, WI) at a 1:50 protease/protein ratio (by mass). 400 nL of the digested proteins was applied to a stainless steel MALDI target and 400 nL of 1 M 2,5-dihydroxybenzoic acid (DHB) (Lancaster, Pelham, NH) in acetonitrile/water/TFA (50%:50%:0.1%, v/v/v) was added as the MALDI matrix.

Mass spectra were collected using a BioApex 7 tesla FTICR mass spectrometer equipped with an intermediate pressure Scout 100 MALDI source (Bruker Daltonics Inc, Billerica, MA). Ions from up to 10 laser desorption events were accumulated in the source hexapole ion guide, and then the ions were released from the ion guide and transmitted to the analyzer cell with electrostatic ion optics. The delay (D2) between ion extraction from the source hexapole and the gating of the cell entrance electrode potentials, EV1 (-7.0 V) and EV2 (-1.5 V), back to the trapping potential was set to 4.0 ms to enhance the detection of heavier ions [27]. No sidekick deflection was employed (DEV2 = 0). The ions were excited by using either a chirp waveform or a SWIFT waveform [29]. The chirp excitation consisted of 95 frequencies with a dwell time of

20 μs per step, and frequency increment of 2000 Hz, yielding a 1.9 ms signal duration that covered the frequency range of 21,511 to 215,277 Hz, corresponding to *m/z* 5000 to 500. The optimal excitation voltage was 100 $V_{p-p}$ amplitude. For SWIFT excitation experiments, waveforms were produced by a PXI box with a PXI-8184 embedded controller and a PXI-5412 arbitrary waveform generator (National Instruments, Austin, TX). The power spectrum was designed to have flat amplitude over the frequency range of 21,511 to 215,277 Hz (corresponding to *m/z* 500-5000). A quadratic phase versus frequency function was used for the inverse Fourier transform calculation to yield a SWIFT excitation waveform that had relatively constant amplitude versus time [29, 30]. The excitation waveform was digitized o 32,768 points, which were read out at rate of 16,667,000 points/s to yield an excitation signal with a duration of 1.966 ms. After ion excitation, a 512K point transient was acquired, apodized with a sinebell function and padded with one zero-fill before fast Fourier transformation and magnitude calculation. Only monoisotopic peaks a with signal-to-noise ratio (S/N) >3 were used to study the mass accuracy.

For stepwise-external calibration, the first mass spectrum is acquired at a low trapping potential (0.60 V) using external calibration, eq 1, where the average mass accuracy is quite good (RMS error <0.5 ppm). The second mass spectrum is acquired for the same sample at higher trapping potential (1.10 V) to increase the detection dynamic range by a factor of 5-10 [26]. For peaks that appear in both mass spectra, the fairly accurate mass values from the first mass spectrum are used to calibrate the masses in the higher trapping potential spectrum using eq 2. Multi-linear regression is used to obtain the calibration constants by fitting the measured frequencies and intensities to the masses that are determined from the first mass spectrum. All statistical data were analyzed using Microsoft Excel. Multi-linear regression to obtain the

constants for the calibration equations was preformed using software developed in our laboratory (available for download at http://amstersgi.chem.uga.edu/html/software.html).

$$\frac{m}{z} = \frac{A}{f + B} \qquad (1)$$

$$\left(\frac{m}{z}\right)_i = \frac{A}{f_i + B + C \cdot I_i} \qquad (2)$$

## RESULTS AND DISCUSSION

### Error Analysis for Calibrant Points

Experiments were first performed to compare the standard deviations of mass errors that result from using chirp and SWIFT excitation waveforms with internal calibration. Five mass spectra of the tryptic peptides of BSA were acquired for chirp or SWIFT excitation at different values of excitation amplitude, normalized to a value of 100% for the power required to eject all ions. In this manner, we were able to compare chirp and SWIFT excitation of ions to a similar cyclotron radius. From each spectrum, 8 monoisotopic peaks were selected and fit by linear regression to eq 1, where *f* is the measured frequency of the ions, *m/z* is the theoretical mass-to-charge value, and *A* and *B* are constants that are related to the magnetic field strength and radial components of trapping potential and the global space charge field, respectively [19].

Figure 3.1 shows the deviation between measurements of the eight masses used for calibration and the values calculated from eq 1, plotted against normalized ion excitation power for chirp and SWIFT excitation; the height of the error bars represent 1 SD based on five replicate measurements at each power value. As can be seen from this plot, the SD of the mass measurement error for both excitation methods depends on the ion excitation power and the smallest errors are achieved when ion excitation power is optimized (~50%). When the ion

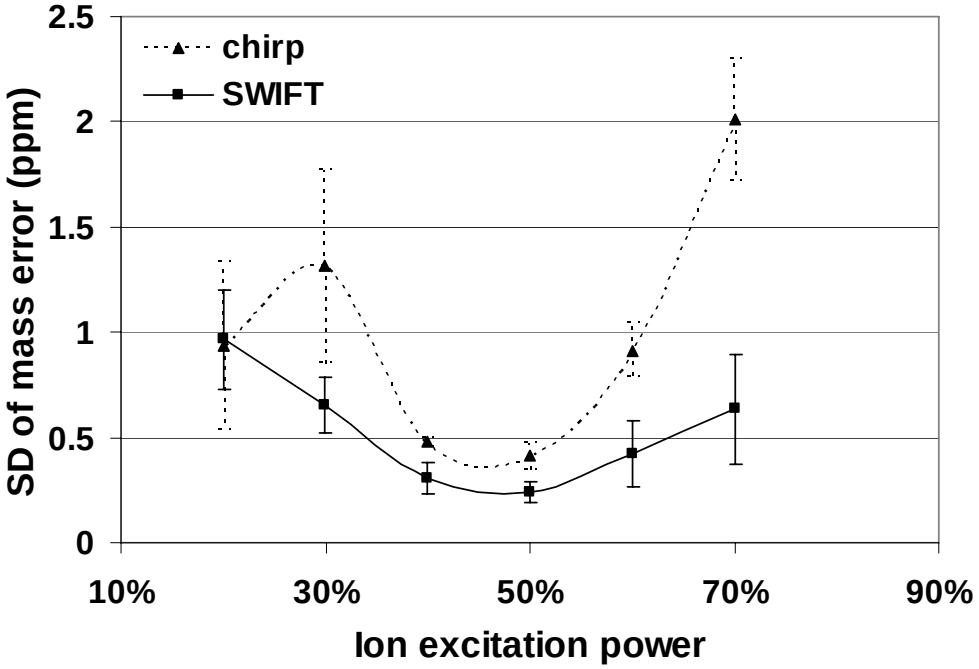



**Figure 3.1** Standard deviation of calibration errors obtained from the linear fit to eq 1 versus ion excitation power (the power to eject all ions is normalized to 100%) using chirp or SWIFT excitation. Five measurements were averaged at each condition and the error bar representing 1 SD of the mean.

excitation power is either larger or smaller than the optimal value, the deviation from the calibration equation becomes larger. The curve for data collected using chirp displays an irregular trend when the ion excitation power is 20% of the value required for ejection, which results from the peaks at high mass region falling below the minimum S/N (<3) and being excluded as calibrants. SWIFT excitation is found to provide a smaller SD than chirp at all excitation powers and maintains an acceptably small value across a much wider range of excitation values. This shows that mass accuracy is less sensitive to tuning of the excitation power for SWIFT versus chirp measurements. While the data shown in Figure 3.1 was collected at low trapping potential (0.6 V), the same results are observed at the higher trapping potential used for stepwise calibration (1.1 V).

The total ion intensity is also compared for the same data set using chirp and SWIFT. They are essentially the same when the same ion excitation power is used and the highest total ion intensity is achieved when the ion excitation power is around 50% (data not shown). These results indicate that accurate mass measurements can be obtained under conditions of optimal sensitivity via both chirp and SWIFT. SWIFT excitation provides better mass accuracy than chirp and provides good mass accuracy even when the ion excitation power is not tuned to its optimal value.

**Stepwise-External Calibration for Ions up to *m/z* 4000**

Stepwise-external calibration was examined for higher mass tryptic peptides, using an ovalbumin tryptic digest. Thirty mass spectra were collected at 0.60 V and 1.10 V cell trapping potential using either chirp or SWIFT excitation with 50% ion excitation power yielding a total 120 spectra. By using a low trapping potential (0.60 V for this experiment), the space charge

induced frequency shifts are significantly reduced due to the smaller ion capacity of the analyzer cell, so that highly accurate mass values can be obtained using external calibration [26]. However, the lower ion capacity of the analyzer cell reduces the S/N and dynamic range of the mass spectra obtained in this manner [26]. The mass spectra acquired at higher trapping potential (1.10 V) are used to recover the lost dynamic range in the lower trapping potential experiment. Stepwise-external calibration is achieved using the mass values measured in low trapping potential as calibration reference masses for the spectrum acquired at higher trapping potential via eq 2, where $I_i$ is the peak intensity of a particular ion, $(m/z)_i$. Parameter $C$ acts as a correction factor for local space-charge effects, i.e. the different interaction between ions of the same mass-to-charge versus ions of different mass-to-charge [24]. The calibration coefficients $A$, $B$, and $C$ are independent of the index $i$. This calibration method takes into account both global and local space charge effects and has been demonstrated to improve the mass accuracy when compared to application of the calibration eq 1 [26]. To calibrate a mass spectrum acquired at higher trapping potential using eq 2, one need at least three reference values from the low potential mass spectrum. The corresponding peaks are identified in the high potential mass spectrum, and their frequencies, *m/z* values, and intensities are used to obtain the calibration constants for eq 2. To get the best overall fit for multi-linear regression, all isotopic peaks with S/N >20 were used to do the stepwise-external calibration and 12 monoisotopic peaks over a wide range (from *m/z* 1346 to *m/z* 3863) were selected from each high potential spectrum to evaluate MMA.

Mass errors for each reference monoisotopic *m/z* value obtained for the external and stepwise-external calibration are plotted in Figure 3.2. Figure 3.2a and 3.2c show the mass errors obtained at high trapping potential (1.10 V) for standard external calibration using eq 1 with chirp and SWIFT excitation, respectively. Figure 3.2b and 3.2d show the mass errors obtained

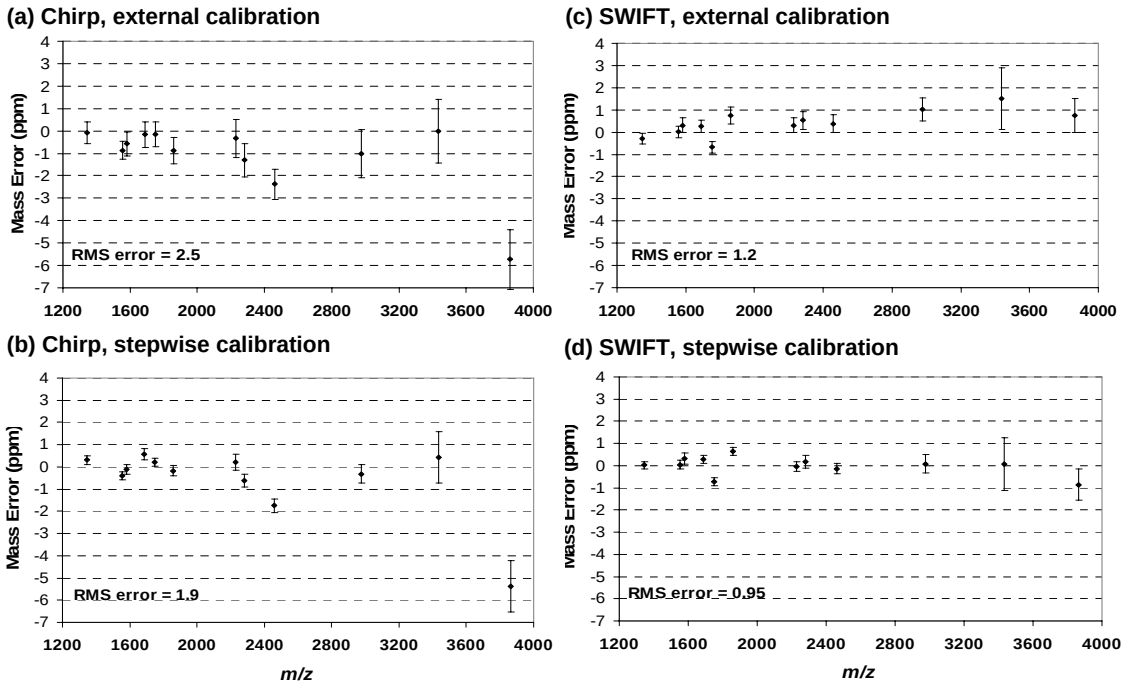



**Figure 3.2** Average calibration errors obtained from external calibration (eq 1) and stepwise-external calibration (eq 2), with the error bar being 1 SD for 30 replicates. (a) chirp for external calibration, (b) chirp for stepwise-external calibration, (c) SWIFT for external calibration, (d) SWIFT for stepwise-external calibration.

for the same data sets but calibrated with stepwise-external calibration using eq 2 with chirp and SWIFT excitation, respectively. The height of the error bars indicates 1 SD for 30 spectra obtained under the same conditions. As seen in Figure 3.2, the mass errors are noticeably reduced when stepwise-external calibration is applied in data collected from both chirp and SWIFT. As expected, the mass error (RMS error = 2.5) is the largest with external calibration using chirp excitation waveform (Figure 3.2a) and SWIFT-excite data calibrated with stepwise-external calibration yields the smallest mass error (RMS error = 0.95) (Figure 3.2d). In addition, the mass errors for ions with *m/z* >2000 indicated that better mass accuracy is achieved by using SWIFT, for both external calibration and stepwise-external calibration (Figures 3.2c and 3.2d).

Table 3.1 compares the SD and RMS of mass errors that were calculated for 30 chirp-excite spectra and 30 SWIFT-excite spectra obtained at high trapping potential with external calibration, internal calibration and stepwise-external calibration. Three different mass ranges were considered for data from each excitation waveform and all errors are expressed in ppm. The SD and RMS of mass errors from external calibration are noticeably reduced when internal calibration or stepwise-external calibration is used, and the errors for stepwise-external calibration results are only slightly larger than by internal calibration. As previously observed [26], by using stepwise-external calibration, the values of RMS and SD are around 0.9 ppm for chirp excitation of ions less than *m/z* 2500, a ~2-fold improvement in mass accuracy compared to external calibration. For ions in the high mass range (2500<*m/z*<4000), stepwise-external calibration with chirp excitation produces are relatively small improvement in MMA (RMS 3.6 ppm versus 4.2 ppm for standard external calibration). However, even internal calibration produces large errors at these higher masses when using chirp excitation (RMS 2.8 ppm). In contrast, for SWIFT excitation, internal calibration produces RMS errors of only 1.2 ppm.

**Table 3.1** Error analysis for data collected using chirp and SWIFT excitation with external calibration, internal calibration and stepwise-external calibration.

| | $m/z < 2500$ | | $2500 < m/z < 4000$ | | All Ions | |
|---|---|---|---|---|---|---|
| | SD | RMS | SD | RMS | SD | RMS |
| Chirp External | 1.4 | 1.6 | 3.6 | 4.2 | 2.2 | 2.5 |
| Chirp Internal | 0.54 | 0.55 | 2.7 | 2.8 | 1.4 | 1.4 |
| Chirp Stepwise | 0.83 | 0.85 | 3.2 | 3.6 | 1.9 | 1.9 |
| SWIFT External | 0.74 | 0.78 | 1.9 | 2.1 | 1.2 | 1.2 |
| SWIFT Internal | 0.45 | 0.46 | 1.2 | 1.2 | 0.71 | 0.71 |
| SWIFT Stepwise | 0.55 | 0.56 | 1.7 | 1.7 | 0.95 | 0.95 |

Stepwise-external calibration with SWIFT excitation produces RMS error values of 0.56, 1.7, 0.95 ppm for the low mass range, high mass range and the entire mass range, respectively. For the same mass range and the same calibration method, the values of SD and RMS from chirp-excite experiments are almost double those obtained using SWIFT. Thus, it can be confidently stated that mass errors are reduced by using SWIFT excitation, particularly for ions of higher mass-to-charge. Through the comparison of these waveforms and calibration procedures, we are able to improve MMA to sub ppm for ions up to *m/z* 3000, and to less than 2 ppm for ions up to *m/z* 4000 by using SWIFT excitation and the stepwise-external calibration method.

These results suggest substantial advantages of this methodology for proteomics analysis by MALDI-FTICR/MS. Batch tryptic digests of proteomes produce many thousands of peptides components, and even after offline liquid chromatography, individual fractions are highly complex. Such samples are highly amenable to the stepwise calibration procedure, as the larger number of peaks that appear in the mass spectrum acquired at low trapping potential allow a greater number of pseudo-calibrants for the stepwise calibration procedure, resulting in a more accurate calibration curve. Furthermore, the larger the number of components, the more uniform is the total ion intensity from fraction to fraction, which decreases space-charge frequency shifts that might otherwise limit mass accuracy in the first step of the stepwise calibration procedure, which relies on external calibration. Conversely, this procedure will have little utility for online-HPLC ESI-FTICR/MS measurements, where the number of components present in a mass spectrum is relatively low, and the variation in total ion intensity is higher (in the absence of automatic gain control.)

## CONCLUSIONS

The utilization of a SWIFT excitation waveform reduced the mass errors significantly compared to chirp excitation, particularly for ions of higher mass-to-charge. SWIFT provides a flat power distribution across the range of *m/z* values that are undergoing excitation, leading to more uniform cyclotron radii for all ions [28, 29]. Stepwise-external calibration yields a RMS error of 0.95 ppm for ions up to *m/z* 4000 using SWIFT excitation. These data demonstrate that sub-ppm MMA can be achieved when SWIFT excitation is combined with stepwise-external calibration. This is particularly significant for MALDI measurements where there is considerate variation in ion intensity from spectrum to spectrum. We are currently working on extending these findings to proteomics measurements using MALDI-FTICR/MS, as this level of MMA will allow us to perform searches at a mass tolerance of 3 ppm (3 times the RMS) with 99.7% confidence in the mass accuracy. It is noteworthy that these experiments were performed with a 7 T magnetic field; the MMA is expected to improve even further by using higher magnetic fields [31]. The improvement of MMA for ions up to *m/z* 4000 should benefit analysis of complex mixtures such as batch proteolytic digests of protein mixtures, where higher dynamic range and higher mass accuracy is required.

## REFERENCES


1 Gross, M. L. Accurate Masses for Structure Confirmation. *Journal of the American Society for Mass Spectrometry* **1994,** *5*, 57-57.

2 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.

3 Williams, J. D.; Flanagan, M.; Lopez, L.; Fischer, S.; Miller, L. A. D. Using Accurate Mass Electrospray Ionization-Time-of-Flight Mass Spectrometry with in-Source Collision-Induced Dissociation to Sequence Peptide Mixtures. *J. Chromatogr. A* **2003,** *1020*, 11-26.

4 Wyatt, M. F.; Stein, B. K.; Brenton, A. G. Investigation into Accurate Mass Capability of Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry, with Respect to Radical Ion Species. *Journal of the American Society for Mass Spectrometry* **2006,** *17*, 672-675.

5 Yates, J. R.; Cociorva, D.; Liao, L.; Zabrouskov, V. Performance of a Mass Analyzer with Orbital Trapping for Peptide Analysis. *Anal. Chem.* **2006,** *78*, 493-500.

6 Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal. Chem.* **2006,** *78*, 2113-2120.

7 Comisarow, M. B.; Marshall, A. G. Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chemical Physics Letters* **1974,** *25*, 282-283.

8 Comisarow, M. B.; Marshall, A. G. Frequency-Sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chemical Physics Letters* **1974,** *26*, 489-490.

9 Gorshkov, M. V.; Guan, S.; Marshall, A. G. Masses of Stable Neon Isotopes Determined at Parts Per Billion Precision by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *International Journal of Mass Spectrometry and Ion Processes* **1993,** *128*, 47-60.

10 Rodgers, R. P.; White, F. M.; Hendrickson, C. L.; Marshall, A. G.; Andersen, K. V. Resolution, Elemental Composition, and Simultaneous Monitoring by Fourier Transform Ion

Cyclotron Resonance Mass Spectrometry of Organosulfur Species before and after Diesel Fuel Processing. *Anal. Chem.* **1998,** *70*, 4743-4750.

11 He, F.; Hendrickson, C. L.; Marshall, A. G. Baseline Mass Resolution of Peptide Isobars: A Record for Molecular Mass Resolution. *Anal. Chem.* **2001,** *73*, 647-650.

12 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.

13 Burton, R. D.; Matuszak, K. P.; Watson, C. H.; Eyler, J. R. Exact Mass Measurements Using a 7 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer in a Good Laboratory Practices-Regulated Environment. *Journal of the American Society for Mass Spectrometry* **1999,** *10*, 1291-1297.

14 Williams, D. K.; Muddiman, D. C. Parts-Per-Billion Mass Measurement Accuracy Achieved through the Combination of Multiple Linear Regression and Automatic Gain Control in a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer. *Anal. Chem.* **2007,** *79*, 5058-5063.

15 Easterling, M. L.; Mize, T. H.; Amster, I. J. Routine Part-Per-Million Mass Accuracy for High- Mass Ions: Space-Charge Effects in MALDI-FTICR. *Anal. Chem.* **1999,** *71*, 624-632.

16 Muddiman, D. C.; Oberg, A. L. Statistical Evaluation of Internal and External Mass Calibration Laws Utilized in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2005,** *77*, 2406-2414.

17 Liu, T.; Belov, M. E.; Jaitly, N.; Qian, W. J.; Smith, R. D. Accurate Mass Measurements in Proteomics. *Chemical Reviews* **2007,** *107*, 3621-3653.

18 Zhang, L. K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. Accurate Mass Measurements by Fourier Transform Mass Spectrometry. *Mass Spectrometry Reviews* **2005,** *24*, 286-309.

19 Francl, T. J.; Sherman, M. G.; Hunter, R. L.; Locke, M. J.; Bowers, W. D.; McIver, R. T. Experimental Determination of the Effects of Space Charge on Ion Cyclotron Resonance Frequencies. *International Journal of Mass Spectrometry and Ion Processes* **1983,** *54*, 189-199.

20 Ledford, E. B.; Rempel, D. L.; Gross, M. L. Space-Charge Effects in Fourier-Transform Mass-Spectrometry - Mass Calibration. *Analytical Chemistry* **1984,** *56*, 2744-2748.

21 Hannis, J. C.; Muddiman, D. C. A Dual Electrospray Ionization Source Combined with Hexapole Accumulation to Achieve High Mass Accuracy of Biopolymers in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2000,** *11*, 876-883.

22 O'Connor, P. B.; Costello, C. E. Internal Calibration on Adjacent Samples (InCAS) with Fourier Transform Mass Spectrometry. *Anal. Chem.* **2000,** *72*, 5881-5885.

23 Mize, T. H.; Amster, I. J. Broad-Band Ion Accumulation with an Internal Source MALDI-FTICR-MS. *Analytical Chemistry* **2000,** *72*, 5886-5891.

24 Masselon, C.; Tolmachev, A. V.; Anderson, G. A.; Harkewicz, R.; Smith, R. D. Mass Measurement Errors Caused By "Local" Frequency Perturbations in Fticr Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2002,** *13*, 99-106.

25 Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. Novel Linear Quadrupole Ion Trap/Ft Mass Spectrometer: Performance Characterization and Use in

the Comparative Analysis of Histone H3 Post-Translational Modifications. *J. Proteome Res.* **2004,** *3*, 621-626.

26 Wong, R. L.; Amster, I. J. Sub Part-Per-Million Mass Accuracy by Using Stepwise-External Calibration in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2006,** *17*, 1681-1691.

27 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.

28 Marshall, A. G.; Wang, T. C. L.; Ricca, T. L. Tailored Excitation for Fourier Transform Ion Cyclotron Mass Spectrometry. *J. Am. Chem. Soc.* **1985,** *107*, 7893-7897.

29 Guan, S.; Marshall, A. G. Stored Waveform Inverse Fourier Transform (SWIFT) Ion Excitation in Trapped-Ion Mass Spectrometry: Theory and Applications. *International Journal of Mass Spectrometry and Ion Processes* **1996,** *157-158*, 5-37.

30 Guan, S. General Phase Modulation Method for Stored Waveform Inverse Fourier Transform Excitation for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Chem. Phys.* **1989,** *91*, 775-777.

31 Marshall, A. G.; Guan, S. Advantages of High Magnetic Field for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Communications in Mass Spectrometry* **1996,** *10*, 1819-1823.

# CHAPTER 4

## SHOTGUN PROTEOMIC ANALYSIS USING ACCURATE MASS MEASUREMENT AND NITROGEN STOICHIOMETRY — A HPLC-MALDI-FTICR/MS APPROACH

---

Chunyan Li*, Li Jing*, Richard L. Wong, Bryan A. Parks, Yuchen Liu, William B. Whitman, Roman A. Zubarev, Joshua S. Sharp and I. Jonathan Amster. To be submitted to *J. Am. Soc. Mass Spectrom.*

*These two authors contributed equally to this work

**ABSTRACT**

A shotgun proteomics approach is described for simultaneous identification and quantitation of the proteins from a proteome using accurate mass measurement and nitrogen stoichiometry. This method was tested on a $^{15}N$-metabolically labeled proteome sample from *Methanococcus maripaludis* using offline HPLC-MALDI-FTICR/MS. We demonstrate here the utilities of $^{15}N$-metabolic labeling for protein identification when using nitrogen stoichiometry as an additional search constraint and for protein quantitation from determining the intensity ratios of light/heave ($^{14}N/^{15}N$) peptide pairs. Our previous work has shown that sub part-per-million (ppm) mass accuracy can be achieved by combining stepwise-external calibration with stored waveform inverse Fourier transform (SWIFT) excitation for MALDI-FTICR/MS analyses of the peptides from protein standards over a wide mass range (up to *m/z* 4000) [1, 2]. We present here the results by extending these findings to proteomic measurements and demonstrate that this approach significantly improves the mass measurement accuracy (MMA) and therefore greatly enhances the protein identification.

**INTRODUCTION**

Proteomics is the systematic study of the proteins expressed by a cell, tissue or organism. The technology for proteomic analysis integrates the separation science for separation of proteins and peptides, analytical science for identification and quantitation of proteins, and bioinformatics for data management and analysis. The proteomics is conventionally performed by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) for protein separation, mass spectrometry (MS) for protein detection and genome sequence database for protein identification [3]. Although 2D-PAGE technology has been demonstrated a powerful tool for protein

separation, it has inherent weaknesses, such as the difficulty of automation, low throughput, and the significant sensitivity and dynamic range limitations [4-10]. To overcome these problems, researchers have moved to the use of liquid chromatography (LC) as an alternative separation method to 2D-PAGE in proteomics. Mass spectrometry plays a central role in proteomics, and a variety of MS-based proteomics approaches has been developed in past twenty years. Currently, the most effective approaches for protein characterization using mass spectrometry are "bottom-up" strategies, which involve proteolytically digesting proteins into smaller peptide fragments prior to mass spectrometric analysis [11]. In recent years, of all "bottom-up" approaches, shotgun proteomics has gained the widest acceptance. This approach is based on a batch digestion of an unseparated protein mixture, separation of the resulting peptides by one- or multidimensional liquid chromatography, and protein identification from mass spectrometric analysis of peptides [12].

Two mass spectrometric approaches are commonly used in shotgun proteomics for protein identification. One approach is based on the fragmentation of one or more peptides generated from a protein using tandem mass spectrometry (tandem MS or MS/MS), and the other is based upon the accurate mass measurement of these peptides using high resolution mass spectrometer (*e.g.* Fourier transform ion cyclotron resonance mass spectrometer, FTICR/MS). In the first approach, protein identification is made from both peptide mass and the sequence information obtained by tandem MS. When analyzing complex biological samples, tandem MS is usually combined with one- or multidimensional liquid chromatography to reduce sample complexity. Yates and coworkers have pioneered the use of multidimensional liquid chromatography and tandem MS for examining large number of peptides from batch digestion of a complex biological sample. With this approach, referred as to multidimensional protein identification technology

(MudPIT) [12, 13], peptides are first separated by two-dimensional liquid chromatography which integrates strong cation exchange (SCX) resin and reversed-phase resin in a biphasic column, and then the elutes are directly introduced into an electrospray ionization (ESI) source for MS and MS/MS analyses, leading to a significant number of proteins identified from the proteome studied. Tandem MS has been broadly used in proteomics and is still the chief analytical technique for protein identification to data; however, tandem MS has some limitations. First, tandem MS is low throughput because it requires the generation of a fragmentation spectrum for each peptide in a mixture that may contain thousands of components. Secondly, tandem MS is computationally intensive due to the need for processing enormous amount of the data generated from tandem MS measurements. Thirdly, data-dependent LC-MS/MS has an inherent "undersampling" limitation whereby only a portion of the species observed in the first stage of MS is selected for fragmentation [14].

In contrast to tandem MS, protein identification by accurate mass measurement is based on the correlation of measured peptide masses with ones for all possible proteolytic peptides predicted from an *in silico* digestion of the genome. With good enough mass measurement accuracy (MMA), the molecular masses of a fraction of peptides from a batch digestion can uniquely identify the protein origin [4, 8, 15-18]. Fourier transform ion cyclotron resonance mass spectrometer (FTICR/MS) is typically used in accurate mass measurement for protein identification due to its high resolution and mass accuracy. It has been reported that a few part-per-million (ppm) of mass accuracy can be routinely obtained by using external calibration [19, 20] for FTICR/MS, and that sub ppm of mass accuracy can be achieved with internal calibration [18, 19, 21-24] or by regulating the ion populations in the analyzer cell using automatic gain control (AGC) [25]. Smith and coworkers have developed an approach for high throughput

proteomics which utilizes accurate mass measurement by FTICR/MS in conjunction with accurate elution time measurement by high resolution liquid chromatography [8, 16, 26-29]. This strategy introduces the concept of "accurate mass and time (AMT)" tags [8], which are first established by measuring the mass and elution time for each of the peptides pre-identified using LC-MS/MS from a batch digestion of a biological sample and then validated by accurate mass measurement using LC-FTICR/MS. These validated AMT tags constitute a reference "look-up" table for subsequent peptide identification. In this approach, LC-MS/MS analysis only needs to be performed once to create AMT tags for a particular biological system, thus the analysis time is significantly reduced. Accurate mass measurement has many advantages over the tandem MS approach in that it is a high throughput method, and that it provides better sensitivity, and that the data produced are easily to be interpreted; however, it still has some limitations. The specificity of protein identification based on this strategy is lower than that using tandem MS, as tandem MS provides the sequence information in addition to the mass of peptide. Previous work has shown that fragmentation of a single peptide is sufficient to accurately assign its identity to the parent protein using tandem MS [4, 30-32], whereas three to six peptide masses are necessary to make an unambiguous protein identification when using accurate mass measurement [4]. Thus, to improve the confidence of protein identification for accurate mass measurement, more constraints should be considered in database searches. This has been demonstrated by a number of publications in which the isoelectric point of peptide [33], retention time [34, 35], presence of cysteine [36], or partial sequence information [37, 38] was used as supplementary information to the mass of peptide.

AMT has been used for proteomic analysis of very complex samples such as proteome sample form human [39, 40]. Studies suggest a need for a higher throughput method for

bacterial/archeal systems which have lower complexity and less post-translational modifications (PTMs), so the large effort investment in constructing AMT library can be omitted. It is known that the mass difference between $^{14}N/^{15}N$ peak pair indicates the number of nitrogen atoms present in the peptide [41, 42], so the nitrogen stoichiometry of peptides can be served as a searching constraint for peptide identification. Here we present the results of protein identification and quantitation using an accurate mass HPLC-MALDI-FTICR/MS approach combined with nitrogen stoichiometry. First, we demonstrate the utility of nitrogen stoichiometry for protein identification by using nitrogen stoichiometry as an additional search constraint. Secondly, we compare three calibration methods (external calibration, stepwise-external calibration and internal calibration) in proteomic measurements and demonstrate that protein identification is greatly enhanced with better mass measurement accuracy when using stepwise-external calibration. Finally, we present the quantitative results and discuss the reliability of this approach by correlating MALDI-MS results with ESI-MS/MS results.

## EXPERIMENTAL

### Materials

Ammonium bicarbonate, chicken egg albumin and insulin chain B were purchased from Sigma (St. Louis, MO). Tris (2-carboxyethyl)-phosphine hydrochloride (TCEP) was purchased from Pierce (Rockford, IL). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI). 2,5-Dihydroxybenzoic acid (DHB) was purchased from Lancaster Synthesis (Pelham, NH). Acetonitrile (HPLC grade) and trifluoroacetic acid (TFA) were purchased from Fisher Scientific (Fair Lawn, NJ). Water used in the experiments was purified using a NanopureInfinity ultrapure water system (Barnstead/Thermolyne, Dubuque, IA).

**Sample Preparation**

The proteome sample was a whole cell lysate from *Methanococcus maripaludis* that was grown on minimal media with ammonium sulfate as the sole source of nitrogen. Cells of *M. maripaludis* Δ*lrp* mutant (S102) and wild-type (S2) were grown using ammonium sulfate with naturally occurring isotopic composition (99.6% $^{14}N$) and with 99% $^{15}N$-enriched composition, respectively. The cells were collected by centrifugation at 10,000 *g* for 10 min. The cell pellets were resuspended in 0.1 M ammonium bicarbonate and lysed with three cycles of freeze-thaw. Phenylmethanesulphonyl fluoride (PMSF) was added to the cell lysates at a final concentration of 1 mM to protect protein lysis. DNA was digested and removed from the extract by incubation with DNAase. Unbroken cells were removed by centrifugation. Protein concentrations were determined spectrophotometrically at 562 nm using a bicinchoninic acid protein assay kit (Pierce, Rockford, IL). Equal amounts of protein extracts from the mutant and wild-type cultures were mixed together before batch trypsinolysis. Small molecules were removed from proteome sample using a centrifugal filter with 3,000 MWCO (Millipore Corporation, Bedford, MA). The sample was heat denatured at 95 °C for 10 min and reduced by TCEP. Denatured sample was digested overnight at 37 °C using trypsin at a 1:50 protease : protein ratio (by mass). The resulting peptide mixtures were separated by reversed-phase HPLC and analyzed by MALDI-FTICR mass spectrometry.

**High Performance Liquid Chromatography**

Reversed-phase high performance liquid chromatography (RP-HPLC) separation of peptide mixtures was performed on an UltiMate Plus system (Dionex LC Packings, Sunnyvale, CA), using a 75 μm i.d. × 15 cm C18 column with 3 μm particles and 300 Å pore size (Dionex LC

Packings, Sunnyvale, CA). Mobile phase A and B were 95/5/0.1 and 20/80/0.1 water/acetonitrile/trifluoroacetic acid (by volume), respectively. A 120 min gradient was used to separate peptide mixtures at a flow rate of 300 nL/min. Mobile phase B started at 0 % and was ramped to 50 % at 90 min, 100 % at 100 min, and kept 100 % for 15 min. The eluate was collected onto two stainless steel MALDI targets at 60-s intervals using a Probot Micro Fraction Collector (Dionex LC Packings, Sunnyvale, CA), resulting in total 90 fractions. Samples were allowed to dry and then added with 0.5 μL of the MALDI matrix solution, which was prepared by dissolving 10 mg of DHB in 50 μL of 50/50/0.1 water/acetonitrile/trifluoroacetic acid (by volume).

**MALDI-FTICR Mass Spectrometry**

Mass spectra were collected on a 7 tesla Bio-Apex Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with an intermediate pressure Scout100 MALDI source. A nitrogen laser was employed and the mass spectra were recorded in the positive mode with the sum of 96 laser shots. Ions generated from MALDI source were accumulated in a hexapole ion guide and then directed to the FTICR analyzer cell through a series of electrostatic ion optics. Ions were excited using a stored waveform inverse Fourier transform (SWIFT) waveform [43] which was produced by a PXI chasis with a PXI-8184 embedded controller and a PXI-5412 arbitrary waveform generator (National Instruments, Austin, TX). The SWIFT excitation was designed to have nearly uniform excitation power over the frequency range of 10,752 to 215,283 (corresponding to *m/z* 10,000 to 500) with a signal duration of 1.966 ms. The optimal exaction voltage was 0.3 Vp-p amplitude.

After ion excitation, a 512 K point transient was acquired, apodized with a sinebell function, and padded with one zero-fill before fast Fourier transformation and magnitude calculation.

## Calibration Methods

### *Method 1. External Calibration*

External calibration was performed through Bruker XMASS 7.0.8 software (Bruker Daltonics, Billerica, MA) using a peptide mixture of insulin chain B and the tryptic digest of chicken egg albumin via eq 1 ('CAL2' calibration equation [44]). The calibration constants generated from XMASS, A and B, are used to convert *m/z* values into frequencies which are used later in the process of stepwise-external calibration and internal calibration.

### *Method 2. Stepwise-External Calibration*

Stepwise-external calibration [1] was accomplished by recording a mass spectrum at a low cell trapping potential (0.60 V), where the ion capacity of the cell is rather low, but the mass accuracy using external calibration is quite good. The trapping potential was then raised to 1.10 V, where the ion capacity of the cell increases by a factor of 5-10, and the mass spectrum was recorded again for the same sample. The fairly accurate mass values from the low trapping potential spectrum were used to calibrate the masses in the second spectrum via eq 2, using the software developed in our laboratory (available upon requested). The calibration constants, A, B and C, were determined by multi-linear fitting the measured frequencies and intensities of the peaks that appear in both mass spectra to the masses measured at low trapping potential. A minimum of four calibrant peaks with a signal-to-noise ratio (S/N) > 5 were used for each mass spectrum.

$$\frac{m}{z} = \frac{A}{f + B} \quad (1)$$

$$\left(\frac{m}{z}\right)_i = \frac{A}{f_i + B + C \cdot I_i} \quad (2)$$

***Method 3. Internal Calibration***

Internal calibration was performed by using the peptides identified from MALDI-FTICR/MS analysis of the proteome sample as internal calibrants. The theoretical mass values of the peptides identified within10 ppm at high trapping potential (1.10 V) were used to calibrate the measured masses in the original spectrum using eq 2. Multi-linear regression was used to determine the calibration constants, A, B and C, by fitting the measured frequencies and intensities of the identified peptides to their theoretical masses. Again, a minimum of four calibrant points were used for each mass spectrum.

**Protein Identification**

The molecular weight of the peptides and their nitrogen stoichiometry were determined by using MALDI-FTICR/MS. The number of nitrogen atoms present in each peptide was calculated from the mass spacing between the monoisotopic peaks of the peptide containing natural isotopic composition and its $^{15}$N-enriched counterpart. The data were processed using the in-house developed software to identify the proteins from which the peptides 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 from the proteins of *M. maripaludis*, including all possible peptides with one missed cleavage, and two missed cleavages if two basic residues (lysine or arginine) are next to each other. A peptide is

considered to be identified when only one predicted tryptic peptide has a mass that lies within a specified mass tolerance of the measured molecular weight and has same nitrogen stoichiometry as the measured value. We define the peptides that match only one entry in the database as "unique" peptides, and ones that match more than one entry as "non-unique" peptides. A protein is considered to be identified when at least one "unique" peptide is observed from that protein.

**Protein Quantitation**

The intensity ratio for each of $^{14}N/^{15}N$ peak pairs was normalized based on the average intensity ratio calculated from all the peptides identified in each measurement. The data from three measurements were combined to determine the differential protein levels. The average intensity ratio for each protein was calculated based on the normalized ratios for all the peptide pairs generated from that protein. The number of peptide pairs per ORF (n), standard deviation of ratio (S.D.) and the confidence level of 95 % were used as a measure of reliability and analytical precision. Differential protein levels are regarded as significant if n is equal to or greater than 2 and if the average intensity ratio ($^{14}N/^{15}N$, mutant/wild) for a given protein is above 1.5 (up-regulation) or below 0.67 (down-regulation).

**RESULTS AND DISCUSSION**

**Enhanced Protein Identification from Nitrogen Stoichiometry**

$^{15}N$-metabolic labeling is well established as a tool for quantitative proteomics. Smith and coworkers suggested that nitrogen stoichiometry can be derived from the mass spacing between pairs of unlabeled/labeled tryptic peptides [41], and this has been demonstrated by Parks [45-47] and Snijders *et al*. [48]. By mixing labeled and unlabeled proteins, batch digestion yields a

mixture of peptide pairs that differ in mass by a variable amount, equal to the number of nitrogen atoms present in the molecular formula. We demonstrate below the utility of nitrogen stoichiometry in protein identification when combined with accurate mass measurement using FTICR mass spectrometer. The archaeon *Methanococcus maripaludis*, with 1722 open reading frames (ORFs), is $^{15}N$-metabolically labeled and used as a model organism to test this approach. The number of nitrogen atoms present in each peptide is calculated by dividing the mass difference between $^{14}N/^{15}N$ peak pairs by 0.997 amu (the mass difference between $^{14}N$ and $^{15}N$), and this number is used as a constraint in database searching in order to improve the specificity of peptide assignments.

To examine the usefulness of $^{15}N$-metabolic labeling in peptide assignments by accurate mass measurement, we performed the statistical analysis of the predicted tryptic peptides with up to 1 missed cleavage from *M. maripaludis* at different level of mass accuracy. It was found that 15 % of the predicted peptides could be identified by their molecular weight alone at 10 ppm search tolerance, and that this identification rate improved to 43 % when nitrogen stoichiometry was used as an additional search constraint. Figure 4.1 compares the specificity of peptide molecular weight for identifying the proteins from which the peptides derived at a search tolerance of 10 ppm, 5 ppm and 1 ppm. With 5 ppm mass accuracy, 33 % of the peptides correspond to a single protein when only using their molecular weight as a search constraint, increasing to 55 % when nitrogen stoichiometry is used. At a search tolerance of 1 ppm, the specificity of peptide assignments improves to 62 % when nitrogen stoichiometry is used as an additional search constraint. Although the improvement in peptide assignments using nitrogen stoichiometry at 1 ppm is not very significant compared to that at 5 ppm and 10 ppm, $^{15}N$-metabolic labeling has other important practical advantages. This method allows one to

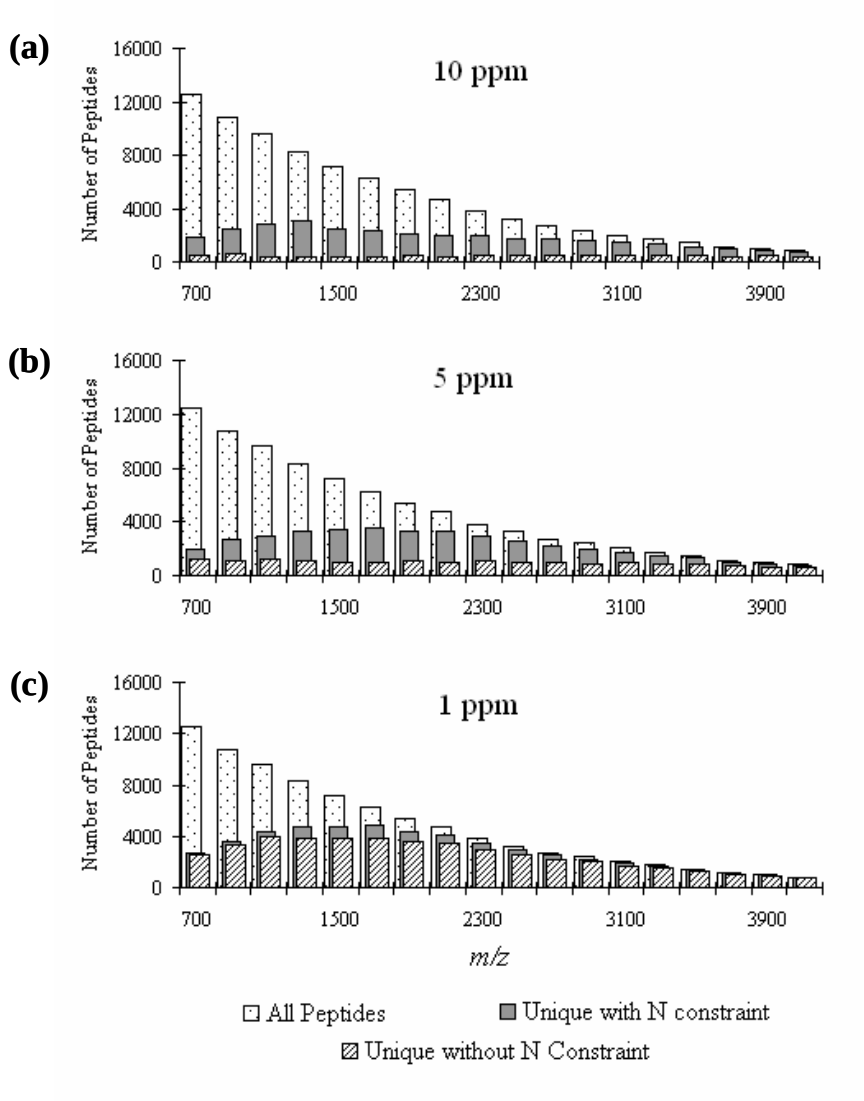



**Figure 4.1** Specificity of peptide assignments from *in silico* tryptic digestion of the proteins from *M. maripaludis* over the range 700-4100 Da using (a) 10 ppm, (b) 5 ppm and (c) 1 ppm as search tolerance. The bars filled with dots represent the total number of peptides produced for each 200 amu mass bin. The grey bars represent the number of peptides that are identified by their molecular weight alone at a given search tolerance. The bars filled with up diagonals represent the number of peptides identified using both molecular weight and nitrogen stoichiometry of the peptides.

distinguish the proteins that are produced by the organism being examined from exogenous proteins (e.g. keratins, trypsin autolysis products) as the latter do not produce pairs of peaks. More significantly, $^{15}N$ metabolic labeling allows one to quantitatively determine changes in protein expression.

In our work, the $^{15}N$-labeled proteome sample was initially analyzed by an offline HPLC-MALDI-FTICR/MS approach using external calibration (The other two calibration methods are used and discussed below). Protein identifications were made by searching against the database using a mass tolerance of 10 ppm. Triplicate analysis was performed for the same sample in order to evaluate the reliability of the approach used. Unless specified, the results shown are from the combined data sets. Using the molecular weight of peptides as search constraint, 1,132 out of the 18,274 detected peptides are uniquely assigned to a single protein, leading to 280 protein identifications, 16,824 of them correspond to two or more proteins, and the remaining 318 peptides do not match any predicted peptides from *M. maripaludis*. In contrast, when the nitrogen stoichiometry is used as an additional search constraint, 5,592 peptides are uniquely assigned, about four-fold increase in the number of identified peptides compared to that using molecular weight alone, leading to 659 protein identifications. These increased numbers indicate that combining nitrogen stoichiometry with accurate mass measurement can greatly improve the specificity of peptide assignments and therefore enhance the protein identification.

**Improved Protein Identification by Higher Mass Measurement Accuracy**

Accurate mass measurement of proteolytic peptides by FTICR mass spectrometry offers a means for rapidly identifying the components of a proteome sample. In this method, protein identification largely depends on the mass accuracy that can be achieved under the experimental

conditions. As shown in Figure 4.1, the identification rate of peptides is predicted to increase when the nitrogen stoichiometry is used as an additional constraint at a specified search tolerance, and it also improves with better mass accuracy. We use HPLC-MALDI-FTICR/MS as primary means to perform proteomic analysis, and each HPLC separation produces around 90 factions to be analyzed. As it is impractical to add an internal calibrant to this large number of samples, it is important to be able to use external calibration and still achieve high mass accuracy. It is well known that MALDI is characterized by substantial shot-to-shot variations in ion intensity, and such variations are known to lead to space-charge induced frequency shifts when using external calibration [19]. For this reason, it is more challenging to achieve low ppm mass accuracy in FTICR/MS by MALDI than by ESI, where the ion population can be controlled, e.g. by automatic gain control [25]. Based on these considerations, we have developed an external calibration procedure for MALDI-FTICR/MS, stepwise-external calibration, which produces a mass accuracy close to that of internal calibration for the peptides generated from protein standards [1]. In addition, we have developed a calibration equation that takes into account both local and global space-charge effects [1]. By combining this procedure with the new calibration equation, we have achieved a RMS error of 0.9 ppm for MALDI-FTICR/MS analyses of the peptides up to *m/z* 2500 [1]. Recently, we have reported that a RMS error of 0.95 ppm can be achieved for the higher mass peptides (up to m/z 4000) when combining stepwise-external calibration with SWIFT excitation [2]. Here we present the results by extending these findings to proteomic measurements using HPLC-MALDI-FTICR/MS.

### *I. Stepwise-External Calibration vs. External Calibration*

Given the above-mentioned findings for protein standards, stepwise-external calibration was applied to a proteome sample in order to examine its usefulness for a more complicated system. To evaluate the performance of stepwise-external calibration, the mass error distribution of the peptides identified within 10 ppm was examined and compared with that using conventional external calibration. Figure 4.2 shows the distribution of mass errors from one measurement based on 1,079 and 1,082 peptides for stepwise-external calibration and external calibration, respectively. The stepwise-external calibration yields -0.077 ± 2.75 ppm (average error ± standard deviation), which is much better than that obtained by using external calibration, 0.99 ± 2.96 ppm. As we can see from Figure 4.2, the distribution of mass errors becomes narrower after applying stepwise-external calibration, and the centroid of the distribution shifts from 1.0 ppm to nearly 0 ppm, indicating that the space-charge effects are significantly reduced.

### *II. Stepwise-External Calibration vs. Internal Calibration*

Cramer and coworkers have developed an automatic internal calibration method for LC-ESI-FTICR/MS analysis of whole-cell digests using the tryptic peptides identified by concurrent tandem mass spectroscopy as internal calibrants, providing a few ppm mass measurement accuracy [18]. Here we employed the same strategy to internally calibrate the peptide masses which were measured at high trapping potential (1.10 V), where the mass accuracy is relatively poor. Some modifications were made in our work to meet the experimental requirements. First, the peptides used as internal calibrants are those identified by LC-MALDI-FTICR/MS using 10 ppm mass tolerance constrained by nitrogen stoichiometry in database searching. Secondly, rather than using eq 1 in Cramer's work, we applied eq 2 in calibration as the latter provides

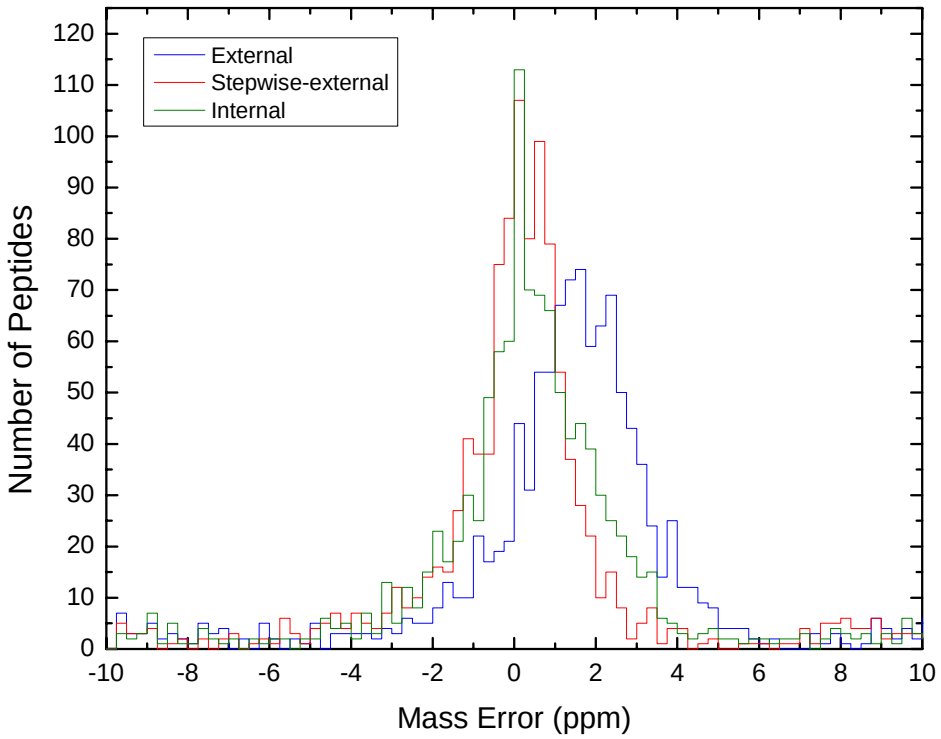



**Figure 4.2** Distribution of mass measurement errors of the peptides identified from a *M. maripaludis* proteome sample using 10 ppm search tolerance constrained by nitrogen stoichiometry by applying external calibration (blue), stepwise-external calibration (red) and internal calibration (green), with a mass error bin of 0.25 ppm. The average mass errors produced by using external calibration, stepwise-external calibration and internal calibration, are 0.99 ± 2.96 ppm, -0.077 ± 2.75 ppm and 0.12 ± 2.77 ppm, respectively.

better mass accuracy due to the fact that local space-charge effects are taken into account [1]. After applying internal calibration on the same data, mass error analysis was performed based on 1,092 identified peptides, showing a similar distribution to that using stepwise-external calibration (Figure 4.2). The distribution of mass measurement errors produced by internal calibration exhibit 0.12 ± 2.8 ppm, which is close to that using stepwise-external calibration. Both methods effectively reduce the mass measurement errors compared to external calibration; however, stepwise-external calibration has some advantages over internal calibration. First, any peaks in the low trapping potential mass spectrum could be used as reference points for stepwise-external calibration. This allows for achieving better mass accuracy as more data points are used during multi-linear regression. In contrast, there is only limited number of reference points that can be used in internal calibration, as only the peptides identified from the initial measurement can be used as calibrants. Figure 4.3 illustrates the distribution of the number of calibrants for each mass spectrum during HPLC separation for both calibration methods. As expected, there are more calibrant points used in stepwise-external calibration than internal calibration for each mass spectrum. It is also noted that 3 out of 90 mass spectra can not be internally calibrated due to the insufficient number of calibrants. This primarily happens at the beginning or the end of the separation where only a few peptides are eluted. Another advantage of stepwise-external calibration is that it is especially useful to analyze a biological sample with unknown origin, as peptide assignments would become more difficult for such sample, resulting in an even lower number of calibrants that can be used for internal calibration.

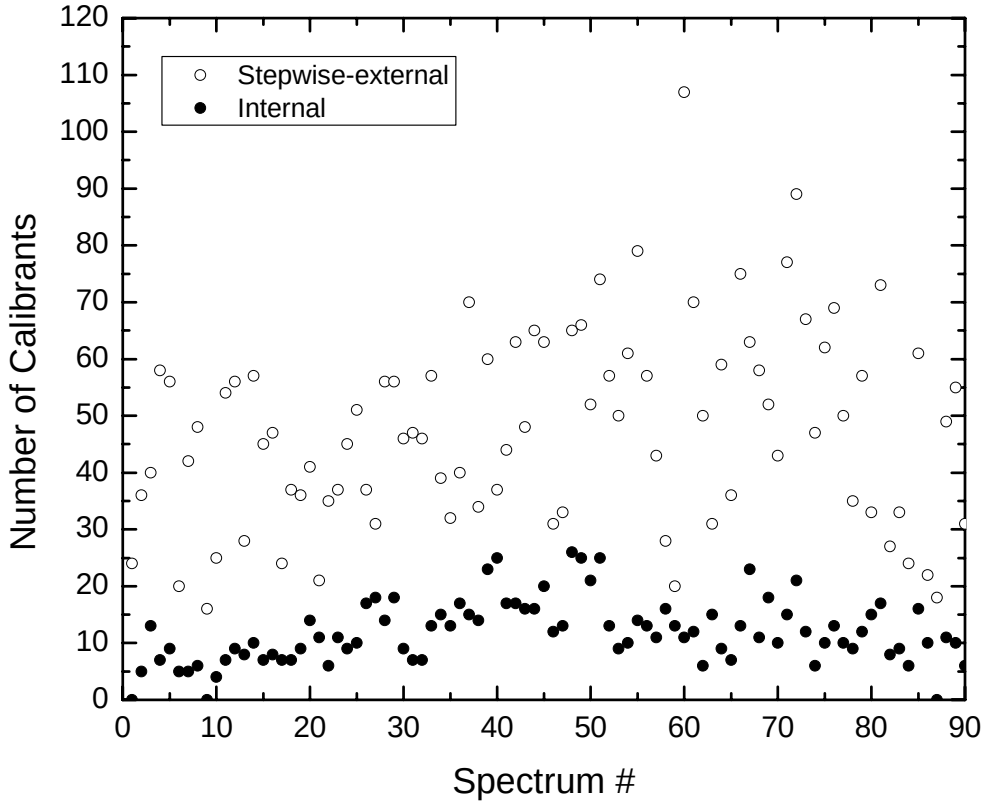



**Figure 4.3** Distribution of the number of calibrants found during the liquid chromatographic run for stepwise-external calibration (open circles) and for internal calibration (filled circles).

### *III. Protein Identification*

We have previously performed similar error analysis for measurements of peptides from protein standards using stepwise-external calibration, and found that the error distribution is a Gaussian fashion [1]. In contrast, here we find for proteomic data that there is a central Gaussian-like distribution, but also there is a low rate of much larger errors on the wings of the main distribution. These outliers suggest that some of the "unique" assignments are false positives. This could be explained as resulting from an incomplete set of entries in the list of predicted peptides. Another evidence that indicates the incompletion of our list of potential peptides is that 15 ~ 20 % of the detected peptides are found to be unassignable, i.e. their mass and nitrogen stoichiometry match none of the entries in the database. The unassigned peptides found in our experiments exhibit $^{14}N/^{15}N$ peak pairs in mass spectrum, suggesting that they derive from the proteins produced by the organism analyzed. The unassigned peptides could arise from biological processes, such as post-translational modification or mistranslation of the protein. However the related organism *M. janaschii* is known to have a low rate of post-translational modification [49], and *M. maripaludis* is expected to behave the same. Thus, it would be difficult to explain the high rate of unassigned peptides based on post-translational modification alone. The unassigned peaks may also be artifacts from sample treatment or from the analysis procedure, such as deamidation of asparagine or glutamine residues, oxidation of methionine, nonspecific trypsin digestion or metastable decomposition of the peptides by MALDI. The determination of the origin of the unassigned peptides is currently being performed in our laboratory by using tandem mass spectrometry.

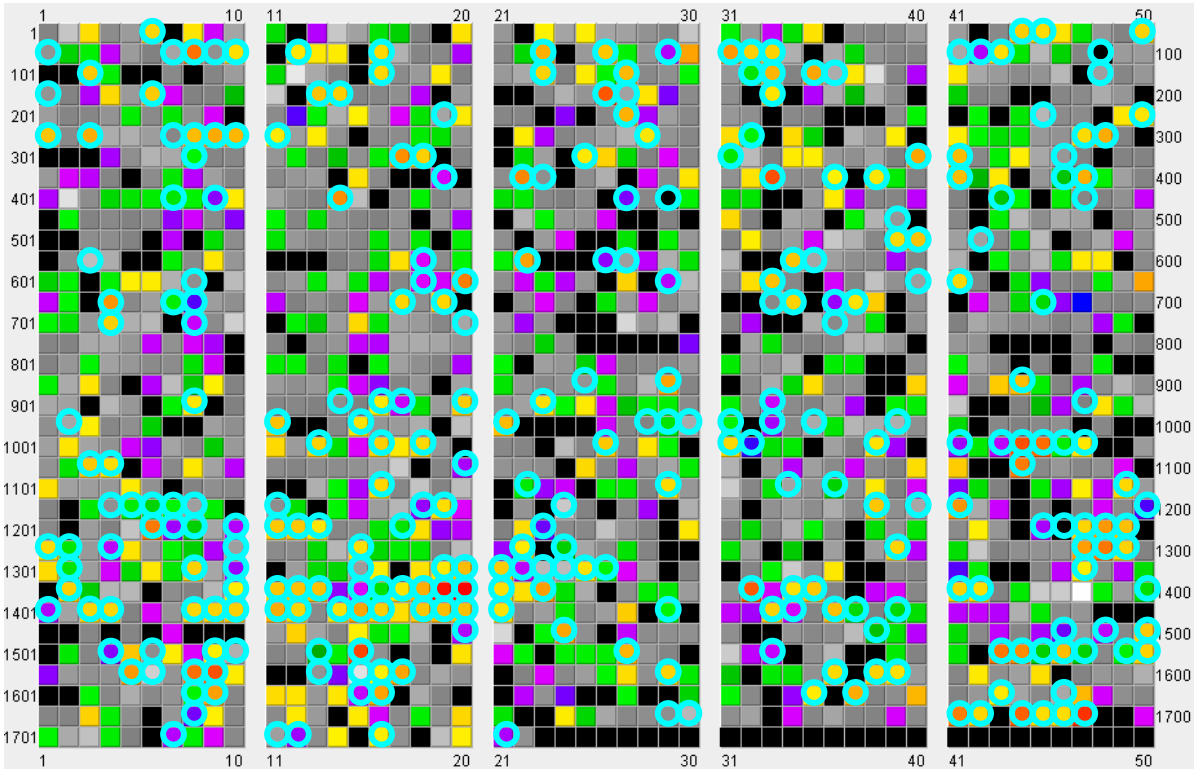
To reduce the rate of false positives, a smaller search tolerance, close to the actual mass accuracy of the measurement, should be used for peptide assignments. This also allows us to

more accurately evaluate the performance of stepwise-external calibration as the most of the misassigned peptides can be excluded from the error analysis. Thus, we calculated the average RMS error from three measurements based only on the peptides that fall within the main distribution (± 5 ppm). As a result, stepwise-external calibration yields a RMS error of 1.5 ppm for proteomic data. This level of mass accuracy allows us to perform database searches at a mass tolerance of 5 ppm (3 times the RMS error of 1.5 ppm) with 99.7 % confidence in the accuracy of the masses. Using 5 ppm mass tolerance combined with nitrogen stoichiometry, 7,154 out of the 18,274 detected peptide are uniquely assigned, leading to 680 protein identifications, a significant improvement in protein identification compared to that using external calibration. Figure 4.4 shows the ORF map of *M. maripaludis* in which the identified proteins are denoted by the orange, green and purple squares and numbered according to the occurrence of ORFs in the sequence database. Green square, purple square and orange square represent the protein identified from unique peptides in one of MALDI-MS experiments, in two of MALDI-MS experiments and in all three MALDI-MS experiments, respectively.

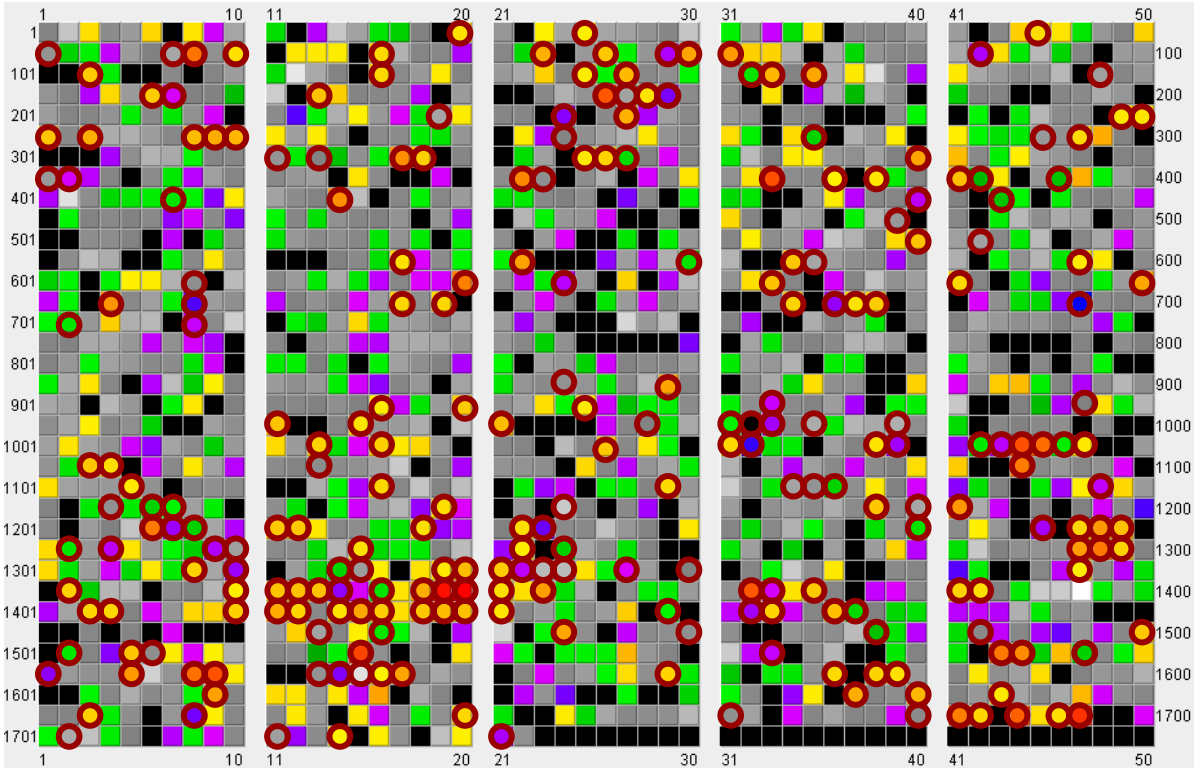
**Protein Quantitation**

$^{15}$N-metabolic labeling provides a means to make quantitative measurements of the changes in protein expression based on the intensity ratios of $^{14}$N/$^{15}$N peak pairs. With 5 ppm mass accuracy, 680 proteins are identified, representing 40 % of the coding capacity of the genome of *M. maripaludis*. Table 4.1 summarizes the genes that show statistically significant expression changes between mutant (S102) and wild-type (S2). 26 out of the 680 proteins are up-regulated, 33 of them are down-regulated, and the remainder lacks statistical support for regulation. We find a number of coregulated proteins which are expected to be expressed at the same level. For

**(a)**


1
10
11
20
21
30
31
40
41
50
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301
1401
1501
1601
1701
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700

**(b)**


1
10
11
20
21
30
31
40
41
50
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301
1401
1501
1601
1701
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700

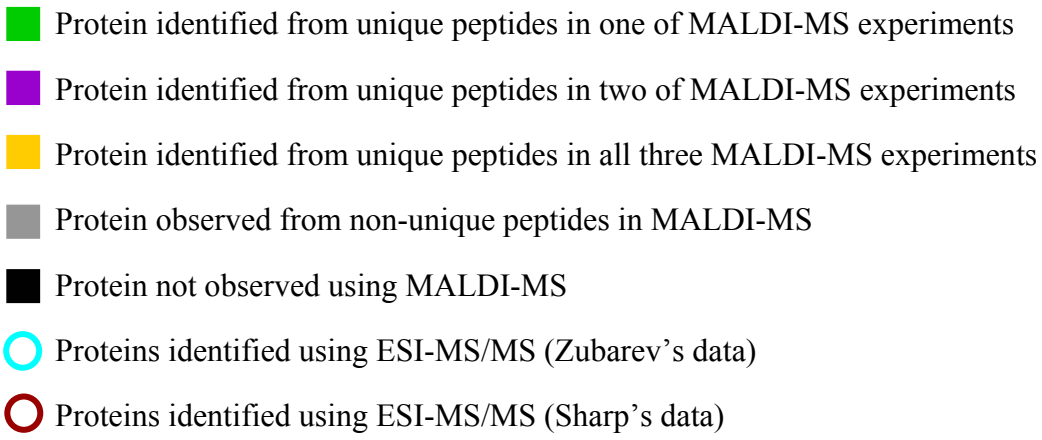



**Figure 4.4** Map of protein assignments from triplicate analysis of a *M. maripaludis* proteome sample using MALDI-FTICR/MS with 5 ppm mass tolerance constrained by nitrogen stoichiometry in database searching (squares) and ESI-MS/MS analysis of the same sample using ESI-MS/MS with 10 ppm peptide mass tolerance and 0.02 Da fragment mass tolerance in database searching (blue circles: Zubarev's data; brown circles: Sharp's data). Each square or circle represents a protein, numbered according to the occurrence of the open reading frames (ORFs) in the sequence database. MALDI-FTICR/MS, Zubarev's ESI-MS/MS and Sharp's ESI-MS/MS experiments respectively yield 680 (orange/green/purple squares), 299 (blue open circles) and 255 (brown open circles) protein identifications.

**Table 4.1** Expression of the regulated genes by shotgun proteomic analyses using HPLC-MALDI-FTICR/MS.

| ORF | Description[1] | Average $^{14}N/^{15}N$ Intensity Ratio[2] | n | SD |
|---|---|---|---|---|
| MMP0020 | Conserved hypothetical protein | 0.53 | 3 | 0.07 |
| MMP0058 | Coenzyme F420-dependent N5,N10-methylenetetrahydromethanopterin reductase | **1.67** | **74** | **0.26** |
| MMP0125 | Conserved archaeal fibrillarin homolog | **1.69** | **7** | **0.25** |
| MMP0209 | Transcriptional regulator | **1.84** | **9** | **0.21** |
| MMP0215 | Hypothetical protein | **1.97** | **6** | **1.14** |
| MMP0253 | Acetyl-CoA synthetase (ADP-forming), alpha and beta subunits | 0.58 | 17 | 0.11 |
| MMP0291 | AIR synthase related protein Archaea | 0.52 | 3 | 0.05 |
| MMP0358 | Conserved hypothetical archaeal protein | 0.64 | 2 | 0.16 |
| MMP0369 | Conserved hypothetical protein | 0.64 | 2 | 0.04 |
| MMP0386 | Archaeal histone B | 0.63 | 59 | 0.09 |
| MMP0443 | SSU ribosomal protein | 0.58 | 3 | 0.09 |
| MMP0507 | Molybdate-binding periplasmic protein | **1.67** | **2** | **0.40** |
| MMP0618 | Conserved hypothetical protein | 0.34 | 3 | 0.07 |
| MMP0645 | Malate dehydrogenase, MDHII (NADP+-dependent) | 0.50 | 6 | 0.10 |
| MMP0648 | RNA 3'-terminal phosphate cyclase | 0.61 | 2 | 0.13 |
| MMP0686 | Undefined product | 0.56 | 4 | 0.16 |
| MMP0692 | Conserved hypothetical protein | 0.58 | 3 | 0.11 |
| MMP0866 | Putative glycine betaine/L-proline ABC transporter, solute-binding protein | 0.54 | 8 | 0.10 |
| MMP0885 | TPR repeat:ATP/GTP-binding site motif A (P-loop) | **1.75** | **15** | **0.35** |
| MMP0961 | Conserved hypothetical protein | 0.58 | 8 | 0.10 |
| MMP0965 | Formylmethanofuran dehydrogenase subunit E related protein | 0.48 | 12 | 0.17 |

| | | | | |
|---|---|---|---|---|
| MMP0976 | Protein of unknown function DUF166 | 0.40 | 4 | 0.08 |
| MMP1016 | Conserved Hypothetical protein with 4 CBS domains. | 0.57 | 19 | 0.17 |
| MMP1028 | Undefined product | **1.54** | **6** | **0.35** |
| MMP1032 | OB-fold nucleic acid binding domain | **1.69** | **3** | **0.34** |
| MMP1044 | A1A0 ATPase, subunit A | 0.64 | 80 | 0.15 |
| MMP1045 | A1A0 ATPase, subunit B | 0.60 | 85 | 0.11 |
| MMP1077 | Phosphoglucomutase/phosphomannomutase:Calcium-binding EF-hand | 0.53 | 2 | 0.10 |
| MMP1123 | Conserved hypothetical protein | 0.63 | 3 | 0.12 |
| MMP1124 | AIR synthase related protein | **1.85** | **2** | **0.15** |
| MMP1138 | Hydroxyethylthiazole kinase family | **2.68** | **7** | **0.55** |
| MMP1208 | Gamma translation initiation factor aIF-2, subunit gamma | 0.66 | 4 | 0.10 |
| MMP1211 | Undefined product | 0.50 | 13 | 0.20 |
| MMP1212 | Undefined product | 0.41 | 15 | 0.09 |
| MMP1213 | Undefined product | 0.44 | 4 | 0.04 |
| MMP1271 | 2-oxoisovalerate oxidoreductase subunit alpha | 0.61 | 4 | 0.14 |
| MMP1313 | 5'3'-Exonuclease N- and I-domains:DNA repair protein (XPGC)/yeast Rad:Helix-hairpin-helix DNA-binding, class 2 | **1.60** | **2** | **0.11** |
| MMP1341 | ATP/GTP-binding site motif A (P-loop):SMC protein, N terminal:ABC transporter | **1.69** | **2** | **0.23** |
| MMP1360 | RNA polymerase H/23 kD subunit | 0.63 | 5 | 0.07 |
| MMP1365 | Ribosomal protein | 0.67 | 4 | 0.14 |
| MMP1366 | KH domain:KH domain, type 1 | 0.65 | 2 | 0.06 |
| MMP1367 | SSU ribosomal protein S12 | 0.64 | 4 | 0.12 |
| MMP1372 | Phosphoglucomutase/phosphomannomutase | **1.53** | **4** | **0.38** |
| MMP1382 | Coenzyme F420-reducing hydrogenase subunit alpha | **1.63** | **104** | **0.34** |
| MMP1383 | Coenzyme F420-reducing hydrogenase delta subunit | **1.94** | **3** | **0.20** |
| MMP1385 | Coenzyme F420-reducing hydrogenase subunit beta | **1.64** | **13** | **0.32** |
| MMP1398 | Succinyl-diaminopimelate desuccinylase | 0.54 | 2 | 0.07 |
| MMP1457 | Energy conserving hydrogenase A integral membrane | **1.53** | **10** | **0.29** |

| | | | | |
|---|---|---|---|---|
| | subunit | | | |
| MMP1466 | Conserved hypothetical protein | **1.96** | **2** | **0.44** |
| MMP1477 | Cobyrinic acid a,c-diamide synthase | **1.51** | **5** | **0.20** |
| MMP1520 | ATP/GTP-binding site motif A (P-loop):HypB/UreG, nucleotide-binding:Hydrogenase accessory protein HypB | **1.61** | **6** | **0.30** |
| MMP1521 | Conserved hypothetical archaeal protein | **2.64** | **4** | **1.20** |
| MMP1559 | Methyl-coenzyme M reductase I, alpha subunit | **1.50** | **143** | **0.29** |
| MMP1569 | Conserved hypothetical archaeal protein | 0.64 | 3 | 0.04 |
| MMP1624 | Polyferredoxin | **1.72** | **4** | **0.66** |
| MMP1634 | Protein of unknown function UPF0033:DsrE-like protein | **1.51** | **5** | **0.16** |
| MMP1662 | Precorrin-4 C11-methyltransferase | **1.57** | **6** | **0.22** |
| MMP1692 | Polyferredoxin, associated with F420-non-reducing hydrogenase | **1.51** | **19** | **0.24** |
| MMP1714 | Conserved hypothetical protein | 0.65 | 4 | 0.19 |

[1] The ORF description is derived from the genome annotation.

[2] $^{14}N/^{15}N$ intensity ratios represent the average S102/S2 (mutant/wild) ratios. Bold indicates significantly higher expression in the mutant; unbold indicates lower expression.

example, two proteins involved in ATP synthesis, ORFs 1044 and 1045, are down-regulated. The proteins involved in methanogenesis, ORFs 1382, 1383 and 1385, are up-regulated. All these proteins play an important role in energy production. In addition, a group of ribosomal proteins, ORFs 1365-1367, and the hypothetical proteins, ORFs 1211-1213, are down-regulated, but the function of the latter is unknown.

**Correlation of MALDI-MS Results with ESI-MS/MS Results**

Accurate mass MALDI-FTICR/MS combined with nitrogen stoichiometry has been under development for years in our laboratory as the primary analytical tool in proteomic analysis, and its utilities for protein identification and quantitation have been demonstrated with various biological samples for different analytical purposes. In order to evaluate the reliability and effectiveness of this approach, our results are correlated with ESI-MS/MS analysis for the same sample from Zubarev's group and Sharp. For Zubarev's data, collisionally induced dissociation and electron capture dissociation were used for protein identification by searching against the NCBInr database using Mascot search engine with 10 ppm mass tolerance, and yields 299 protein identifications (Figure 4.4a). Among these proteins, 240 of them have been identified, 56 of them have been observed from the non-unique peptides, and 3 proteins have never been found using our approach. For Sharp's data, proteins were identified by using collisionally induced dissociation and *M. maripaludis* S2 database were used in Mascot. 255 proteins were identified with 10 ppm mass tolerance (Figure 4.4b). Among these proteins, 207 of them have been identified, 47 of them have been observed from the non-unique peptides, and only 1 protein has never been found using our approach. The high overlapping in protein identification in this interlaboratory comparison studies suggests that accurate mass MALDI-FTICR/MS combined

with nitrogen stoichiometry provides a reliable means for proteomic measurements. Moreover, compared to ESI-MS/MS approach, this technique appears to be more attractive as the analysis time is greatly reduced by obviating the tedious tandem mass spectrometric analysis.

## CONCLUSIONS

The method presented here provides a means for improving the specificity of protein identification in shotgun proteomics by using accurate mass measurement and nitrogen stoichiometry. We have demonstrated that protein identification can be greatly enhanced by $^{15}N$-metabolic labeling when the nitrogen stoichiometry is used as a constraint along with the molecular weight of peptides in database searching. With only external calibration applied, 30.6 % of the peptides observed from a *M. maripaludis* proteome sample are identified within 10 ppm when both constraints are used, a significant improvement compared to 6.2 % when the molecular weight is used alone. We have also shown that protein identification can be improved by using stepwise-external calibration as it significantly improves the mass measurement accuracy. To evaluate the performance of stepwise-external calibration in proteomic measurements, it was compared with the other two calibration methods, external calibration and internal calibration. Statistical mass error analyses show that stepwise-external calibration produces much better mass accuracy than that using external calibration and nearly identical mass accuracy to that using internal calibration. Stepwise-external calibration yields a RMS error of 1.5 ppm, allowing us to perform database searches at a mass tolerance of 5 ppm (three times the RMS error) with 99.7 % confidence in the mass accuracy. With this level of mass accuracy, 7,154 out of the 18,274 detected peptides are assigned, leading to 680 protein identifications. Quantitative measurements of the changes in protein expression were made for the proteome

sample analyzed. By calculating the intensity ratios of $^{14}N/^{15}N$ peptide pairs, 26 out of 680 found proteins are found to be up-regulated, and 33 of them are observed to be down-regulated.

**REFERENCES**


1 Wong, R. L.; Amster, I. J. Sub Part-Per-Million Mass Accuracy by Using Stepwise-External Calibration in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2006**, *17*, 1681-1691.

2 Jing, L.; Li, C. Y.; Wong, R. L.; Kaplan, D. A.; Amster, I. J. Improved Mass Accuracy for Higher Mass Peptides by Using SWIFT Excitation for MALDI-FTICR Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 76-81.

3 Peng, J. M.; Gygi, S. P. Proteomics: The Move to Mixtures. *J. Mass Spectrom.* **2001**, *36*, 1083-1091.

4 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.

5 Gygi, S. P.; Aebersold, R. Mass Spectrometry and Proteomics. *Curr. Opin. Chem. Biol.* **2000**, *4*, 489-494.

6 Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of Two-Dimensional Gel Electrophoresis-Based Proteome Analysis Technology. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9390-9395.

7 Yates, J. R. Mass Spectrometry - from Genomics to Proteomics. *Trends in Genetics* **2000**, *16*, 5-8.

8 Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y. F.; 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.

9 McDonald, W. H.; Yates, J. R. Shotgun Proteomics and Biomarker Discovery. *Disease Markers* **2002**, *18*, 99-105.

10 Lane, C. S. Mass Spectrometry-Based Proteomics in the Life Sciences. *Cmls-Cellular and Molecular Life Sciences* **2005**, *62*, 848-869.

11 Bogdanov, B.; Smith, R. D. Proteomics by FTICR Mass Spectrometry: Top Down and Bottom Up. *Mass Spectrom. Rev.* **2005**, *24*, 168-200.

12 Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001**, *73*, 5683-5690.

13 Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001**, *19*, 242-247.

14 Tabb, D. L.; MacCoss, M. J.; Wu, C. C.; Anderson, S. D.; Yates, J. R. Similarity among Tandem Mass Spectra from Proteomic Experiments: Eetection, Significance, and Utility. *Anal. Chem.* **2003**, *75*, 2470-2477.

15 Clauser, K. R.; Baker, P.; Burlingame, A. L. Role of Accurate Mass Measurement (+/- 10 ppm) in Protein Identification Strategies Employing MS or MS MS and Database Searching. *Anal. Chem.* **1999**, *71*, 2871-2882.

16 Strittmatter, E. F.; Ferguson, P. L.; Tang, K. Q.; Smith, R. D. Proteome Analyses Using Accurate Mass and Elution Time Peptide Tags with Capillary LC Time-of-Flight Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2003**, *14*, 980-991.

17 He, F.; Emmett, M. R.; Hakansson, K.; Hendrickson, C. L.; Marshall, A. G. Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement. *Journal of Proteome Research* **2004**, *3*, 61-67.

18 Palmblad, M.; Bindschedier, L. V.; Gibson, T. M.; Cramer, R. Automatic Internal Calibration in Liquid Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Protein Digests. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3076-3080.

19 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.

20 Bruce, J. E.; Anderson, G. A.; Brands, M. D.; Pasa-Tolic, L.; Smith, R. D. Obtaining More Accurate Fourier Transform Ion Cyclotron Resonance Mass Measurements without Internal Standards Using Multiply Charged Ions. *J. Am. Soc. Mass. Spectrom.* **2000**, *11*, 416-421.

21 Burton, R. D.; Matuszak, K. P.; Watson, C. H.; Eyler, J. R. Exact Mass Measurements Using a 7 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer in a Good Laboratory Practices-Regulated Environment. *J. Am. Soc. Mass. Spectrom.* **1999**, *10*, 1291-1297.

22 Hannis, J. C.; Muddiman, D. C. A Dual Electrospray Ionization Source Combined with Hexapole Accumulation to Achieve High Mass Accuracy of Biopolymers in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2000**, *11*, 876-883.

23 O'Connor, P. B.; Costello, C. E. Internal Calibration on Adjacent Samples (InCAS) with Fourier Transform Mass Spectrometry. *Anal. Chem.* **2000**, *72*, 5881-5885.

24 Mize, T. H.; Amster, I. J. Broad-Band Ion Accumulation with an Internal Source MALDI-FTICR-MS. *Anal. Chem.* **2000**, *72*, 5886-5891.

25 Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-Translational Modifications. *J. of Proteome Res.* **2004**, *3*, 621-626.

26 Qian, W. J.; Camp, D. G.; Smith, R. D. High-Throughput Proteomics Using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Expert Review of Proteomics* **2004**, *1*, 87-95.

27 Jacobs, J. M.; Monroe, M. E.; Qin, W. J.; Shen, Y. F.; Anderson, G. A.; Smith, R. D. Ultra-Sensitive, High Throughput and Quantitative Proteomics Measurements. *Int. J. Mass spectrom.* **2005**, *240*, 195-212.

28 Qian, W. J.; Jacobs, J. M.; Liu, T.; Camp, D. G.; Smith, R. D. Advances and Challenges in Liquid Chromatography-Mass Spectrometry-Based Proteomics Profiling for Clinical Applications. *Mol. Cell. Proteomics* **2006**, *5*, 1727-1744.

29 Zimmer, J. S. D.; Monroe, M. E.; Qian, W. J.; Smith, R. D. Advances in Proteomics Data Analysis and Display Using an Accurate Mass and Time Tag Approach. *Mass Spectrom. Rev.* **2006**, *25*, 450-482.

30 McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R. Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-Femtomole Level. *Anal. Chem.* **1997**, *69*, 767-776.

31 Ducret, A.; Van Oostveen, I.; Eng, J. K.; Yates, J. R.; Aebersold, R. High Throughput Protein Characterization by Automated Reverse-Phase Chromatography Electrospray Tandem Mass Spectrometry. *Protein Sci.* **1998**, *7*, 706-719.

32 Yates, J. R. Mass Spectrometry and the Age of the Proteome. *J. Mass Spectrom.* **1998**, *33*, 1-19.

33 Cargile, B. J.; Stephenson, J. L. An Alternative to Tandem Mass Spectrometry: Isoelectric Point and Accurate Mass for the Identification of Peptides. *Anal. Chem.* **2004**, *76*, 267-275.

34 Palmblad, M.; Ramstrom, M.; Markides, K. E.; Hakansson, P.; Bergquist, J. Prediction of Chromatographic Retention and Protein Identification in Liquid Chromatography/Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 5826-5830.

35 Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Pasa-Tolic, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E. F.; Shen, Y. F.; Zhao, R.; Smith, R. D. Use of Artificial Neural Networks for the Accurate Prediction of Peptide Liquid Chromatography Elution Times in Proteome Analyses. *Anal. Chem.* **2003**, *75*, 1039-1048.

36 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.

37 Green, M. K.; Johnston, M. V.; Larsen, B. S. Mass Accuracy and Sequence Requirements for Protein Database Searching. *Anal. Biochem.* **1999**, *275*, 39-46.

38 Spengler, B. De Novo Sequencing, Peptide Composition Analysis, and Composition-Based Sequencing: A New Strategy Employing Accurate Mass Determination by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2004**, *15*, 703-714.

39 Qian, W.-J.; Monroe, M. E.; Liu, T.; Jacobs, J. M.; Anderson, G. A.; Shen, Y.; Moore, R. J.; Anderson, D. J.; Zhang, R.; Calvano, S. E.; Lowry, S. F.; Xiao, W.; Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Camp, D. G., II; Smith, R. D. Quantitative Proteome Analysis of

Human Plasma Following in Vivo Lipopolysaccharide Administration Using 16O/18O Labeling and the Accurate Mass and Time Tag Approach. *Mol. Cell. Proteomics* **2005**, *4*, 700-709.

40 Diamond, D. L.; Jacobs, J. M.; Paeper, B.; Proll, S. C.; Gritsenko, M. A.; Carithers, R. L.; Larson, A. M.; Yeh, M. M.; Camp, D. G.; Smith, R. D.; Katze, M. G. Proteomic Profiling of Human Liver Biopsies: Hepatitis C Virus-Induced Fibrosis and Mitochondrial Dysfunction. *Hepatology* **2007**, *46*, 649-657.

41 Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Pasa-Tolic, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. Quantitative Analysis of Bacterial and Mammalian Proteomes Using a Combination of Cysteine Affinity Tags and 15N-Metabolic Labeling. *Anal. Chem.* **2001**, *73*, 2132-2139.

42 Jing, L.; Amster, I. J. Rapid and Automated Processing of MALDI-FTICR/MS Data for 15N-Metabolic Labeling in a Shotgun Proteomics Analysis. *Int. J. Mass spectrom.* **2009**, *In press*.

43 Guan, S. H.; Marshall, A. G. Stored Waveform Inverse Fourier Transform (SWIFT) Ion Excitation in Trapped-Ion Mass Spectometry: Theory and Applications. *Int. J. Mass spectrom.* **1996**, *157*, 5-37.

44 Francl, T. J.; Sherman, M. G.; Hunter, R. L.; Locke, M. J.; Bowers, W. D.; McIver, R. T. Experimental-Determination of the Effects of Space-Charge on Ion-Cyclotron Resonance Frequencies. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54*, 189-199.

45 Parks, B. A.; Boltz, S. A.; Porat, I.; Whitman, W. B.; Easterling, M. L.; Speir, J. P.; Amster, I. J. Shotgun Proteomics of *Methanococcus Maripaludis* Using $^{15}$N-Labeling and Accurate Mass Measurement. *51st ASMS Conference on Mass Spectrometry and Allied Topics* **2003**, A031870.

46 Parks, B. A.; Porat, I.; Kim, W.; Whitman, W. B.; Amster, I. J. Shotgun Proteomic Analysis of Membrane Proteins from *Methanococcus Maripulidis* Using Accurate Mass Measurement and $^{15}$N-Metabolic Labeling. *52nd ASMS Conference on Mass Spectrometry and Allied Topics* **2004**, A041808.

47 Parks, B. A.; Ferguson, J. T.; Du, Y.; Burke, P. V.; Kwast, K. E.; Marshall, A. G.; Hendrickson, C. L.; Schaub, T. M.; Kelleher, N. L. High-Throughput Identification of Intact Proteins and Comparative Analysis of *Sacchromyces Cerevisiae* Using $^{14}$N/$^{15}$N Metabolic Labeling and Top-Down Proteomics. *54th ASMS Conference on Mass Spectrometry and Allied Topics* **2006**, A062109.

48 Snijders, A. P. L.; de Vos, M. G. J.; Wright, P. C. Novel Approach for Peptide Quantitation and Sequencing Based on N-15 and C-13 Metabolic Labeling. *J. Proteome Res.* **2005**, *4*, 578-585.

49 Forbes, A. J.; Patrie, S. M.; Taylor, G. K.; Kim, Y. B.; Jiang, L. H.; Kelleher, N. L. Targeted Analysis and Discovery of Posttranslational Modifications in Proteins from Methanogenic Archaea by Top-Down MS. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2678-2683.

# CHAPTER 5

# PROTEOME ANALYSIS BY AN ON-TARGET DIGESTION MALDI-FTICR APPROACH


## ABSTRACT

Tryptic digestion of proteins is one of essential steps in bottom-up proteomics. A fast digestion protocol needs to be developed to meet the demands of high-throughput proteomics approaches. Here we demonstrate the effectiveness of tryptic digestion of protein standards on a MALDI target combined with MALDI-FTICR/MS analysis for rapid protein identification. This approach, on-target digestion, significantly reduces the digestion time from several hours required by conventional in-solution digestion to a few minutes. Moreover, on-target digestion greatly enhances the protein identification by reducing the number of unassignable peaks in mass spectrum. On-target digestion of a HPLC separated proteome with MALDI-FTICR/MS analysis offers an alternative to the standard method of shotgun analysis.


## INTRODUCTION

Proteomics is the systematic analysis of the proteins in a cell, tissue and organism for their identities, quantities and functions [1]. Mass spectrometry (MS) in combination with a wide variety of separation methods plays a central role in proteomic studies. With the continuous improvements of mass spectrometers in mass accuracy, resolution and sensitivity, significant advances have been made by MS-based proteomics. Nonetheless, the demands of developing high-throughput methods for rapid protein identification and quantitation still exist. There are two fundamental strategies used in MS-based proteomics. Top-down strategy involves the

ionization of intact proteins or large protein fragments followed by tandem mass spectrometric analysis [2-5], and more commonly, bottom-up strategy refers to proteolysis of protein mixtures followed by mass spectrometric analysis of peptides [6-8]. In either case, separation techniques such as two-dimensional electrophoresis (2-DE) or multidimensional liquid chromatography (LC) are often employed to reduce the sample complexity. Over the last decade, significant efforts have been made to develop high-throughput proteomics approaches. Tandem mass spectrometry has been broadly used in proteomics for characterization of proteins, but it is limited by its tedious process. Smith and coworkers [9-12] developed a method called "accurate mass and time (AMT)", in which tandem mass spectrometric analysis needs to be performed only once for creating peptide AMT tags for subsequent protein identification and therefore the throughput of the analysis is significantly improved. Digestion of proteins is one of essential steps in bottom-up proteomics. Trypsin is the most commonly used enzyme for digestion as it consistently generate uniform peptides that are easily analyzed by mass spectrometry [13] and produces more specific cleavages compared to the other enzymes. However, tryptic digestion typically requires several hours to effectively cleave the proteins into small peptide fragments, increasing the occurrence rate of trypsin autolysis [14] and non-specific cleavages [15]. Thus a number of rapid digestion strategies have been developed to overcome these problems. One accelerated digestion method that uses immobilized trypsin reactor was developed for coupling with electrospray ionization (ESI) analysis [16-19]. Another approach involves microwave-assisted protein digestion by using trypsin-immobilized magnetic nanoparticles [20-26]. Previous works also show that addition of a certain percentage of organic solvents (e.g. acetonitrile) during the digestion can reduce the digestion time with less number of non-specific tryptic cleavages [27]. Nonetheless,

growing demands of high-throughput proteomics approaches keeps promoting the development of other fast digestion protocols, particularly the approaches that are suitable to MALDI analysis.

Here we demonstrate the effectiveness of tryptic digestion of protein standards on a MALDI target combined with MALDI-FTICR/MS analysis for rapid protein identification. The potentials of integrating the HPLC protein separation, on-target digestion, and MALDI-FTICR mass spectrometry for shotgun analysis of complex biological samples are also discussed.

## EXPERIMENTAL

### Materials

Hemoglobin, β-lactoglobulin, chicken egg albumin (ovalbumin), transferring, and ammonium bicarbonate were purchased from Sigma (St. Louis, MO). Bovine serum albumin (BSA) was purchased from Calbiochem (San Diego, CA). Tris (2-carboxyethyl)-phosphine hydrochloride (TCEP) was purchased from Pierce (Rockford, IL). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI). 2,5-Dihydroxybenzoic acid (DHB) was purchased from Lancaster Synthesis (Pelham, NH). Acetonitrile (HPLC grade) and trifluoroacetic acid (TFA) were purchased from Fisher Scientific (Fair Lawn, NJ). Water used in the experiments was purified using a NanopureInfinity ultrapure water system (Barnstead/Thermolyne, Dubuque, IA).

### Sample Preparation

To evaluate the performance of on-target digestion by comparison with in-solution digestion, the following five protein standards were tested:

Hemoglobin: MW = 15964.4 Da, pI = 6.37

β-Lactoglobulin: MW = 19883.4 Da, pI = 4.93

Chicken egg albumin (ovalbumin): MW = 42881.5 Da, pI = 5.19

Bovine serum albumin (BSA): MW = 69293.9 Da, pI = 5.60

Transferrin: MW = 77753.7 Da, pI = 6.80

Each of these proteins was dissolved in 20 mM ammonium bicarbonate solution at the concentration of 1 mg/mL. 100 μL of protein solution was reduced by addition of 50-fold molar excess of 50 mM TCEP, denatured by heating at 95 ºC for 10 minutes, and allowed to cool down to the room temperature. The protein solution was then split into two fractions for the comparison between on-target digestion and in-solution digestion. For on-target digestion, three deposition methods that differ in the order of deposition of protein and trypsin were compared. Trypsin was deposited before protein for the predeposited method, mixed with protein during deposition for the mixing method, and deposited after protein for the postdeposited method. In either case, 0.5 μL of protein solution and 0.5 μL of trypsin solution were applied onto a stainless steel MALDI target. The protein-to-trypsin ratio was varied from 1 : 1 to 100 : 1 (m/m), and the digestion was allowed to occur at the room temperature for 5 minutes. Upon completion of the digestion, 0.5 μL of 200 mg/mL DHB solution was deposited and allowed to dry. For in-solution digestion, 10 μL of 0.1 mg/mL trypsin solution was added to the protein solution to make a 50 : 1 (m/m) protein-to-trypsin ratio, and the digestion was allowed to take place at 37 ºC for 12 hours. The peptides generated from two digestion protocols were analyzed by MALDI-FTICR mass spectrometry. Triplicate analysis was performed for each of the proteins at a given condition. A protein mixture was prepared by mixing these five protein standards at the concentration of 1mg/mL. A simple comparison was made between on-target digestion and in-solution digestion with this sample.

### MALDI-FTICR Mass Spectrometry

Mass spectrometric analysis of peptide mixtures was performed on a 7 tesla Bio-Apex Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with an intermediate pressure Scout100 MALDI source and a nitrogen laser. The mass spectra were recorded at trapping potential of 0.80 V using 80 laser shots and externally calibrated via Bruker XMASS 7.0.8 software (Bruker Daltonics, Billerica, MA) using the tryptic digest of chicken egg albumin.

### Peptide Assignment and Protein Identification

Peptide assignment was made by picking the monoisotopic peptide masses from the mass spectrum via DataAnalysis Version 3.4 Build 191 (Bruker Daltonics, Billerica, MA) and then comparing the generated list against a list of the predicted tryptic peptides from the original protein through MS-Fit (http://prospector.ucsf.edu/), with 10 ppm mass tolerance and 2 maximum missed cleavages selected.

## RESULTS AND DISCUSSION

### Optimization of On-target Digestion

On-target tryptic digestion was tested and optimized using BSA. First, the digestion was performed on a MALDI target using three deposition methods: spotting trypsin before protein deposition (predeposited method), spotting trypsin with protein (mixing method), and spotting trypsin after protein deposition (postdeposited method). The digested BSA was analyzed by MALDI-FTICR mass spectrometry, and the results are shown in Figure 5.1. Although the relative abundance of peptides varies, the three deposition methods yield a similar digestion

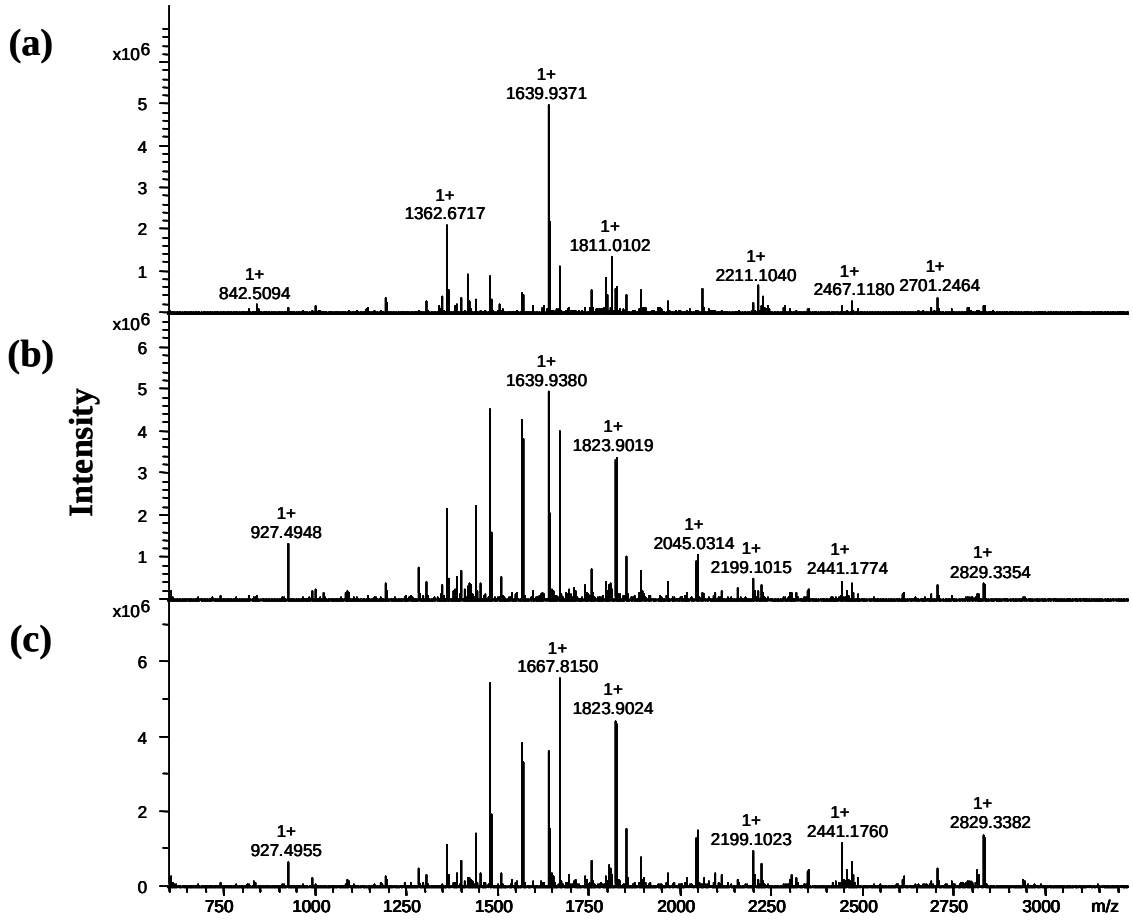



**Figure 5.1** On-target digestion of BSA using (a) predeposited method, (b) mixing method, and (c) postdeposited method with the protein-to-trypsin ratio of 10 : 1 (m/m).

pattern, indicating that the order of deposition does not significantly affect the digestion quality. The predeposited method seems to be less attractive than the other two methods because a few more trypsin autolysis products were observed when using this method. In contrast to the mixing method, the postdeposited method is more easily implemented on the MALDI target and therefore was chosen to be used in the following experiments. Second, the protein-to-trypsin ratio was varied from 1:1 to 100:1 (m/m) in order to determine the optimal ratio for on-target digestion. Results show that the ratio of 10:1 yields the highest sequence coverage for protein identification and less number of trypsin autolysis products (Figure 5.2). Third, the influence of incubation on the digestion was investigated by placing the MALDI target in a humid chamber at about 40 $^{o}$C for 15 minutes during the digestion. Compared to the on-target digestion without incubation, a similar set of peptides were produced when using incubation (data not shown), indicating that incubation can be avoided for on-target digestion. Based on all these observations, on-target digestion should be carried out using the postdeposited method and the protein-to-trypsin ratio of 10 : 1 (m/m) to ensure a good digestion quality.

**Comparison of On-Target Digestion with In-Solution Digestion**

To evaluate the performance of on-target digestion, five protein standards (hemoglobin, ß-lactoglobulin, ovalbumin, BSA and transferrin) with the molecular weights ranging from16 kDa to 78 kDa were tested. A comparison was made for each of these proteins by comparing the database search results obtained from on-target digestion with that from in-solution digestion. As shown in Table 5.1, on-target digestion yield comparable sequence coverage to that obtained using in-solution digestion for most of the proteins tested (with the exception of ovalbumin). In addition, on-target digestion significantly improves the percentage of the matched masses,

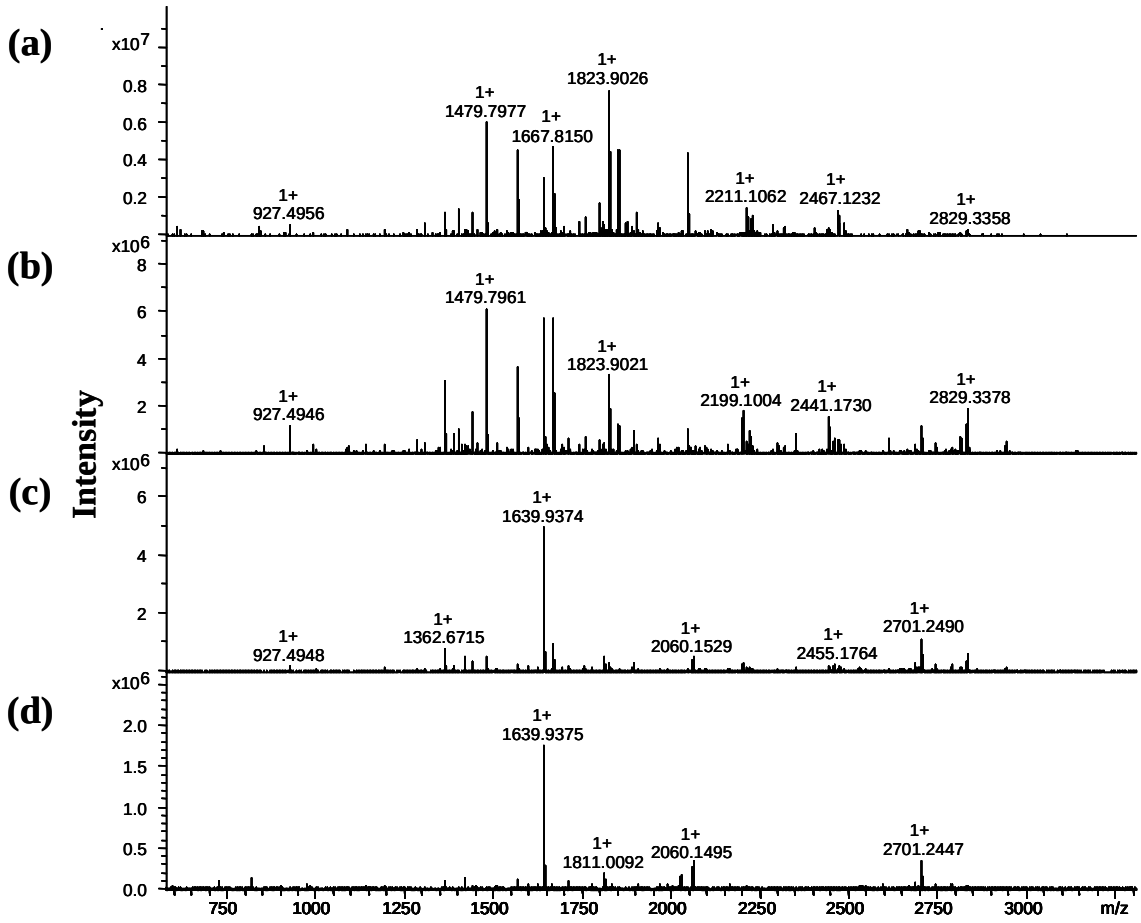



**Figure 5.2** On-target digestion of BSA using the postdeposited method with the protein-to-trypsin ratio of (a) 1 : 1, (b) 10 : 1, (c) 20 : 1, and (d) 70 : 1 (m/m).

**Table 5.1** Comparison of database search results of five protein standards obtained using on-target digestion with that using in-solution digestion.

| Protein Name | In-Solution Digestion | | | On-Target Digestion | | |
|---|---|---|---|---|---|---|
| | Total Intensity[1] | Matched Masses (%)[2] | Sequence Coverage (%) | Total Intensity[1] | Matched Masses (%)[2] | Sequence Coverage (%) |
| Hemoglobin | 5.1E+07 | 22 | 84 | 5.4E+06 | 24 | 70 |
| β-Lactoglobulin | 6.3E+07 | 13 | 73 | 3.3E+07 | 25 | 74 |
| Ovalbumin | 8.8E+07 | 28 | 74 | 2.3E+07 | 38 | 38 |
| BSA | 1.1E+08 | 44 | 62 | 7.4E+07 | 54 | 73 |
| Transferrin | 3.6E+07 | 63 | 60 | 5.5E+07 | 73 | 68 |

[1] The total intensity was calculated by summarizing the intensities of all matched peaks.

[2] The percentage of matched masses was calculated by dividing the number of matched masses by the number of masses submitted.

suggesting that the number of unassignable peaks in mass spectrum is reduced. This feature will benefit the protein identification for a complex biological sample where a large number of peaks can not be assigned in a typical shotgun experiment. Certainly, the major advantage of on-target digestion over in-solution digestion is the rapid digestion process. On-target digestion generally takes place in a few minutes while in-solution digestion requires at least several hours to ensure an effective digestion. The shorten digestion time decreases the occurrence rate of trypsin autolysis and non-specific cleavages, and therefore reduces the number of unassignable peaks generated during the digestion.

After testing with single protein standards, the effectiveness of on-target digestion on a complex protein mixture was investigated. For this purpose, five protein standards were mixed, processed by on-target digestion, and analyzed using MALDI-FTICR mass spectrometry. The results were evaluated by comparing with those obtained for the same sample using in-solution digestion. As shown in Figure 5.3, no significant difference in the number and types of peptides was observed between two digestion protocols. The database search results in Table 5.2 demonstrate again that on-target digestion produces satisfactory sequence coverage with less number of unassignable peaks.

**Experimental Challenges**

After optimizing the experimental conditions with protein standards, an attempt was made to test this approach for the analysis of a proteome sample obtained from *Methanococcus maripaludis* by integrating the HPLC protein separation, on-target digestion, and MALDI-FTICR mass spectrometry (a modified procedure, Path 2 in Figure 5.4). This sample was previously analyzed using a standard procedure for shotgun proteomics, in which the whole

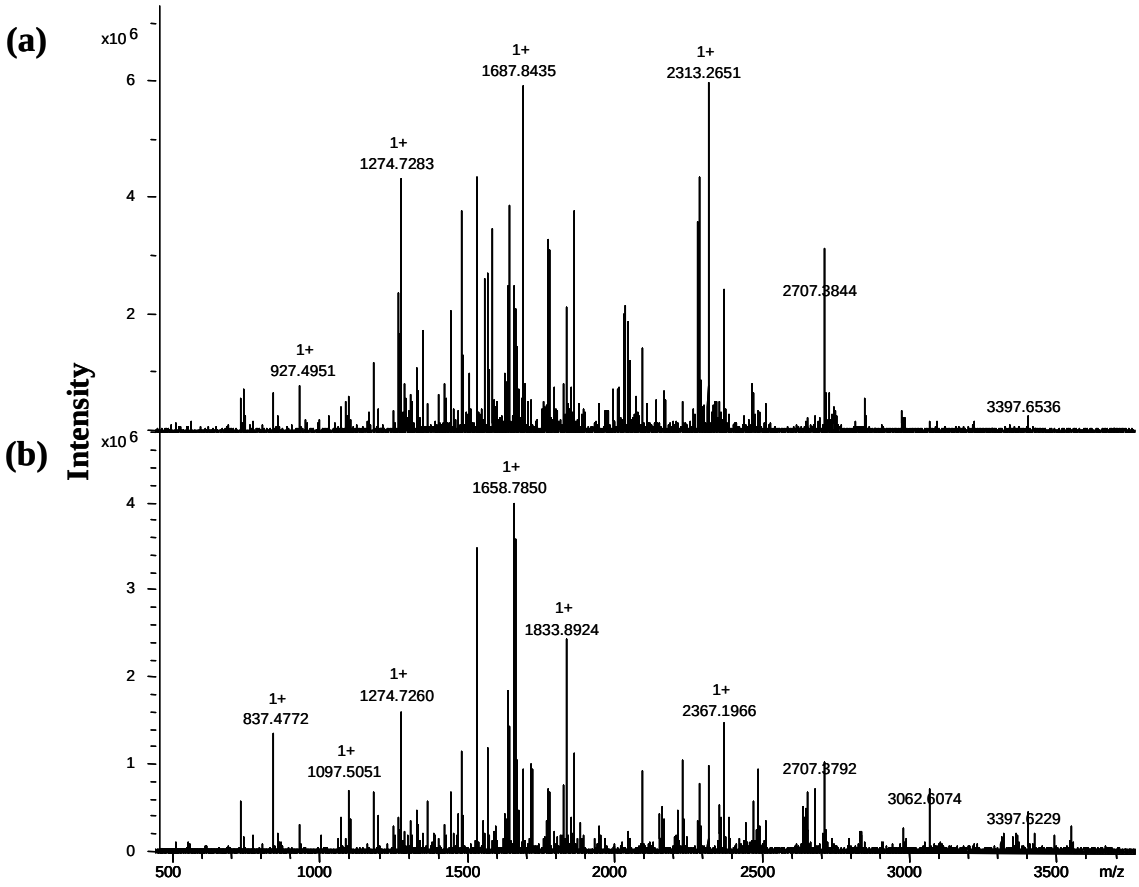



**Figure 5.3** (a) In-solution digestion and (b) on-target digestion of a protein mixture using the postdeposited method with the protein-to-trypsin ratio of 10 : 1 (m/m).

**Table 5.2** Comparison of database search results of a protein mixture obtained using on-target digestion with that using in-solution digestion.

| Protein Name | In-Solution Digestion | | On-Target Digestion | |
|---|---|---|---|---|
| | Matched Masses (%)[1] | Sequence Coverage (%) | Matched Masses (%)[1] | Sequence Coverage (%) |
| Hemoglobin | 4 | 81 | 6 | 53 |
| β-Lactoglobulin | 4 | 63 | 9 | 77 |
| Ovalbumin | 9 | 61 | 8 | 39 |
| BSA | 13 | 53 | 21 | 54 |
| Transferrin | 9 | 36 | 24 | 45 |

[1] The percentage of matched masses was calculated by dividing the number of matched masses by the number of masses submitted.

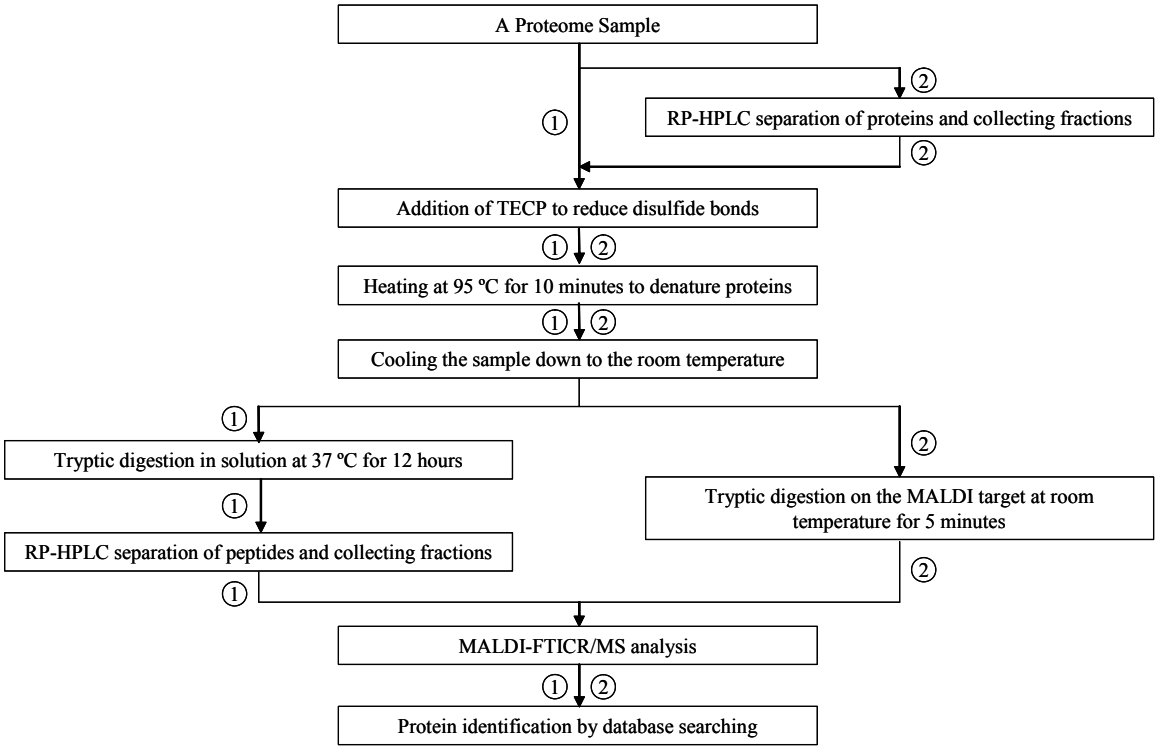



**Figure 5.4** The standard procedure (Path 1) and the modified procedure (Path 2) for shotgun proteomic analysis of a proteome sample using MALDI-FTICR mass spectrometry.

proteome sample was batch digested in solution, and the resulting peptides were subjected to the separation using RP-HPLC and MALDI-FTICR/MS analysis (Path 1 in Figure 5.4). It is of great interest to compare the results obtained using the standard procedure with that using the modified procedure, especially focused on the differences in the protein coverage and in the classes of proteins that are favored by each approach. Unfortunately, after several trials on HPLC separation of the proteins in the proteome sample, we only observed a few weak broad peaks in the HPLC chromatogram, indicating that the majority of the proteins were not eluted from the analytical column. Although the HPLC used in this experiment has shown excellent performance when tested with a mixture that contains five small protein standards, it was not able to successfully separate the proteins in our proteome sample. The failure could arise from the strong binding between the large proteins present in the sample and the packing material (nonporous C18) of the analytical column. Since the protein separation is an initial and essential step in the modified procedure, a more powerful HPLC that is capable to resolve complex biological samples should be used to ensure the experiment's success.

**Potentials of Combining Protein Separation with On-Target Digestion**

The combination of the HPLC protein separation, on-target digestion and MALDI-FTICR/MS analysis of peptides provides an alternative to the standard method of shotgun analysis. This approach is particularly useful for protein identification using accurate mass measurement for a highly complex biological sample. Batch digestion of protein mixtures can produce thousands to hundreds of thousands of peptides, making it difficult to uniquely assign the peptides solely based on their molecular weights. With this new approach, the proteins in the mixture are separated first, resulting in one or several proteins for each fraction collected.

Peptide assignment can be greatly facilitated by correlating the peptide masses with the proteins in the fraction from which the peptides are derived. This combination also allows for a variety of mass spectrometric analyses to be performed, such as measurement of intact protein mass, peptide mass fingerprinting, and sequencing the peptide by tandem mass spectrometry. Rapid and unambiguous identification of proteins can be achieved by integrating all the information obtained from the above-mentioned experiments. This was demonstrated by Getie-Kebtie and coworkers who established a unique method that combines RP-HPLC separation of proteins, on-target digestion and MALDI mass spectrometry for the analysis of a complex biological sample, MRC5 cell lysate [14].

## CONCLUSIONS

An on-target digestion protocol was tested and optimized with protein standards. The best digestion quality can be achieved when using the postdeposited method and the protein-to-trypsin ratio of 10 : 1 (m/m). On-target digestion yields comparable sequence coverage to in-solution digestion with less number of unassignable peaks. Moreover, on-target digestion significantly reduces the digestion time from several hours required by in-solution digestion to a few minutes, holding a promise in high-throughput proteomics. The combination of the HPLC protein separation, on-target digestion, and MALDI-FTICR/MS analysis offers an alternative to the standard method of shotgun analysis.

## REFERENCES

1 Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; HumpherySmith, I.; Williams, K. L.; Hochstrasser, D. F. From

Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. *Bio-Technology* **1996,** *14*, 61-65.

2 Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top-Down Mass Spectrometry of a 29-Kda Protein for Characterization of Any Posttranslational Modification to within One Residue. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99*, 1774-1779.

3 Kelleher, N. L. Top-Down Proteomics. *Analytical Chemistry* **2004,** *76*, 196A-203A.

4 Bogdanov, B.; Smith, R. D. Proteomics by FTICR Mass Spectrometry: Top Down and Bottom Up. *Mass Spectrometry Reviews* **2005,** *24*, 168-200.

5 Wehr, T. Top-Down versus Bottom-up Approaches in Proteomics. *Lc Gc North America* **2006,** *24*, 1004.

6 Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Analytical Chemistry* **2001,** *73*, 5683-5690.

7 Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nature Biotechnology* **2001,** *19*, 242-247.

8 Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P. Multidimensional Separations-Based Shotgun Proteomics. *Chemical Reviews* **2007,** *107*, 3654-3686.

9 Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y. F.; 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.

10 Qian, W. J.; Camp, D. G.; Smith, R. D. High-Throughput Proteomics Using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Expert Review of Proteomics* **2004,** *1*, 87-95.

11 Jacobs, J. M.; Monroe, M. E.; Qin, W. J.; Shen, Y. F.; Anderson, G. A.; Smith, R. D. Ultra-Sensitive, High Throughput and Quantitative Proteomics Measurements. *International Journal of Mass Spectrometry* **2005,** *240*, 195-212.

12 Zimmer, J. S. D.; Monroe, M. E.; Qian, W. J.; Smith, R. D. Advances in Proteomics Data Analysis and Display Using an Accurate Mass and Time Tag Approach. *Mass Spectrometry Reviews* **2006,** *25*, 450-482.

13 Angel, P. M.; Orlando, R. Trypsin Is the Primary Mechanism by Which the O-18 Isotopic Label Is Lost in Quantitative Proteomic Studies. *Analytical Biochemistry* **2006,** *359*, 26-34.

14 Getie-Kebtie, M.; Franke, P.; Aksamit, R.; Alterman, M. A. Experimental Evaluation of Protein Identification by an LC/MALDI/On-Target Digestion Approach. *Journal of Proteome Research* **2008,** *7*, 3697-3707.

15 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.

16 Hedstrom, M.; Andersson, M.; Galaev, I. Y.; Mattiasson, B. Fast on-Column Protein Digestion with Subsequent Peptide Mapping Using Tandem Mass Spectrometry with Information Dependent Acquisition. *Journal of Chromatography A* **2005,** *1080*, 117-123.

17 Duan, J. C.; Liang, Z.; Yang, C.; Zhang, J.; Zhang, L. H.; Zhang, W. B.; Zhang, Y. K. Rapid Protein Identification Using Monolithic Enzymatic Microreactor and LC-ESI-MS/MS. *Proteomics* **2006,** *6*, 412-419.

18 Duan, J. C.; Sun, L. L.; Liang, Z.; Zhang, J.; Wang, H.; Zhang, L. H.; Zhang, W. B.; Zhang, Y. K. Rapid Protein Digestion and Identification Using Monolithic Enzymatic Microreactor Coupled with Nano-Liquid Chromatography-Electro Spray Ionization Mass Spectrometry. *Journal of Chromatography A* **2006,** *1106*, 165-174.

19 Zhang, K.; Wu, S.; Tang, X. T.; Kaiser, N. K.; Bruce, J. E. A Bifunctional Monolithic Column for Combined Protein Preconcentration and Digestion for High Throughput Proteomics Research. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2007,** *849*, 223-230.

20 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.

21 Lin, S. S.; Wu, C. H.; Sun, M. C.; Sun, C. M.; Ho, Y. P. Microwave-Assisted Enzyme-Catalyzed Reactions in Various Solvent Systems. *Journal of the American Society for Mass Spectrometry* **2005,** *16*, 581-588.

22 Sun, W.; Gao, S.; Wang, L.; Chen, Y.; Wu, S.; Wang, X.; Zheng, D.; Gao, Y. Microwave-Assisted Protein Preparation and Enzymatic Digestion in Proteomics. *Molecular & Cellular Proteomics* **2006,** *5*, 7.

23 Chen, W. Y.; Chen, Y. C. Acceleration of Microwave-Assisted Enzymatic Digestion Reactions by Magnetite Beads. *Analytical Chemistry* **2007,** *79*, 2394-2401.

24 Lill, J. R.; Ingle, E. S.; Liu, P. S.; Pham, V.; Sandoval, W. N. Microwave-Assisted Proteomics. *Mass Spectrometry Reviews* **2007,** *26*, 657-671.

25 Sandoval, W. N.; Pham, V.; Ingle, E. S.; Liu, P. S.; Lill, J. R. Applications of Microwave-Assisted Proteomics in Biotechnology. *Combinatorial Chemistry & High Throughput Screening* **2007,** *10*, 751-765.

26 Lin, S.; Yun, D.; Qi, D. W.; Deng, C. H.; Li, Y.; Zhang, X. M. Novel Microwave-Assisted Digestion by Trypsin-Immobilized Magnetic Nanoparticles for Proteomic Analysis. *Journal of Proteome Research* **2008,** *7*, 1297-1307.

27 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.

# CHAPTER 6

# CONCLUSIONS

Proteomics is a systematic study of the whole complement or subset of the proteins expressed in a cell, tissue and organism [1]. Mass spectrometry (MS) in combination with a variety of separation methods has become a primary tool for proteomic studies. Of all proteomic approaches, shotgun proteomics has gained tremendous popularity in recent years. This method involves batch proteolysis of a protein mixture, separation of peptides by single or multidimensional liquid chromatography (LC), and protein identification from mass spectrometric analysis of peptides [2-4]. Fourier transform ion cyclotron resonance mass spectrometry (FTICR/MS) has been broadly used in proteomics because of its unprecedented high mass accuracy, resolution and sensitivity. This dissertation describes three approaches to facilitate shotgun proteomic analysis by accurate mass measurement using MALDI-FTICR/MS.

The first approach described in this dissertation, referred to as mass defect labeling (MDL) [5], is based on the derivatization of tryptophan residues in proteins with a reagent that alters their mass defects. With this approach, protein identification can be significantly improved by reducing the congestion of mass spectra for a complex protein mixture. Chapter 2 reports the development and application of a new tryptophan MDL reagent, 4,6-dibromo-(trifluoromethyl) benzenesulfenyl chloride. Initially, 2-nitrobenzenesulfenyl chloride has been widely used for many years as a tryptophan derivatizing reagent; however,

its photosensitivity limits its applications in MALDI analysis. To address this problem, we have developed a new tryptophan tag, 2-(trifluoromethyl) benzenesulfenyl chloride [6]. This compound shows identical reaction specificity and efficiency toward tryptophan compared to 2-nitrobenzenesulfenyl chloride. Most significantly, it exhibits high photostability in MALDI analysis. To serve as a MDL reagent, the dibromo derivative of 2-(trifluoromethyl) benzenesulfenyl chloride has been synthesized. This new tryptophan MDL reagent, 4,6-dibromo-(trifluoromethyl) benzenesulfenyl chloride, holds high promise as a tool for proteomic analysis using MALDI-MS strategy.

Chapter 3 describes an approach that employs the stored waveform inverse Fourier transform (SWIFT) excitation [7, 8] in MALDI-FTICR/MS analysis to improve the mass accuracy of higher mass peptides. A comparison was made between SWIFT and chirp to investigate the effect of either of these excitation methods on the mass measurement accuracy (MMA) of the peptides up to mass-to-charge ratio (*m/z)* 4000, with or without applying stepwise-external calibration [9]. Results showed that SWIFT provides significantly better mass accuracy than that can be obtained by chirp, particularly for ions of higher *m/z* ratio, and that sub part-per-million (ppm) can be achieved when combining SWIFT with stepwise-external calibration [10]. These findings are expected to benefit analysis of complex mixtures such as batch proteolytic digests of protein mixtures, where higher dynamic range and higher mass accuracy are required.

In Chapter 4, a HPLC-MALDI-FTICR/MS approach that combines accurate mass measurement with nitrogen stoichiometry for shotgun proteomics is presented. This method allows for simultaneous identification and quantitation of the proteins in a $^{15}$N-metabolically

labeled proteome sample that was obtained from *Methanococcus maripaludis*. We demonstrated that protein identification can be significantly improved when using the nitrogen stoichiometry as a constraint along with the molecular weight of peptides in database searching. Further enhancement in protein identification can be achieved by applying SWIFT excitation and stepwise-external calibration, due to the improved mass measurement accuracy. Quantitative measurements of the proteome sample were also made by determining the intensity ratios of $^{14}N/^{15}N$ peptide pairs.

Chapter 5 describes an approach that combines on-target digestion with MALDI-FTICR/MS analysis for rapid protein identification. An on-target digestion protocol was tested and optimized with protein standards. The best digestion quality can be achieved when using the postdeposited method and the protein-to-trypsin ratio of 10 : 1 (m/m). On-target digestion yields comparable sequence coverage to in-solution digestion with less number of unassignable peaks. Moreover, on-target digestion significantly reduces the digestion time from several hours required by in-solution digestion to a few minutes, holding a promise in high-throughput proteomics. The combination of the HPLC protein separation, on-target digestion, and MALDI-FTICR/MS analysis offers an alternative to the standard method of shotgun analysis.

**REFERENCES**

1 Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; HumpherySmith, I.; Williams, K. L.; Hochstrasser, D. F. From Proteins to Proteomes: Large Scale Protein Identification by

Two-Dimensional Electrophoresis and Amino Acid Analysis. *Bio-Technology* **1996,** *14*, 61-65.

2 Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001,** *73*, 5683-5690.

3 Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001,** *19*, 242-247.

4 Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P. Multidimensional Separations-Based Shotgun Proteomics. *Chem. Rev.* **2007,** *107*, 3654-3686.

5 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. *Anal. Chem.* **2006,** *78*, 3417-3423.

6 Li, C. Y.; Gawandi, V.; Protos, A.; Phillips, R. S.; Amster, I. J. A Matrix-Assisted Laser Desorption/Ionization Compatible Reagent for Tagging Tryptophan Residues. *Eur. J. Mass Spectrom.* **2006,** *12*, 213-221.

7 Marshall, A. G.; Wang, T. C. L.; Ricca, T. L. Tailored Excitation for Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry. *J. Am. Chem. Soc.* **1985,** *107*, 7893-7897.

8 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.

9 Wong, R. L.; Amster, I. J. Sub Part-Per-Million Mass Accuracy by Using Stepwise-External Calibration in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2006,** *17*, 1681-1691.

10 Jing, L.; Li, C. Y.; Wong, R. L.; Kaplan, D. A.; Amster, I. J. Improved Mass

Accuracy for Higher Mass Peptides by Using Swift Excitation for MALDI-FTICR Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2008,** *19*, 76-81.