NEW TECHNIQUES FOR PROTEOMIC ANALYSIS WITH FTICR-MS:

INSTRUMENTATION, AUTOMATION AND APPLICATION

by

LI JING

(Under the Direction of I. JONATHAN AMSTER)

ABSTRACT

Shotgun proteomics has been broadly used for high-throughput analysis of proteins in biological systems. In shotgun proteomics, proteins are digested and the resulting peptides are analyzed using a combination of high performance liquid chromatography (HPLC) and mass spectrometry (MS). Our approach relies on accurate mass measurement of peptides by Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). In this work, we explore and develop methodologies to improve proteomic analysis.

First, we develop an algorithm to automatically identify light/heavy peptide pairs and calculate peptide relative abundance ratios in ^{15}N metabolic labeling data. This method provides over 99% accuracy in assigning peak pair and reduces the time of data analysis from over 100 hours to tens of minutes.

In the second approach, we investigate the mass accuracy for higher mass peptides by using stored waveform inverse Fourier transform (SWIFT) excitation for MALDI-FTICR mass spectrometry. Analysis of measurement errors reveals that SWIFT excitation provides smaller deviations from stepwise-external calibration and better mass accuracy than chirp excitation for a wide mass range and for widely varying ion populations.

A shotgun proteomics approach is presented for simultaneous identification and quantitation of the proteins from a proteome using accurate mass measurement and nitrogen stoichiometry. 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/heavy peptide pairs. The combination of stepwise-external calibration and SWIFT excitation is applied and the mass measurement accuracy (MMA) is significantly improved.

Last, we describe a novel calibration method, N15Cal, which corrects the space charge induced frequency shift in FTICR-MS analysis of ^{15}N-metabolically labeled peptides from a batch digested proteome. N15Cal utilizes the information from the mass difference between the ^{14}N/^{15}N peptide peak pairs to correct for space charge induced mass shifts. There is no need to include an internal calibrant or to apply ion population control. N15Cal has been successfully applied to the LC-MALDI-FTICR data of ^{15}N-metabolic labeling proteomics.

INDEX WORDS: Proteomics, FTICR MS, accurate mass measurement, data analysis, algorithm, ^{15}N metabolic label, SWIFT, mass accuracy, mass calibration

NEW TECHNIQUES FOR PROTEOMIC ANALYSIS WITH FTICR-MS: INSTRUMENTATION, AUTOMATION AND APPLICATION

by

LI JING

B.S. Fudan University, China, 2002

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

NEW TECHNIQUES FOR PROTEOMIC ANALYSIS WITH FTICR-MS: INSTRUMENTATION, AUTOMATION AND APPLICATION

by

LI JING

Major Professor:	I. Jonathan Amster
Committee:	John Stickney Lance Wells

Electronic Version Approved:

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

DEDICATION

To my parents, Kangwen Jing and Miaoxian Dong, my husband, Wei Zhong, and my children, Athena Zhong and Colin Zhong for their unconditional support and love.

ACKNOWLEDGEMENTS

My first, and most earnest, acknowledgement must go to my major advisor, Processor I. Jonathan Amster, for his guidance and encouragement, as well as for providing me with excellent facilities and ensuring financial support ever since I joined his lab. Many thanks also go to my committee members Dr. John Stickney and Dr. Lance Wells, for their valuable time and comments, which significantly improved the quality of this dissertation.

I would like to thank Dr. William B. Whitman and Yuchen Liu for providing the ^{15}N-metabolically labeled proteome sample. I would also like to thank my group members, Richard Wong, Jeremy Wolff, Hilda Barry, Chunyan Li, Dhanashri Bagal, Lisabeth Hoffman, Franklin Leach and Melissa Warren for their ideas, support and assistance in the lab.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1. PROTEOMICS AND MASS SPECTROMETRY

Proteomics, the systematic characterization of the proteins expressed in a cell or tissue, is now the focus of many fields of science [1, 2]. The term proteome is defined as being the protein complement to the genome [3]. In general, proteomics deals with the large-scale determination of gene and cellular function directly at the protein level [4, 5]. Different properties have been studied in proteomics and molecular weight is one of the most important and straightforward properties. Mass spectrometry (MS), which is composed of an ion source, mass selective analyzer and detector, is a technique to measure mass-to-charge ratio (*m/z*) of gas phase ions. Techniques for ionization have been key to determining what types of samples can be analyzed by mass spectrometry. Earlier ionization techniques like electron ionization (EI) [6] and chemical ionization (CI) [7] required gas phase analyte molecules to be present, and thus were only suitable for volatile compounds such as organic molecules. EI and CI ionization techniques limited the application of mass spectrometry on non-volatile biological macromolecules. In the late 1980s, two Nobel-winning "soft" ionization methods, matrix-assisted laser desorption/ionization (MALDI) [8] and electrospray ionization (ESI) [9], were developed, which allowed a breakthrough for mass spectrometry in biological sciences and related studies. Mass spectrometry (MS)-based methods for the identification of proteins have become a standard platform in proteomics.

1.1 Matrix Assisted Laser Desorption/Ionization (MALDI)

The term matrix-assisted laser desorption/ionization (MALDI) was coined in 1985 by Franz Hillenkamp, Michael Karas and their colleagues [10]. This group found that the amino acids could be ionized with a pulsed 266 nm laser. In 1988, the applications of this method to large biomolecules were reported by K. Tanaka [11], Karas and Hillenkamp [8], who demonstrated that a protein can be ionized with the proper combination of laser wavelength and matrix.

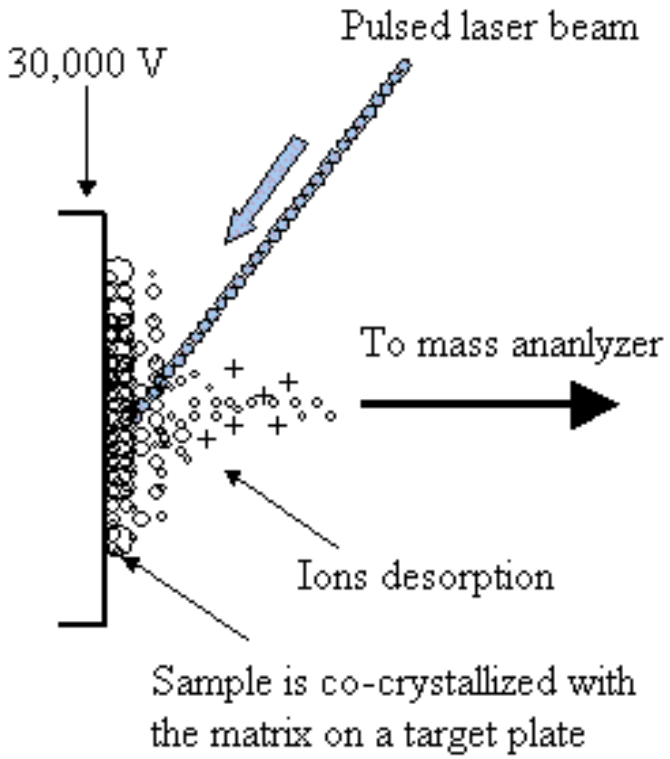


Figure 1.1 Diagram of the MALDI process [12].

In MALDI, analyte is mixed with a highly absorbing matrix, which typically consist an aromatic acid with a chromophore that has a strong absorption at the applied laser wavelength [13]. When the mixture is dried on the sample target, the matrix will form co-crystal with the analyte and can be ablated by intense pulses of laser. The laser irradiation induces an explosive evaporation of matrix molecules, and the analyte is entrained in the gas phase matrix plume [12] (Figure 1.1). The matrix plays an important role in the ionization process. The selection of the best matrix usually depends on both mass analyzer and the nature of the analyte [13-16]. Factors such as the selection and concentration of the matrix can strongly influence the quality of mass spetrum. A typical MALDI spectrum includes mainly the mono-charged molecular ions.

1.2 Electrospray Ionization (ESI)

As introduced by John Fenn [9], the electrospray ionization (ESI) is to apply a strong electric field to a liquid passing through a capillary at atmospheric pressure. This technique produces a fine continuous spray of highly charged droplets, which enable ESI easy to be interfaced with high performance liquid chromatography (HPLC) or capillary electrophoresis (CE).

In the electrospray process, analyte is dissolved in a solution that has equal proportion of aqueous and organic solvents. The high electric field is obtained by applying a potential difference (2-6 kV) between tip and counter electrode, which induce a large charge accumulation at the liquid surface located at the end of the capillary tip. When the surface of liquid drop is broken, the liquid drop changes to “Taylor cone” [17-20] (Figure 1.2). ESI is able to produce multiply charged ions, which make it possible to analyze large molecules at relatively low mass-

to-charge ratios [20, 21]. Multiply charged ion also has advantage in tandem mass spectrometry as compared to fragmentation of a singly charged ion [22].

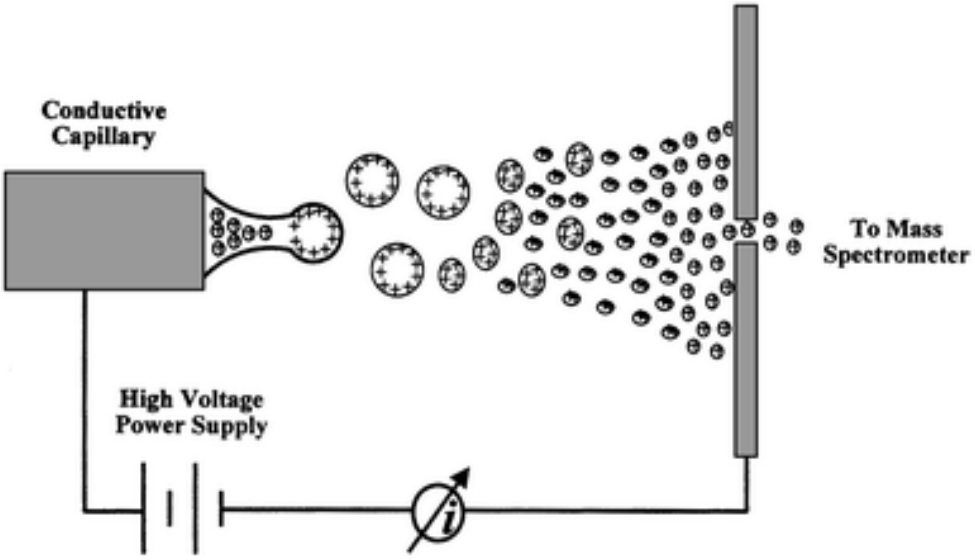


Figure 1.2 Diagram of the ESI process [23].

1.3 Mass Spectrometry in Proteomics

With the development of the modern ionization techniques (MALDI and ESI) and the increasing availability of genomic sequence databases, mass spectrometry has played an increasingly significant role for studying of complex protein samples. Such proteome-wide analysis provides important information of proteins, such as sequence, quantity, post-translational modification (PTM), interactions and structure, which is critical to the study of biological organisms. Modern mass spectrometric techniques coupled with advanced protein/peptide separation methods can provide sensitive, specific means of proteome analysis.

Over the past decade, mass spectrometry has undergone tremendous change culminating in the development of highly sensitive, robust instruments to analyze biomolecules, particularly proteins and peptides.

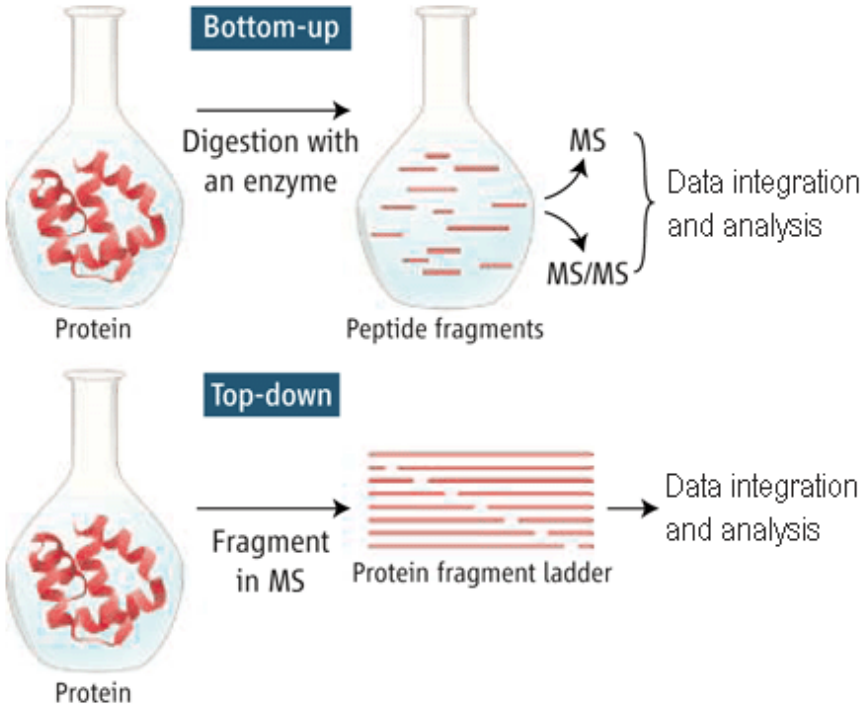


Figure 1.3 Diagram of bottom up and top down proteomics [24].

In general, two different strategies are applied in MS-based proteome analysis, “bottom up” and “top down” [25-27]. “Bottom up” is also known as “shotgun proteomics”, which involves cleaving proteins into peptides for protein identification (Figure 1.3, top panel). This method is able to identify protein based upon mass measurements of proteolytic digestion

products and/or MS fragmentation (MS/MS) of one or more peptides from a protein [28, 29]. Bottom up approaches have been broadly used for high-throughput analysis of proteins in biological systems and it is capable of providing quantitative information of protein expression by using stable isotopic labeling technology. In contrast, The “top down” technology involves direct analysis of intact proteins, without the need for prior proteolytic digestion [24] (Figure 1.3, bottom panel). This method strives to examine site-specific mutations and post-translational modifications (PTMs) of proteins by measuring them intact, rather than measuring peptides produced from them by proteolysis. The top-down technology has made valuable contributions to out knowledge of site-specific PTMs, protein structures and protein-protein interaction networks.

2. PROTEIN IDENTIFICATION AND QUANTIFICATION

2.1 Gel Electrophoresis

Traditionally, mass spectrometry-based proteomics relied upon the generation of relatively low complexity samples, where proteins were initially separated by two-dimensional gel electrophoresis (2D-GE) prior to the identification of peptides using MS [30]. Proteins are separated in one dimension by isoelectric point (PI) and in the other dimension by molecular weight (MW) [31]. Once proteins are separated in gels, they can be visualized by staining techniques such as silver staining, Coomassie blue staining, and SYPRO ruby staining. In-gel digestion can be performed to study the proteins present in spots of interest. Although this method offers high resolution for protein separation, it has a number of limitations. The individual extraction, digestion and analysis of each spot are a tedious and time-consuming process even with current advances in automation. This technique also has significant limitations

in sensitivity and dynamic range, and results in significant biases against hydrophobic proteins, low abundant proteins and proteins having extremes in molecular weight or PI [30, 32-34]. Compare to 2D-GE technology, 1D SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) has a relatively larger sample loading amount, and comparisons of multiple samples can be done in one gel by loading samples lane-by-lane. However, the limited resolving capacity increases the complexity in a single gel band.

2.2 Shotgun Proteomics

With the difficulties of gel-based protein separation, many researchers have explored new methods that are based on peptide separation. Shotgun proteomics utilizes liquid chromatography (LC) as a separation technique for a complex mixture of peptides that is generated by a batch digestion [35, 36]. Typically, liquid chromatography is coupled with tandem mass spectrometry (MS/MS) resulting in high throughput peptide analysis. Each peptide is correlated with its precursor protein, so the peptide can be identified by searching the MS/MS spectrum against the protein database. Currently, Sequest [37] and Mascot [38] are the most frequently used algorithms for peptide identification, which compare experimental MS/MS spectra with in silico spectra generated from the peptide sequences in a database.

Although the measurement of sequence tags by LC-MS/MS technique is a very powerful tool in proteomics for protein identification, alternatives have been introduced which omit the tandem mass spectrometry step and make use of the fact that many tryptic peptides have unique mass. The highly accurate mass measurement capability using high resolution MS has enabled the identifications of protein based on the unique *m/z* of peptide; however, this application is limited to relatively simple proteome sample [39]. This limitation has been largely overcome by

utilizing LC retention time information obtained from reverse phase LC separation [40, 41]. Compared to top-down proteomics techniques, shotgun proteomics avoids the modest separation efficiency and poor mass spectral sensitivity associated with intact protein analysis. This high throughput approach has been one of the most powerful and most popular methods for proteomic study.

2.3 Stable-Isotope Protein Labeling for Quantitative Proteomics

One of the major aims of proteomics is to provide quantitative data on differential protein expression levels for instance in healthy and diseased states. Conveniently, the proteomics approach uses protein separation by 2D-GE, followed by staining of the proteins. The staining intensity of proteins from two samples is compared and 'up-' or 'down-regulated' protein expression levels are identified.

To date, mass spectrometry based methods have been introduced that can provide quantitative data on differential protein expression, mostly using stable isotope labeling methods [4, 42]. Stable isotope incorporation has been achieved in a variety of ways, including metabolic labeling, enzymatically directed methods, and chemical labeling using externally introduced tags [43-45] (Figure 1.4). Use of stable isotope incorporation involves the addition to a sample of chemically identical form of the analyte(s), containing stable heavy isotope (e.g., ^{2}H, ^{13}C, ^{15}N, etc.) as an internal standard. The light/heavy peptides have the same properties and behave with essentially identical characteristics under any isolation or separation step. Pairs of chemically identical analytes of different stable-isotope composition result in light and heavy peaks in mass spectrum, and the ratio of signal intensities for such analyte pairs accurately indicates the

abundance ratio of two analytes [4]. This way, relative protein quantitation is achieved when two proteins or protein mixtures are compared where one serves as a control sample.

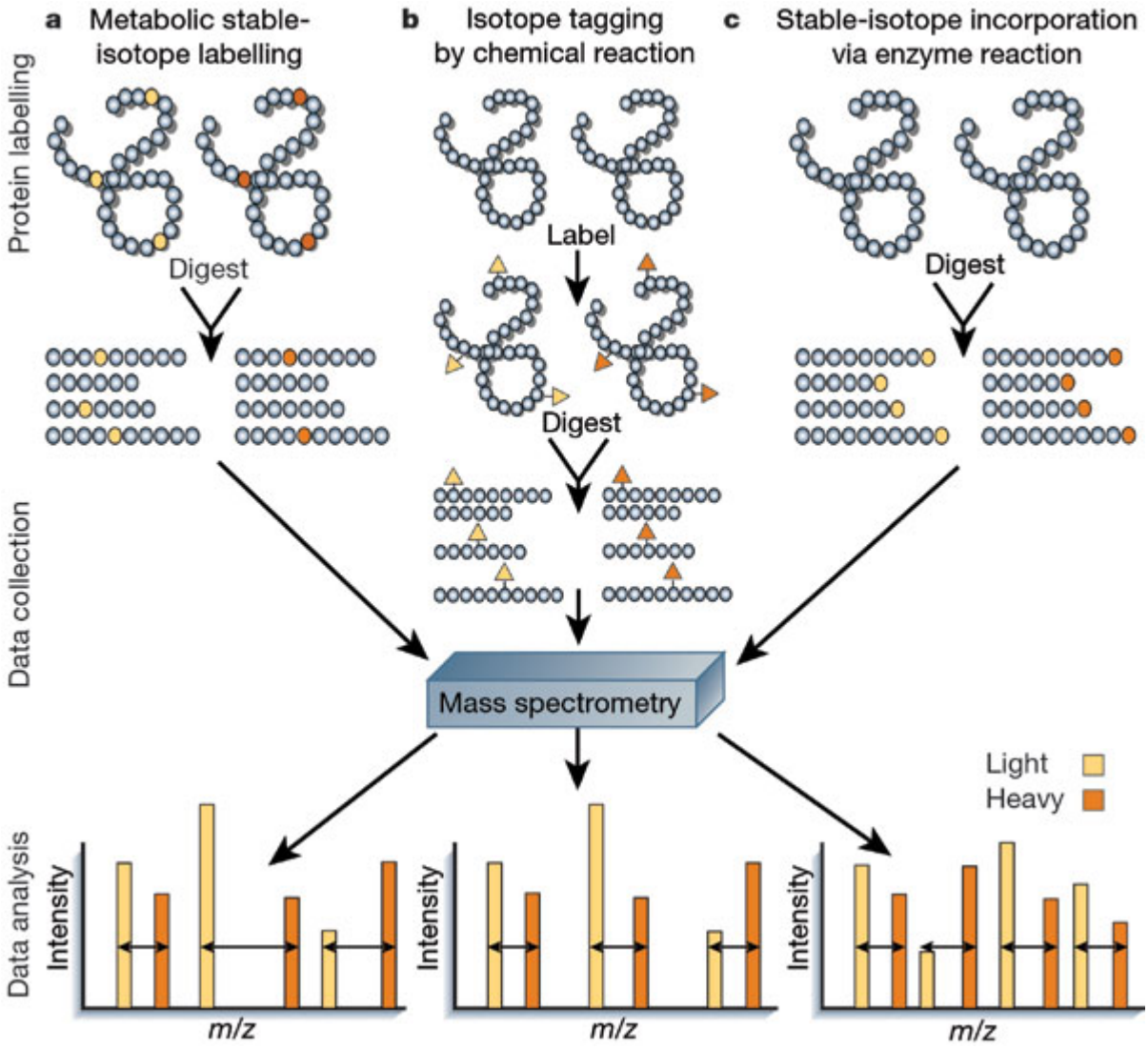


Figure 1.4 Schematic representation of stable-isotope labeling methods [4].

.

In metabolic labeling approach, the proteins in the cells grown under a particular condition and proteins can be uniformly labeled by adding a heavy isotope to the growth media, e.g., with isotope-enriched essential amino acids (SILAC) or salts such as ^{15}N ammonium sulfate. An advantage of metabolic labeling is that the protein lysates are mixed tighter at an early stage in the experiment, minimizing the influence of variations of experimental conditions and improving the precision of quantitation [42]. ^{15}N metabolic labeling is not only been shown to be useful for quantitative proteomics, but also to assist with protein identification. The mass difference between the $^{14}N/^{15}N$ peak pair indicates the number of nitrogen atoms present in the peptide [46, 47], so the nitrogen stoichiometry of peptides can be served as a searching constraint for peptide identification

3. FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETRY (FTICR-MS)

Of all mass spectrometric analyzers, Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS), developed by Comisarow and Marshall [48, 49], offers a unique combination of analytical qualities. FTICR-MS provides the highest resolving power and mass measurement accuracy, which makes it a powerful technology for extremely complex mixture study. The ionization sources used in FTMS are most often an external type, such as the electron-impact (EI) ionization source in the early years and the more recent ESI and MALDI ion sources, and even the combined ESI/MALDI ion source.

3.1 Ion Motion

The ion moving in the presence of a spatially uniform magnetic field (*B*) is subject to a Lorentz force (*F*). The force is described in eq 1, where q is the charge of the ion and *v* is the velocity of the ion. As shown in Figure 1.5, the force is perpendicular to both direction of the ion velocity and the magnetic field, which causes the ion moves in a circular motion, so-called cyclotron motion.

$$F = qv \times B \qquad (1)$$

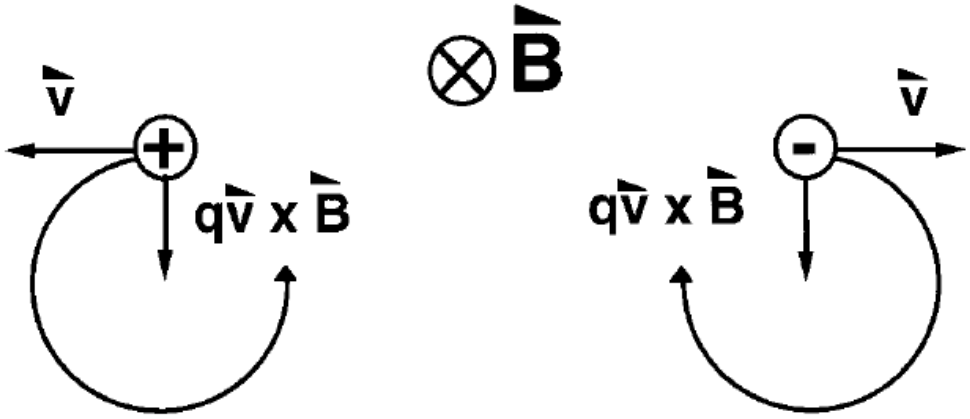


Figure 1.5 Ion cyclotron motion [50].

Cyclotron motion is periodic and is characterized by its cyclotron frequency. The cyclotron frequency (ω_c) of the ion motion can be determined by the magnetic field (*B*) and the mass-to-charge ratio (*m/q*) (eq 2). The purpose of the analyzer cell is to determine the mass-to-charge ratio of the ions trapped in the cell. The remarkable feature of eq 2 is that the cyclotron

frequency is independent of ion velocity, so the cyclotron frequency can be extremely accurate to calculate the *m/q* ratio. For a typical magnetic field of a few Tesla, the cyclotron frequencies of ions range from a few kHz to few MHz for ions of high and low *m/z* values, respectively. The ions also experience two other types of motion, known as trapping motion and magnetron motion, which are discussed elsewhere [51].

$$\omega_c = \frac{qB}{m} \qquad (2)$$

3.2 Ion Excitation and Detection

The heart of FTICR-MS instrument is the analyzer cell, where the excitation and detection of ions occur. There are two types of commonly used ICR cells, cylindrical and cubic cells [52, 53], but generally consists of two trapping electrodes, two excite electrodes and two detection electrodes. The most commonly used commercial ICR cell is cylindrical, which geometry is particularly suitable for fitting into the horizontal bore of a superconducting magnet. As indicated in Figure 1.6, the cylindrical cell consists a pair of endcap electrodes with ~1 volt applied to provide the axial confinement and a cylindrical electrode divided into two opposing pairs of sectors with each pair functioning as excitation and detection plates, respectively.

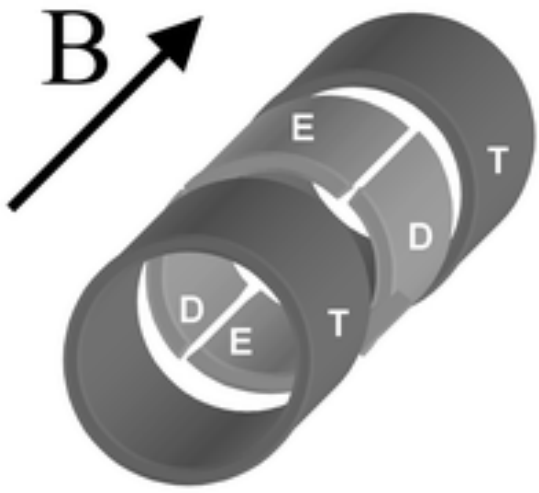


Figure 1.6 FTICR cylindrical cell, showing the two open trap electrodes (*T*), the excitation electrodes (*E*), and the detection electrodes (*D*). The direction of the magnetic field (*B*) is also indicated [54].

Ions that are trapped in the analyzer cell have a relatively small cyclotron orbit (less than 1 mm). In order to detect ions, the ions are exposed to an oscillating electric field that produces a net outward electric force for a limited period of time. This oscillating electric field is created by applying an RF potential to the excitation plates. The ion will only spiral outwards when its cyclotron frequency is in resonance with the applied RF electric field. Ions that are not in resonance do not absorb energy and remain at the center of the cell. To excite all ions trapped in the cell, an RF pulse comprising multiple frequencies is employed, and all ions of different mass-to-charge ratios are exposed to a net outward electric force for the same amount of time.

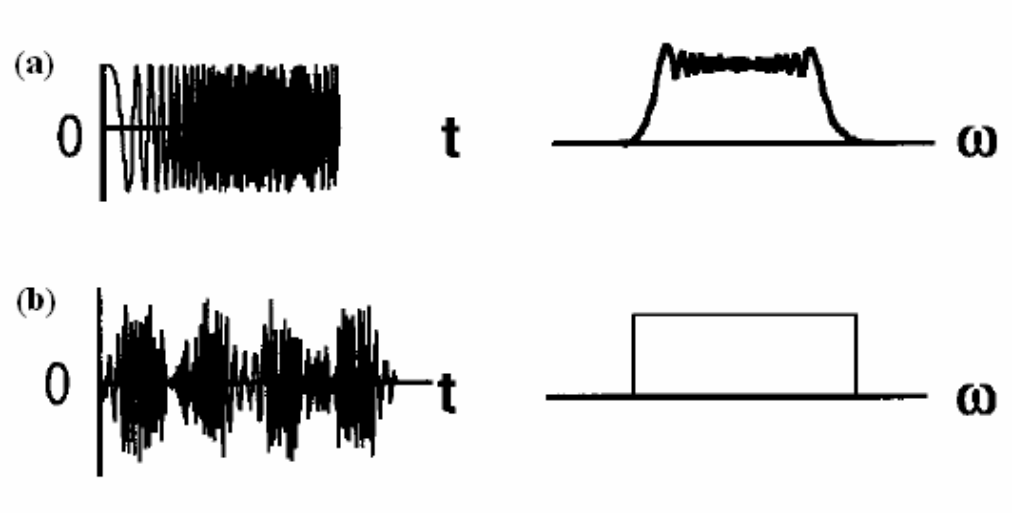


Figure 1.7 Time-domain (left) and frequency-domain (right) excitation waveforms for broadband excitation. (a) Frequency-sweep (chirp). (b) Stored-waveform inverse Fourier transform (SWIFT) waveform [50].

To detect ions of many masses simultaneously (broadband detection), a range of frequencies is applied during the excitation event. A rapid frequency sweep called “chirp” is the most common method for broadband detection (Figure 1.7a). Chirp waveform provides an excitation of a relatively flat magnitude over a broad frequency in a short period of time (1-2 ms). However, the amplitude of chirp excitation waveform is not uniform across the spectrum, and limited mass selectivity at the start and end frequencies of the sweep. Compare to chirp, stored-waveform inverse Fourier transform (SWIFT) [55] waveform excitation shows highly linear excitation/response of ICR, and provides a more uniform power applied across all frequencies leading to a more uniform distribution of radii of gyration for all ions (Figure 1.7b). The utilization of a SWIFT excitation waveform reduces the mass errors significantly compared to chirp excitation, particularly for ions of higher mass-to-charge [56].

Once ions are trapped, all ions of the same mass-to-charge ratio are excited coherently and undergo cyclotron motion as a packet. As they pass the detection electrodes, the packet attracts electrons to the first one and then the other of the two detection plates through the external circuit that joins them. This alternating current is called image current [57] and its detection is a non-destructive process. The signal is then amplified, digitized, and processed to obtain the m/z value. In the case of a mixture of ions with a range of m/z values, all ions are excited to the same orbit simultaneously by applying a broadband excitation waveform. The detected sinusoid is now a mixture of different frequencies and amplitudes. Fast Fourier transformation of the detected signal extracts all the frequency components in real-time and converts them into a mass spectrum with proper calibration.

3.3 Mass Calibration

The image current (transient) that results from ions is a composite of sinusoids of different frequencies and amplitudes, and constitutes the time domain signal. Applying a Fourier transform to the time domain transient yields the frequency spectrum. Then the frequencies are converted into the mass-to-charge values using a calibration equation that is derived from the cyclotron frequency expression. A number of calibration methods have been developed (Table 1.1) to obtain accurate mass measurements [58].

Table 1.1 Proposed calibration procedures [58]

$f = \frac{a}{m}$	Basic law of ions in a B field
$f^2 = \frac{a}{m^2} + \frac{b}{m}$	Beauchamp-Armstrong et al., 1969
$f^2 = \frac{a}{m^2} + \frac{b}{m} + c$	Ledford et al., 1980
$f_{sideband} = \frac{a}{m}$	Allemann et al., 1983
$f = \frac{a}{m} + c$	Francl et al., 1983
$\left(\frac{m}{z}\right) = \frac{a}{f} + \frac{b}{f^2}$	Ledford et al., 1984
$f_{estimated} = f_{measured} + c\left(I_{calibrant} - I_{analyte}\right)$ $\left(\frac{m}{z}\right) = \frac{a}{f_{estimated}} + \frac{b}{f_{estimated}^2} + \frac{c}{f_{estimated}^3}$	Easterling et al., 1999
$M = \left(\frac{kB}{f_n + \Delta f}\right) n - n(M_c)$	Bruce et al., 2000
$\left(\frac{m}{z}\right)_i = \frac{a}{f} + \frac{b}{f^2} + \frac{cI_i}{f^2}$	Masselon et al., 2002
$\frac{m}{z} = \frac{A}{\nu} + \frac{B}{\nu^2} + \frac{C}{\nu^3} + \frac{BC}{A\nu^4}$	Wang et al., 1988

The mostly widely used calibration function [59] is:

$$m/z = \frac{A}{f} + \frac{B}{f^2} \qquad (3)$$

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 [60]. The space charge is assumed to be generated by all ion species present in the ICR cell during collection of the time domain signal. The frequency shift induced by space charge effects depends on changes in the total ion population in the ICR cell during experimental conditions, as compared with the total ion population present during calibration. Although the space charge term is included in eq 3, variations in ion populations severely degrade the ability to predict frequencies of the ions. Thus, the ion population needs to be kept low to reduce the data scattering [61], but spectra under this condition are of poor sensitivity and poor abundance dynamic range. To accommodate higher ion populations, ion intensity has been incorporated as a part of the fitting parameters for the calibration equation:

$$\left(\frac{m}{z}\right)_i = \frac{A}{f_i} + \frac{B}{f_i^2} + C \bullet \frac{I_i}{f_i^2} \qquad (4)$$

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 [61].

Mass calibration procedures employed with MS instrumentation can be separated to internal and external calibration. For FTICR/MS, space-charge is the principal cause of mass measurement error [60, 62]. The best MMA is obtained by using internal calibration, as this eliminates space charge effects [63]. The 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 [61-63], or for matching the ion abundance between the analyte and calibrant spectra, e.g. by automatic gain control (AGC) [64, 65]. However, when a large dynamic range, large ion population is desired, space charge effect irrevocably causes relatively obvious induced frequency shift.

We recently described a two-step external calibration procedure for FTICR, stepwise-external calibration [66], 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. We have demonstrated that sub-ppm MMA can be achieved when SWIFT excitation is combined with stepwise-external calibration [56].

4. OBJECTIVE

The objective of the research work is to develop improved proteomics study. Chapter 2 describes an algorithm, N15Tool, which is able to automatically identify the $^{14}N/^{15}N$ peptide pairs and calculate peptide relative abundance ratios in highly complex mixtures from the proteolytic digest of a whole organism protein extract. This method is demonstrated to provide

over 99% accuracy in assigning peak pair, verified against manual peak picking of the ^{15}N metabolic labeling data. This automated program reduces the time of data analysis from over 100 hours to tens of minutes. Chapter 3 details the improved mass accuracy for higher mass peptides by using SWIFT excitation for MALDI-FTICR mass spectrometry. Mass accuracy obtained by using 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. In chapter 4, a shotgun proteomics approach is described for simultaneous identification and quantitation of the proteins from a proteome using accurate mass measurement and nitrogen stoichiometry. 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/heavy ($^{14}N/^{15}N$) peptide pairs. Combining stepwise-external calibration with SWIFT excitation is applied for MALDI-FTICR/MS analyses and mass measurement accuracy (MMA) is significantly improved. Chapters 5 describes a new calibration method, N15Cal, which corrects the space charge induced frequency shift in FTICR-MS of ^{15}N-metabolic labeled data. There is no need to include an internal calibrant or to apply ion population control. N15Cal has been successfully applied to the LC-MALDI-FTICR data of ^{15}N-metabolic labeling proteomics. Finally chapter 6 recounts conclusions pertinent to the work carried out and our approaches to shotgun proteomics.

REFERENCES

1 Haynes, P. A.; Gygi, S. P.; Figeys, D.; Aebersold, R. Proteome Analysis: Biological Assay or Data Archive? *Electrophoresis* **1998**, *19*, 1862-1871.

2 Jung, E.; Heller, M.; Sanchez, J.-C.; Hochstrasser, D. F. Proteomics Meets Cell Biology: The Establishment of Subcellular Proteomes. *Electrophoresis* **2000**, *21*, 3369-3377.

3 Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I. Progress with Gene-Product Mapping of the Mollicutes: Mycoplasma Genitalium. *Electrophoresis* **1995**, *16*, 1090-1094.

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

5 Patterson, S. D. How Much of the Proteome Do We See with Discovery-Based Proteomics Methods and How Much Do We Need to See? *Curr. Proteomics* **2004**, *1*, 3-12.

6 Bleakney, W. A New Method of Positive Ray Analysis and Its Application to the Measurement of Ionization Potentials in Mercury Vapor. *Phys. Rev.* **1929**, *34*, 157-160.

7 Munson, M. S. B.; Field, F. H. Chemical Ionization Mass Spectrometry. I. General Introduction. *J. Am. Chem. Soc.* **1966**, *88*, 2621-2630.

8 Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60*, 2299-2301.

9 Fenn, J. B. M., Matthias; Meng, Chin Kai; Wong, Shek Fu; Whitehouse, Craig M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246*, 64-71.

10 Karas, M.; Bachmann, D.; Hillenkamp, F. Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Anal. Chem.* **1985**, *57*, 2935-2939.

11 Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. Protein and Polymer Analyses up to *m/z* 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.

12 Gluckmann, M.; Karas, M. The Initial Ion Velocity and Its Dependence on Matrix, Analyte and Preparation Method in Ultraviolet Matrix-Assisted Laser Desorption/Ionization. *J. Mass Spectrom.* **1999**, *34*, 467-477.

13 Karas, M.; Bahr, U.; Gieumann, U. Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Mass Spectrom. Rev.* **1991**, *10*, 335-357.

14 Burton, R. D.; Watson, C. H.; Eyler, J. R.; Lang, G. L.; Powell, D. H.; Avery, M. Y. Proton Affinities of Eight Matrices Used for Matrix-Assisted Laser Desorption/Ionization. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 443-446.

15 Dale, M.; Knochenmuss, R.; Zenobi, R. Two-Phase Matrix-Assisted Laser Desorption/Ionization: Matrix Selection and Sample Pretreatment for Complex Anionic Analytes. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 136-142.

16 Schurenberg, M.; Dreisewerd, K.; Hillenkamp, F. Laser Desorption/Ionization Mass Spectrometry of Peptides and Proteins with Particle Suspension Matrixes. *Anal. Chem.* **1999**, *71*, 221-229.

17 Cech, N. B.; Enke, C. G. Practical Implications of Some Recent Studies in Electrospray Ionization Fundamentals. *Mass Spectrom. Rev.* **2001**, *20*, 362-387.

18 Mora, J. F. d. l.; Berkel, G. J. V.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. Electrochemical Processes in Electrospray Ionization Mass Spectrometry. *Journal of Mass Spectroetry* **2000**, *35*, 939-952.

19 Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. New Developments in Biochemical Mass Spectrometry: Electrospray Ionization. *Anal. Chem.* **1990**, *62*, 882-899.

20 Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R. Principles and Practice of Electrospray Ionization - Mass Spectrometry for Large Polypeptides and Proteins. *Mass Spectrom. Rev.* **1991**, *10*, 359-452.

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

22 Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*; MCC Press: San Diego, 2003.

23 Sanz-Medel, A. Analytical Atomic Spectrometry Going into the Next Millennium: Photons or Ions, Atoms or Molecules? *The Analyst* **2000**, *125*, 35-43.

24 Siuti, N.; Kelleher, N. L. Decoding Protein Modifications Using Top-Down Mass Spectrometry. *Nature methods* **2007**, *4*, 817-821.

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

26 Kettman, J. R.; Frey, J. R.; Lefkovits, I. Proteome, Transcriptome and Genome: Top Down or Bottom up Analysis? *Biomol Eng* **2001**, *18*, 207-212.

27 Chait, B. T. Chemistry: Mass Spectrometry: Bottom-up or Top-Down? *Science* **2006**, *314*, 65-66.

28 Liu, H.; 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 Brock, A.; Horn, D. M.; Peters, E. C.; Shaw, C. M.; Ericson, C.; Phung, Q. T.; Salomon, A. R. An Automated Matrix-Assisted Laser Desorption/Ionization Quadrupole Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for "Bottom-Up" Proteomics. *Anal. Chem.* **2003**, *75*, 3419-3428.

30 Santoni, V.; Molloy, M.; Rabilloud, T. Membrane Proteins and Proteomics: Un Amour Impossible? *Electrophoresis* **2000**, *21*, 1054-1070.

31 O'Farrell, P. H. High Resolution Two-Dimensional Electrophoresis of Proteins. *J. Biol. Chem.* **1975**, *250*, 4007-4021.

32 Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. Correlation between Protein and mRNA Abundance in Yeast. *Mol. Cell. Biol.* **1999**, *19*, 1720-1730.

33 Fountoulakis, M.; Takacs, M.-F.; Berndt, P.; Langen, H.; Takacs, B. Enrichment of Low Abundance Proteins of Escherichia Coli by Hydroxyapatite Chromatography. *Electrophoresis* **1999**, *20*, 2181-2195.

34 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.

35 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.

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

37 Yates, J. R.; Eng, J. K.; McCormack, A. L.; Schieltz, D. Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **1995**, *67*, 1426-1436.

38 Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551-3567.

39 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.

40 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.

41 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.

42 Ong, S.-E.; Mann, M. Mass Spectrometry-Based Proteomics Turns Quantitative. *Nature Chemical Biology* **2005**, *1*, 252-262.

43 Conrads, T. P.; Issaq, H. J.; Veenstra, T. D. New Tools for Quantitative Phosphoproteome Analysis. *Biochem. Biophys. Res. Commun.* **2002**, *290*, 885-890.

44 Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic 18O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001**, *73*, 2836-2842.

45 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 Biotechnol.* **1999**, *17*, 994-999.

46 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.

47 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*.

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

49 Comisarow, M. B.; Marshall, A. G. Frequency-Sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974**, *26*, 489-490.

50 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.

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

52 Campbell, V. L.; Guan, Z.; Vartanian, V. H.; Laude, D. A. Cell Geometry Considerations for the Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Remeasurement Experiment. *Anal. Chem.* **1995**, *67*, 420-425.

53 Beu, S. C.; Laude, D. A. Open Trapped Ion Cell Geometries for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1992**, *112*, 215-230.

54 Heeren, R. M. A.; Kleinnijenhuis, A. J.; McDonnell, L. A.; Mize, T. H. A Mini-Review of Mass Spectrometry Using High-Performance FTICR-MS Methods. *Analytical and Bioanalytical Chemistry* **2004**, *378*, 1048-1058.

55 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.

56 Jing, L.; Li, C.; 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.

57 Comisarow, M. B. Signal Modeling for Ion Cyclotron Resonance. *J. Chem. Phys.* **1978**, *69*, 4097-4104.

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

59 Ledford, E. B.; Rempel, D. L.; Gross, M. L. Space Charge Effects in Fourier Transform Mass Spectrometry. II. Mass Calibration. *Anal. Chem.* **1984**, *56*, 2744-2748.

60 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.

61 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. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 99-106.

62 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.

63 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.

64 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.

65 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.

66 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.

CHAPTER 2

RAPID AND AUTOMATED PROCESSING OF MALDI-FTICR/MS DATA FOR ^{15}N-METABOLIC LABELING IN A SHOTGUN PROTEOMICS ANALYSIS

Li Jing and I. Jonathan Amster, submitted to *Int. J. Mass Spectrom.*

ABSTRACT

Offline high performance liquid chromatography combined with matrix assisted laser desorption and Fourier transform ion cyclotron resonance mass spectrometry (HPLC-MALDI-FTICR/MS) provides the means to rapidly analyze complex mixtures of peptides, such as those produced by proteolytic digestion of a proteome. This method is particularly useful for making quantitative measurements of changes in protein expression by using ^{15}N-metabolic labeling. Proteolytic digestion of combined labeled and unlabeled proteomes produces complex mixtures that with many mass overlaps when analyzed by HPLC-MALDI-FTICR/MS. A significant challenge to data analysis is the matching of pairs of peaks which represent an unlabeled peptide and its labeled counterpart. We have developed an algorithm and incorporated it into a compute program which significantly accelerates the interpretation of ^{15}N metabolic labeling data by automating the process of identifying unlabeled/labeled peak pairs. The algorithm takes advantage of the high resolution and mass accuracy of FTICR mass spectrometry. The algorithm is shown to be able to successfully identify the $^{15}N/^{14}N$ peptide pairs and calculate peptide relative abundance ratios in highly complex mixtures from the proteolytic digest of a whole organism protein extract.

INTRODUCTION

Various stable isotope labeling strategies are being used today in the field of comparative proteomics [1]. The methods of stable isotope labeling can generally be placed in three classes, metabolic labeling [2-4], chemical reaction [5, 6] and enzyme reaction [7, 8]. The differentially labeled samples are then combined and concurrently processed and analyzed. In metabolic labeling method, the proteins in the cells grown in a particular media can be uniformly labeled

by including the isotope of interest as an inorganic compound [2, 4] or already incorporated into amino acids [3]. The mass difference between unlabeled and labeled peptides depends upon the type of stable isotope labeling that is employed. The chemical labeling methods and enzyme reactions usually produce a fixed mass difference between unlabeled and labeled peptides, which facilitates the matching of the peak pairs in a mass spectrum. On the other hand, metabolic labeling usually results in a variable mass difference, such as in the case of ^{15}N-labeling, in which the mass difference between unlabeled and labeled peptides is very close to 1 u per nitrogen atom in the elemental composition.

Efforts in our laboratory are directed at developing high-throughput proteomic analysis methods that allow protein identification and quantitation in the shortest possible time. Reversed-phase liquid chromatography (LC) coupled with Fourier transform ion cyclotron resonance mass spectrometry (FTICR) has been shown to provide high confidence in peptide identification, increased sensitivity, and higher throughput than methods that rely on repetitive MS/MS identification [9, 10]. To measure differences in relative protein abundances, proteomes with two different stable-isotope compositions are combined and analyzed. In this study, *Methanococcus maripaludis* was grown in both normal (^{14}N minimal media) and heavy media (^{15}N minimal media). In the mass spectra, peptides appear as light and heavy pairs of peaks, and the spacing between a heavy/light peak pair corresponds to the number of nitrogen atoms present in the peptide times the mass difference between ^{14}N and ^{15}N, namely 0.9970 u, as shown in Figure 2.1. Thus, ^{15}N metabolic labeling in conjunction with FTICR not only allows quantitative studies, but also provides the nitrogen stoichiometry of the peptides [2], which serves as a search constraint to increase the peptide identification rate. For the organism *M. maripaludis*, with 1722 open reading frames, 15% of the tryptic peptides can be identified by accurate mass measurement with

a 10 ppm mass tolerance, while 43% can be identified if nitrogen stoichiometry is used as a search constraint.

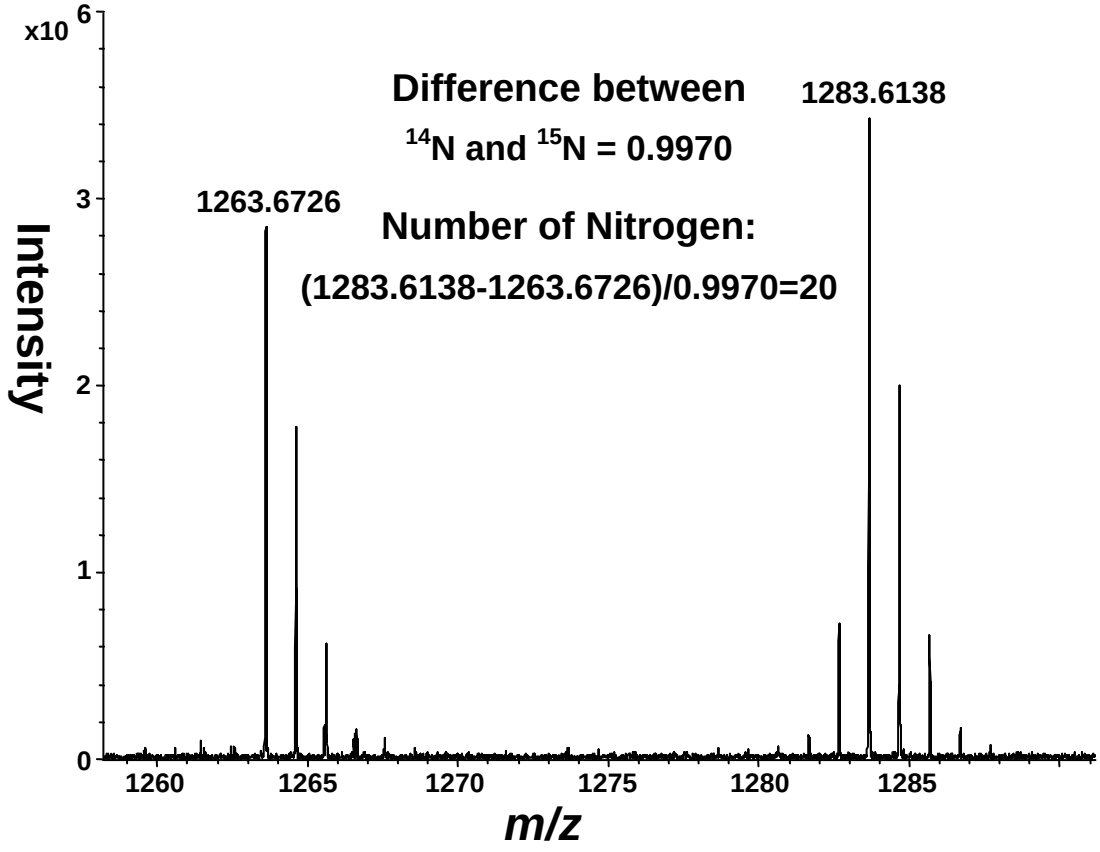


Figure 2.1 $^{15}N/^{14}N$ peak pair in a MALDI-FTICR mass spectrum of proteolytic peptides from *M. maripaludis* grown in normal media and ^{15}N-enriched (99%) media.

At the outset of the development of HPLC-FTICR/MS for shotgun proteomics, the most significant bottleneck in the analysis was manual peak picking data from a ^{15}N metabolic labeling data set. While the most abundant pairs are easily assigned by visual inspection, there

are regions in most mass spectra that are quite congested with moderate and low abundance peptides which overlap each other in mass-to-charge. For a typical proteomic analysis of *M. maripaludis*, 50-100 fractions are collected and each fraction is analyzed by MALDI-FTICR/MS. On average, it takes 2-3 hours to manually pick a single mass spectrum, requiring 150-200 hours for analyzing one data set. Manual interpretation of such complex data is very tedious and time consuming. Several algorithms have been developed to process the stable isotope labeling data [11-14]. However, these algorithms are mostly focused on the quantitative aspects of the measurement, and the peptide identification is based on MS/MS. Until now there has been no development of software that might be used to aid the identification of $^{15}N/^{14}N$ peptide pairs from ^{15}N metabolic labeling experiments. To assist with the study of $^{15}N/^{14}N$ peptides in mass spectra, we have devised and implemented a new computer algorithm. It has been incorporated into a program we call N15Tool, which is able to automatically identify and quantify $^{15}N/^{14}N$ peptide pairs from the mass spectra using *m/z* and peak intensity data directly.

The objective of the work described here is to first consider the challenge presented in analyzing a typical MALDI-FTICR mass spectrum of a tryptic digest of a ^{15}N-metabolically labeled proteome, and then to develop an algorithm that identifies pairs of ^{15}N labeled components, and calculates the ratio of stable isotope coded peptides. The algorithm for pair assignment is based on finding peaks that are separated by an integer value times the $^{15}N/^{14}N$ mass spacing of 0.9970 amu, within a narrow range of error. This method will be demonstrated to provide over 99% accuracy in assigning peak pair, verified against manual peak picking. In addition, light and heavy labeled peaks are compared and normalized to provide relative differential quantitative information at peptide level. This automated program reduces the time of data analysis from over 100 hours to tens of minutes.

EXPERIMENTAL

Differential Labeling ($^{15}N/^{14}N$) of Methanococcus maripaludis

The analyzed proteome was a whole cell lysate extracted from *M. maripaludis*, which was grown on minimal media with ammonium sulfate as the sole source of nitrogen. *M. maripaludis* Δ*lrp* mutant (S102) and wild-type (S2) were cultured in midlogarithmic and stationary stages in ammonium sulfate with naturally occurring isotopic composition (99.6% ^{14}N, 0.4% ^{15}N) and with ^{15}N-enriched composition (>99% ^{15}N), respectively. 40mL of culture grown with ^{14}N mixed with 40 mL of culture grown with ^{15}N to form the cell mixtures. The cell mixtures were centrifuged at 10,000 x g for 30 min at 4 ^{o}C and followed by lysis with a French pressure cell. DNA was digested by DNase and incubated for 15 min at 37 ^{o}C. Cell debris and the unbroken cells were removed by centrifugation at 8000 x g for 30 min at 4 ^{o}C.

The proteome sample was prepared in alkaline solution (10 mM ammonium bicarbonate) at ~1mg/mL concentration 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).

HPLC-MALDI-FTICR/MS Analysis

Separation of peptide mixture were carried out on an UltiMate Plus HPLC system by using a 75 μm i.d. x 15cm C18 nanocloumn with 3 μm particles and 100 Å pore size (Dionex LC packings, Sunnyvale, CA). The mobile phase was water and acetonitrile. Peptides were eluted with increasing acetonitrile (5-80% in 90 min) at approximate flow rate of 300 nL/min. Eluate was detected by UV absorption at 214 nm and was collected onto a stainless steel MALDI target

at a 60-s interval using a Probot Micro Fraction collector (Dionex LC packings, Sunnyvale, CA). Samples were allowed to dry before adding 400 nL of 1 M 2,5-Dihydroxybenzoic acid (DHB) (Lancaster, Pelham, NH) as the MALDI matrix.

Samples were analyzed on a 7-T FTICR mass spectrometer equipped with an intermediate pressure Scout 100 MALDI source (Bruker Daltonics Inc, Billerica, MA). Conditions for operation of the FTICR/MS were similar to those reported previously [15] with minor modifications. The ions generated from 6 MALDI laser shots per scan and 12 scans were co-added for each spectrum. The external mass calibration was established using a peptide mixture generated by tryptic digestion of chicken egg albumin (Sigma, St. Louis, MO).

RESULTS AND DISCUSSION

Algorithm Description

The algorithm for peak matching has been incorporated into a Java program, N15Tool (available from the authors upon request) which consists of three major steps involving data input, processing and output. The input to the program is a text file with two entries per line; the mass value for the monoisotopic peak (charge deconvolution is required for multiply-charged ions) and signal intensity. For the work published here, assignment of monoisotopic peaks for the ^{14}N and ^{15}N labeled peptides (S/N > 3) was performed automatically using the SNAP program in the DataAnalysis software package (version 3.4 build 184, Bruker Daltonics Inc, Billerica, MA). The peak list from each spectrum, sorted by mass-to-charge value, was saved in a general text format, and serve as input to the N15Tool program.

For the analysis, the mass spectrum of each collected fraction is processed separately. The labeled and unlabeled versions of a peptide are not separated by chromatography, and so

both will appear in the same fraction. Thus, it is advantageous to the matching procedure that the data from separate fractions are not combined before the assignment of peak pairs is made.

The algorithm separates the mass range (*m/z* 700-4000) into 100 *m/z* wide segments. For each segment, the minimum and maximum number of nitrogen atoms per peptide is calculated in advance by analysis of *in silico* digestion of all entries in the protein database for the organism of interest. The algorithm begins with the lowest *m/z* value of the first input text file, and proceeds by comparing the mass difference with successively larger *m/z* values, until a match is found. In order to be considered as a potential heavy isotope companion, the mass difference to the second peak must lie between the computed minimum and maximum number of nitrogen atoms for the segment corresponding to the mass-to-charge value of the peak. Furthermore, to be assigned as a light/heavy peptide pair, the mass-to-charge spacing between two peaks must closely match an integer value times the $^{14}N/^{15}N$ mass difference of 0.9970 u. That is, for $n=\Delta m/0.997$, where Δm is the mass difference between peaks, n is the calculated number of nitrogen atoms in the peptide, and it must lie within ±0.006 of an integer value for peptides of *m/z* 1000. The tolerance value scales with mass, and is equal to approximately 6 ppt of the mass of the peptide in question. The entry for an identified ^{15}N-labeled peak is then removed from the list so that it will not be selected as a ^{14}N peak. If no suitable corresponding ^{15}N peak is found within the maximum spacing predicted for the segment, the peak is discarded, and processing of the list continues with the next peak, which is then treated as an unlabeled peptide. This procedure is repeated until all peaks in the file for the HPLC fraction have been processed. The procedure is automatically repeated for each input file, corresponding to each separate HPLC fraction.

Quantitation Issues

As each $^{15}N/^{14}N$ peptide pair is identified, the relative abundance ratio is calculated by the program and stored with the monoisotopic mass of the unlabeled peptide and its nitrogen stoichiometry value. A text format output file is created in which each line contains the three fields of information (mass, nitrogen stoichiometry, and abundance ratio.) ^{15}N metabolic labeling does not cause chromatographic isotope effects, so the peak ratio calculated in each spectrum should preserve the approximation of quantitative information. However, the monoisotopic peak intensities of the $^{15}N/^{14}N$ peptide pair are not equal for a 1:1 mixture because the isotopic peak distribution shape of the same peptide from a sample grown in ^{14}N media (99.6 atom %) and ^{15}N media (97-99 atom %) are not exactly identical, and depend on the exact isotopic enrichment value and the mass of peptide. Also, it is not possible to create a mixture in which proteins from both samples are present in exactly equal amounts, and so even a carefully measured ratio will deviate from 1:1. To determine an accurate ratio of abundances, both these issues must be addressed. The former issue (determining the enrichment value) is easily addressed by matching the isotope peak distribution of abundant labeled peptides with calculated isotope distributions, in which ^{15}N enrichment is varied to produce the best fit. The latter issue, determining the exact amount of unlabeled and labeled proteins in the original mixture, is resolved by plotting the distribution of abundances, and taking the central value as the true mixing ratio. This value is used to normalize all the measured abundance ratios, so that a value of 1.0 corresponds to no change in protein expression between the control and stressed proteomes.

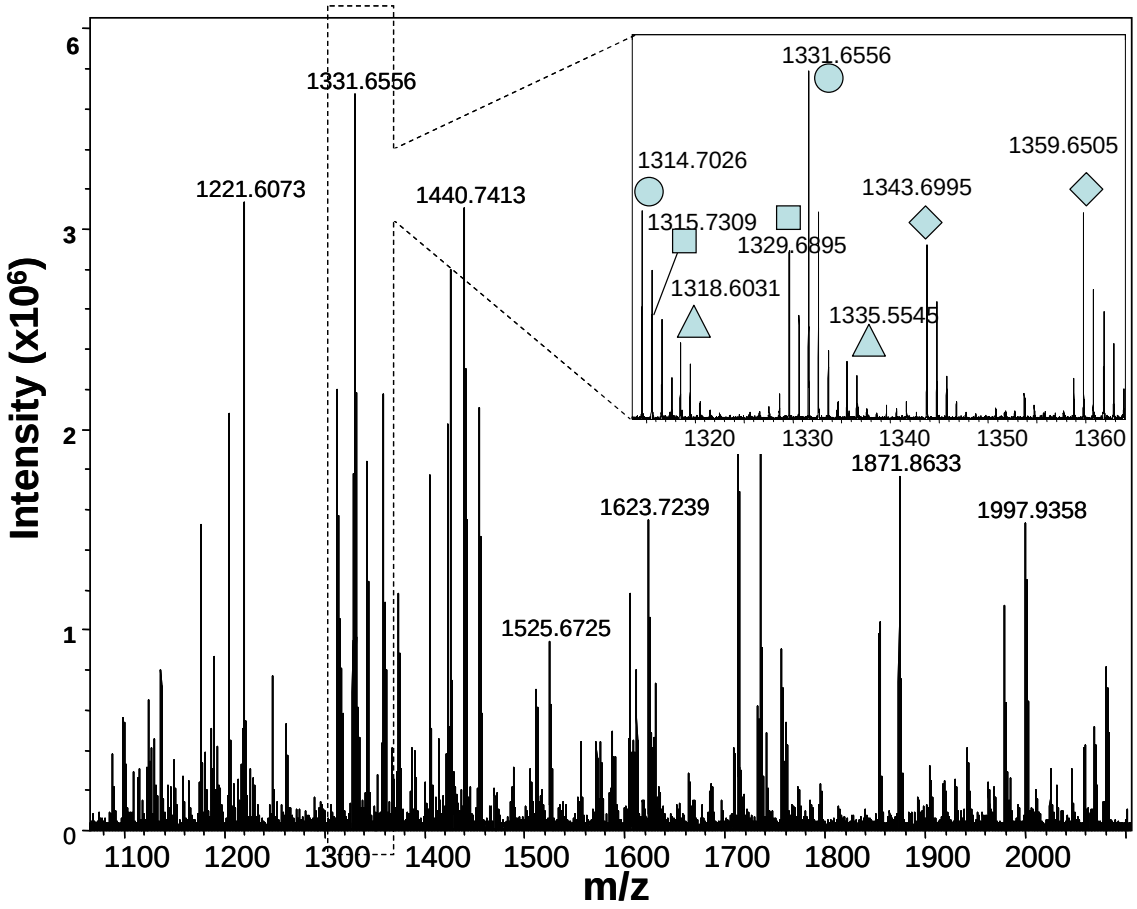


Figure 2.2 MALDI-FTICR mass spectrum of a HPLC fraction from the analysis of a tryptic digest of a proteome with ^{15}N-metabolical labeling. The inset shows an expansion of a congested region of the mass spectrum. Corresponding unlabeled and labeled peptide are marked with the same symbol.

For the most accurate measurement of abundance, the entire isotope envelope should be used [16]. Although summing the entire isotope distribution for a peptide is the most accurate way to measure its abundance, it cannot be accomplished for highly congested mass spectra where there are significant overlaps in mass, as in Figure 2.2. For the proteome experiments, we

use only the monoisotopic peak intensity for determining peptide abundance ratios between labeled and unlabeled pairs. The ratio of monoisotopic peak intensities for peptide pairs ($^{15}N/^{14}N$) varies systematically with the molecular weight of a peptide, and depends also on the exact enrichment value for ^{15}N-labeling. Thus, normalization is required for accurate relative abundance determination. The normalization value and its mass dependence is determined from calculated isotope distributions using the elemental composition of averagine [17]. Errors can occur when a peptide contains different number of sulfur atoms from the number of sulfur atoms predicted by averagine. More accurate quantitative analysis can be done if this peptide results in an unique protein identification, as the specific elemental composition of the peptide can be used to fit the isotopic pattern.

Evaluation of the Algorithm Performance for a Proteome Analysis

The performance of the N15Tool algorithm was tested for a proteome analysis using ^{15}N-metabolic labeling in *M. maripaludis*. Unlabeled and labeled protein extracts were mixed in approximately equal amounts, digested with trypsin, and analyzed by HPLC-MALDI-FTICR/MS. In this experiment, 90 fractions from the HPLC separation were collected directly onto a MALDI target, and then mass analyzed. To test the accuracy of the assignment of $^{15}N/^{14}N$ peptide pairs, all spectra were interpreted manually, a task that consumed about 10 days. In contrast, the same data were analyzed by the N15tool program in about 15 minutes. Of the 2321 peptides identified from the N15Tool, more than 99% of them were also identified in the manually-interpreted data set.

The robustness of the algorithm under the realistic test conditions is remarkable given its simplicity. For example, each time an unlabeled/labeled pair is identified, and the peaks are

removed from the list, the algorithm assumes that that the next peak in the list is an unlabeled peptide. If this assumption is wrong, the procedure will be out of registry, treating labeled peptides as unlabeled peptides. In developing the algorithm, there was concern that such an event would lead to an irrecoverable event in which all subsequent matches in an input file would be invalid. However, in practice, the algorithm recovers from such situations, and quickly finds the next unlabeled peptide to use as the base for finding the matching labeled peptide. We believe that the robustness results from the very precise requirement for the mass spacing between unlabeled and labeled peptides, i.e. for $n=\Delta m/0.997$, n must lie within ±0.006 of an integer value. It appears that when labeled peptides are accidentally treated as unlabeled peptides, no higher mass peak is found to fit the matching constraints. The algorithm discards these unmatched peaks, and thus moves forward to the next peak in the list, and gets back in registry with the unlabeled versus labeled peptides in the list. The ability to recover from situations in which a pair match cannot be established is also useful for other events, for example, when protein expression between the unlabeled and labeled sample changes to such a large extent that only one peak is observed for the unlabeled/labeled peptide pair; or in the case where two peptides overlap in mass sufficiently to be unresolved from each other. Although one can imagine many possible ways in which the algorithm could fail, in practice it recovers gracefully from such events and continues to assign isotopic peak pairs. The robustness of this procedure has been established by its application to hundreds of peak lists and tens of thousands of peptide pairs. The algorithm is particularly remarkable in being able to match peptide pairs in portions of the mass spectrum that are highly congested, as shown in Figure 2.2. The three lowest mass peptide pairs shown in the figure inset would present a significant challenge to manual interpretation, given the mass overlaps of the ions. For example, peptides of nominal mass 1314 and 1315

produce a doublet m/z 1315, with a separation of 0.025 mass-to-charge units. The SNAP algorithm assigns the proper monoisotopic masses to both peaks, and our pair matching algorithm finds a unique set of labeled peaks to assign to these two peptides.

The pair matching algorithm does not rely on differences in the isotopic distribution of labeled versus unlabeled peptides. When the matching process was performed by hand, we used lower ^{15}N-enrichment (97-98 mol %) to help visually distinguish labeled from unlabeled peptides, as shown in Figure 2.3. While this aids manual interpretation, it detracts from automated interpretation and reduces the accuracy of the quantitative measurement. The software that is used for assigning monoisotopic peaks (SNAP, *vide supra*) does not perform well on labeled peptides with isotope distributions that vary significantly from unlabeled peptides. Labeling at 99% enrichment significantly facilitates that automated assignment of monoisotopic peaks for labeled peptides by the SNAP software.

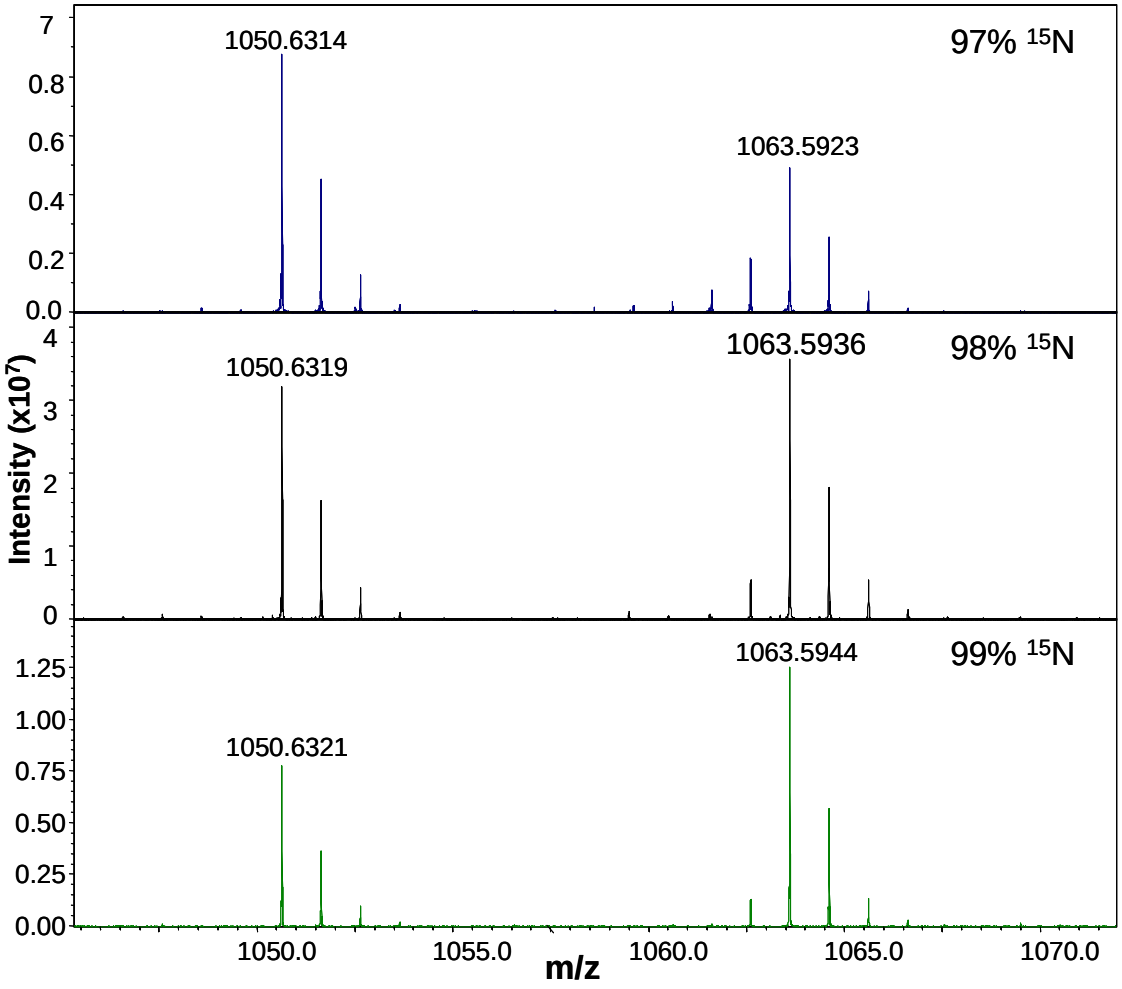


Figure 2.3 The effect of small changes in isotopic enrichment on the isotope distribution of ^{15}N-metabolically labeled peptides is demonstrated for a peptide pair observed in three metabolic labeling experiments, with 97%, 98%, and 99% ^{15}N enrichment.

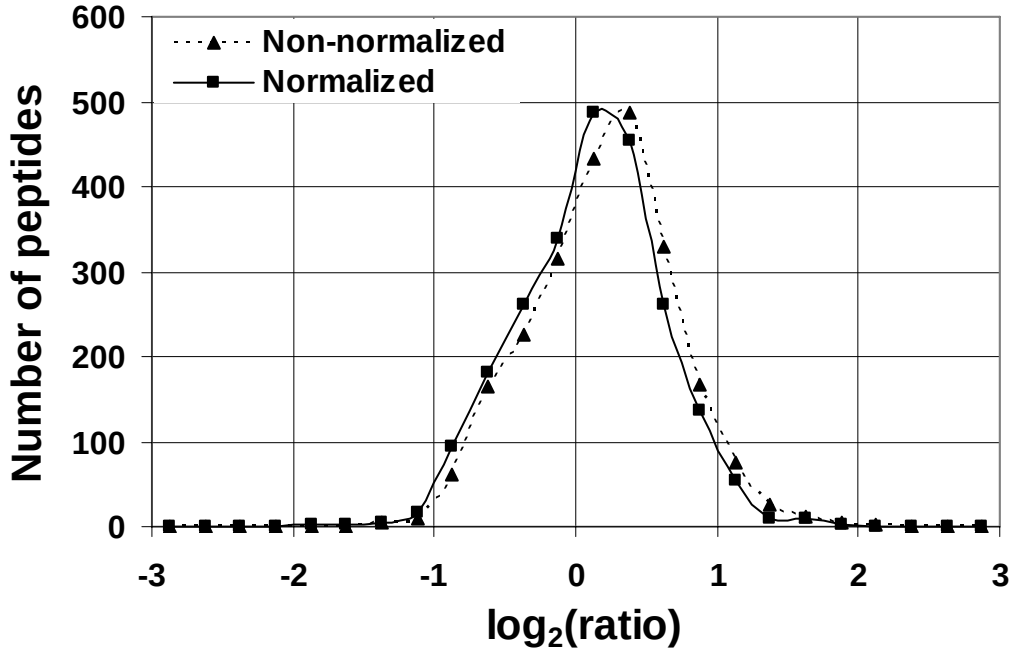


Figure 2.4 Distribution of 2321 heavy/light peptide log (base 2) ratio from an LC-MALDI-MS experiment. The dashed line depicted the peak ratio calculated directly from peak height. The solid line is the normalized ratio calculated using averagine.

To test the effectiveness of the normalization of abundance ratio, non-normalized ratios were compared with the normalized ratios for the same data set. The sample contains equal amount wild type and mutant type of *M. maripaludis*, so the overall relative abundance between the light and heavy peptides is expected to be 1:1. Figure 2.4 shows the distribution of 2321 heavy/light peptide logarithm (base 2) ratio. In this case, the average logarithm (base 2) ratio for these 2321 peptides is 0.1885 if the peak ratio is obtained by direct calculation from the monoisotopic peak height. This value is changed to 0.0846 when normalization is applied by N15tool. The normalized relative abundance of the majority peak pairs indicates a 1:1 ratio

Table 2.1 Partial list of the output from N15Tool.

	^{14}N data		^{15}N data			
File name	*m/z*	Intensity	*m/z*	Intensity	# of N	H/L
45	1315.7272	240946	1329.6846	241305	14	0.99
45	1405.6332	200980	1423.5796	205302	18	1.01
45	1440.7382	211082	1455.6964	96282	15	0.45
45	1610.8758	121238	1629.8221	82527	19	0.66
45	1739.8073	160292	1759.7477	156684	20	0.93
45	1891.0208	80398	1915.9483	115965	25	1.35
45	2007.9773	87532	2027.9152	99054	20	1.03
45	2577.2121	425700	2610.1169	44630	33	0.88

The program produces only one output file even though the number of input files is nearly one hundred. A portion of results from the output file is shown in Table 2.1. Each row of the output contains the information of one identified peptide pair. The output file is a tab delimited text file including the information of the input file name (fraction number), identified peak pairs, and intensity ratio, suitable for import into Excel or other programs. Collectively, these data indicate that the N15Tool program is capable of generating reliable light/heavy peptide assignment and abundance ratios automatically.

CONCLUSIONS

We have developed a simple but robust algorithm that aids the process of ^{15}N metabolic labeling data resulting from LC-MS. This algorithm has been incorporated into a Java program that can process data from a variety of mass spectrometer platforms. It provides an automated method for identifying ^{14}N and ^{15}N peak pairs and calculating the ratios between the light and heavy peptides without prior knowledge of amino acid composition of the peptides. The program significantly reduces the burden of manually interpreting large scale LC-MS data sets from ^{15}N labeling experiments. The current version of the program has been tested for the proteome from *M. maripaludis*. Nevertheless, applications should be expanded to MS based experiments of any other organism with ^{15}N metabolic labeling by apply new theoretical nitrogen numbers of each *m/z* segment in the program. N15Tool is platform independent and can be used in any operating system have the Java runtime installed. It is also proprietary vendor format independent because the input for the software is text files containing the *m/z* and signal intensity information. Finally, the flexibility and versatility of N15Tool allows it to be used with any ^{15}N metabolic labeling data from any instrument that can provide accurate mass measurement, including the Orbitrap or high resolution time-of-flight mass spectrometers.

REFERENCES

1. S.-E. Ong, M. Mann, Nat. Chem. Biol. 1 (2005) 252-262.
2. T.P. Conrads, K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. Udseth, W.B. Chrisler, B.D. Thrall, R.D. Smith, Anal. Chem. 73 (2001) 2132-2139.

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

4. M.P. Washburn, R. Ulaszek, C. Deciu, D.M. Schieltz, J.R. Yates, Anal. Chem. 74 (2002) 1650-1657.

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

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

7. X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836-2842.

8. A. Ramos-Fernandez, D. Lopez-Ferrer, J. Vazquez, Mol. Cell. Proteomics 6 (2007) 1274-1286.

9. R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, H.R. Udseth, PROTEOMICS 2 (2002) 513-523.

10. T. Liu, M.E. Belov, N. Jaitly, W.-J. Qian, R.D. Smith, Chem. Rev. 107 (2007) 3621-3653.

11. X. Li, H. Zhang, J.A. Ranish, R. Aebersold, Anal. Chem. 75 (2003) 6648-6657.

12. M. Cannataro, G. Cuda, M. Gaspari, S. Greco, G. Tradigo, P. Veltri, BMC Bioinformatics 8 (2007) 255.

13. C.J. Mason, T.M. Therneau, J.E. Eckel-Passow, K.L. Johnson, A.L. Oberg, J.E. Olson, K.S. Nair, D.C. Muddiman, H.R. Bergen, III, Mol. Cell. Proteomics 6 (2007) 305-318.

14. V.P. Andreev, L. Li, T. Rejtar, Q. Li, J.G. Ferry, B.L. Karger, J. Proteome Res. 5 (2006) 2039-2045.

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

16. M.J. MacCoss, C.C. Wu, D.E. Matthews, J.R. Yates, Anal. Chem. 77 (2005) 7646-7653.

17. M.W. Senko, S.C. Beu, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 6 (1995) 229-233.

CHAPTER 3

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

Li Jing*, Chunyan Li*, 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].

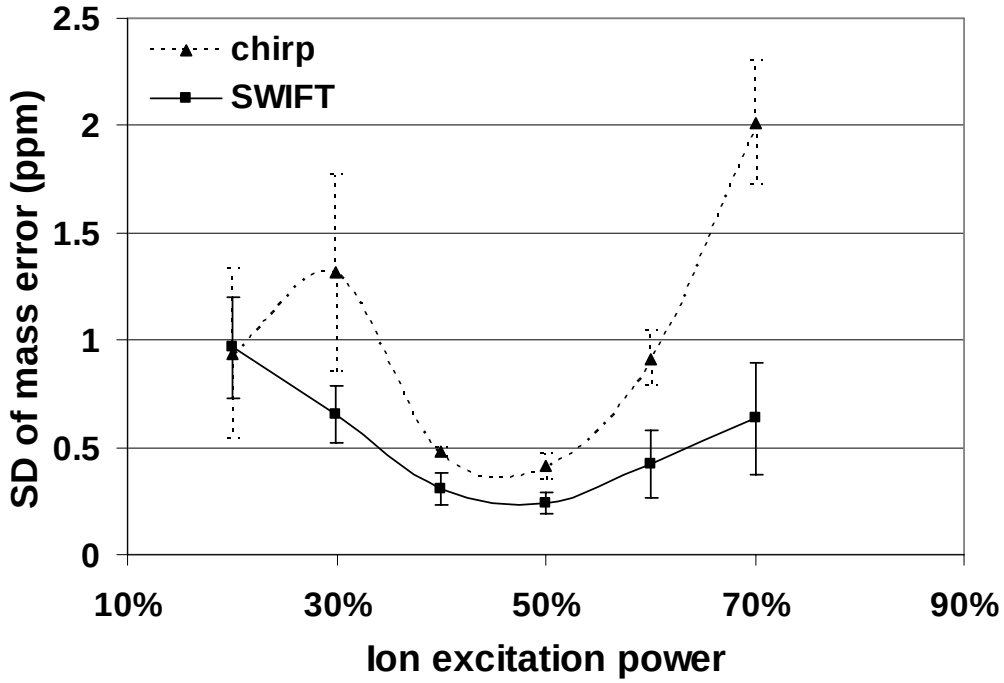


Figure 3.1 Standard deviation of calibration errors obtained from the linear fit to eq 1 versus ion excitation power (the power to eject all ions is normalized to 100%) using chirp or SWIFT excitation. Five measurements were averaged at each condition and the error bar representing 1 SD of the mean.

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 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.

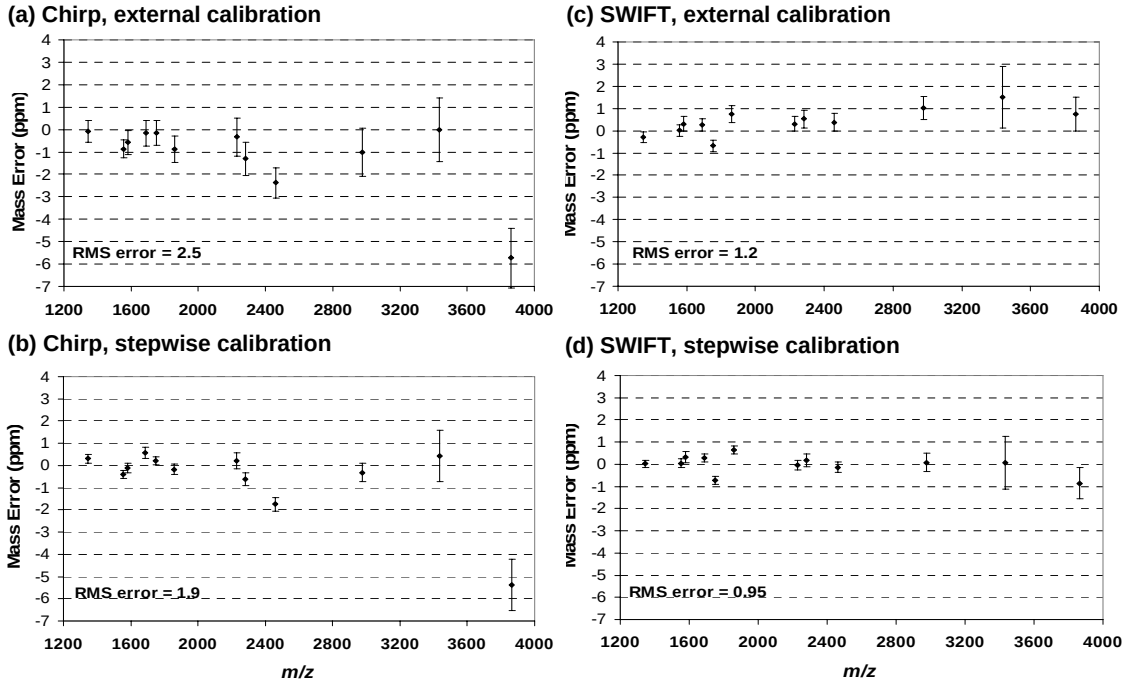


Figure 3.2 Average calibration errors obtained from external calibration (eq 1) and stepwise-external calibration (eq 2), with the error bar being 1 SD for 30 replicates. (a) chirp for external calibration, (b) chirp for stepwise-external calibration, (c) SWIFT for external calibration, (d) SWIFT for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

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 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 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

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. 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. *J. Am. Soc. Mass. Spectrom.* **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. *J. Am. Soc. Mass. Spectrom.* **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. *Chem. Phys. Lett.* **1974**, *25*, 282-283.

8 Comisarow, M. B.; Marshall, A. G. Frequency-Sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **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. *Int. J. Mass Spectrom. 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. *J. Am. Soc. Mass. Spectrom.* **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 FT-ICR. *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. *Chem. Rev.* **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 Spectrom. Rev.* **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. *Int. J. Mass Spectrom. 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. *Anal. Chem.* **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. *J. Am. Soc. Mass. Spectrom.* **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. *Anal. Chem.* **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. *J. Am. Soc. Mass. Spectrom.* **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. *J. Am. Soc. Mass. Spectrom.* **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. *J. Am. Soc. Mass. Spectrom.* **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. *Int. J. Mass Spectrom. 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 Commun. Mass Spectrom.* **1996**, *10*, 1819-1823.

CHAPTER 4

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

Li Jing*, Chunyan Li*, 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

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, 15N 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.

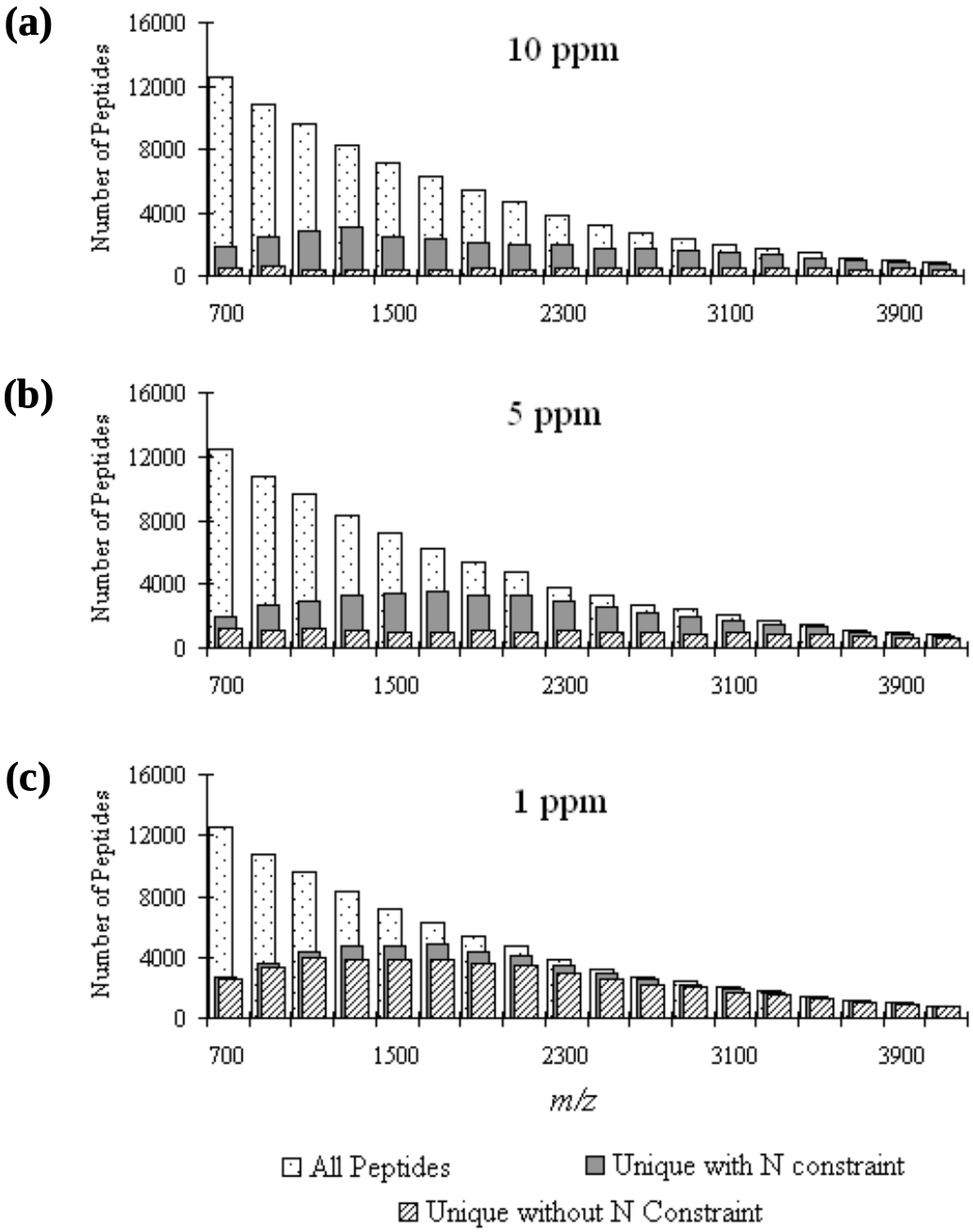


Figure 4.1 Specificity of peptide assignments from *in silico* tryptic digestion of the proteins from *M. maripaludis* over the range 700-4100 Da using (a) 10 ppm, (b) 5 ppm and (c) 1 ppm as search tolerance. The bars filled with dots represent the total number of peptides

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.

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.

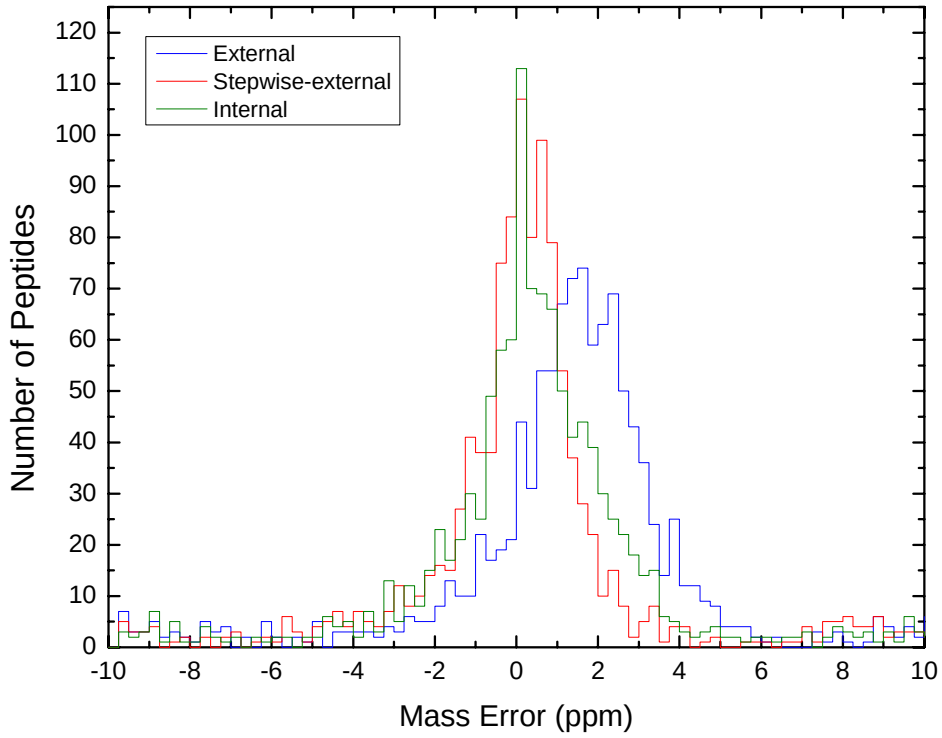


Figure 4.2 Distribution of mass measurement errors of the peptides identified from a *M. maripaludis* proteome sample using 10 ppm search tolerance constrained by nitrogen stoichiometry by applying external calibration (blue), stepwise-external calibration (red) and internal calibration (green), with 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.

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

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.

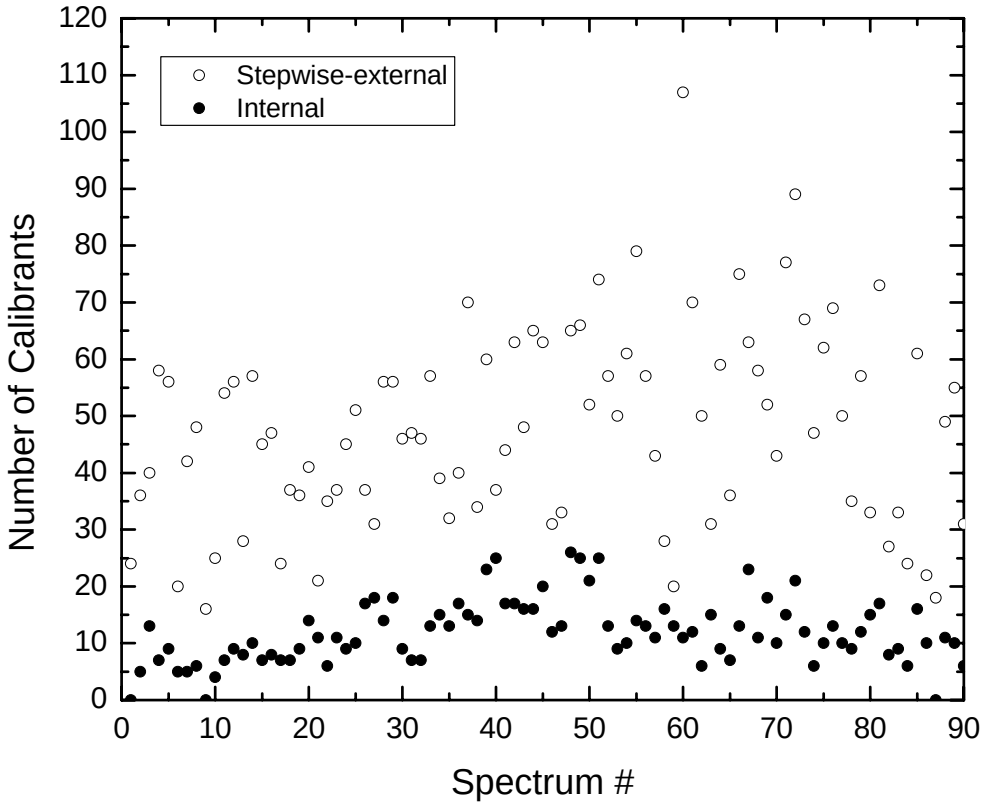


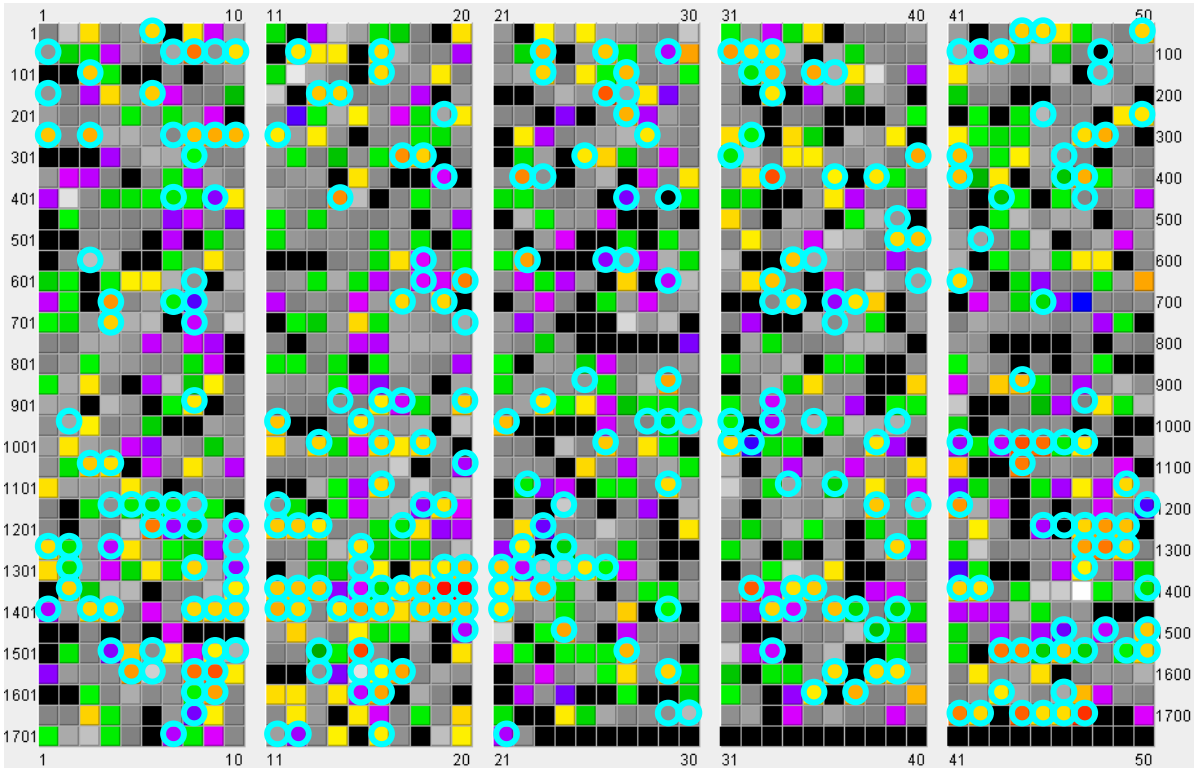
Figure 4.3 Distribution of the number of calibrants found during the liquid chromatographic run for stepwise-external calibration (open circles) and for internal calibration (filled circles).

III. Protein Identification

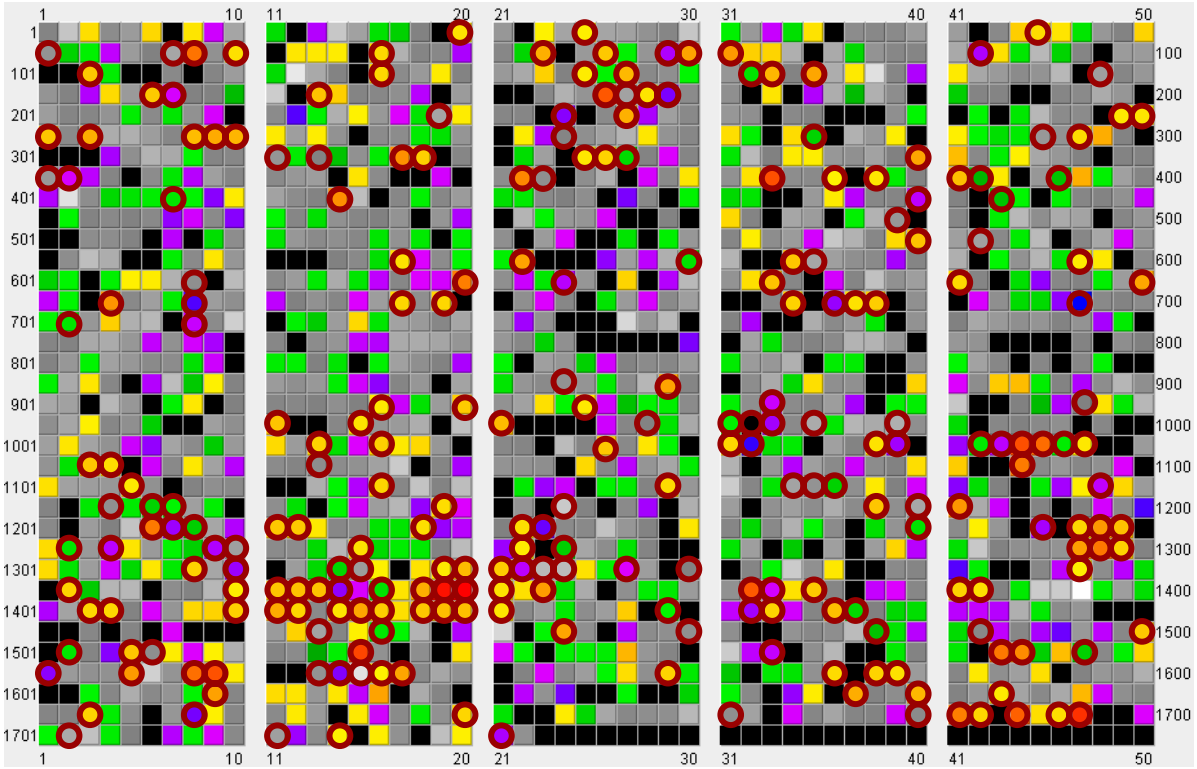
We have previously performed similar error analysis for measurements of peptides from protein standards using stepwise-external calibration, and found that the error distribution is a Gaussian fashion [1]. In contrast, here we find for proteomic data that there is a central Gaussian-like distribution, 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.

(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

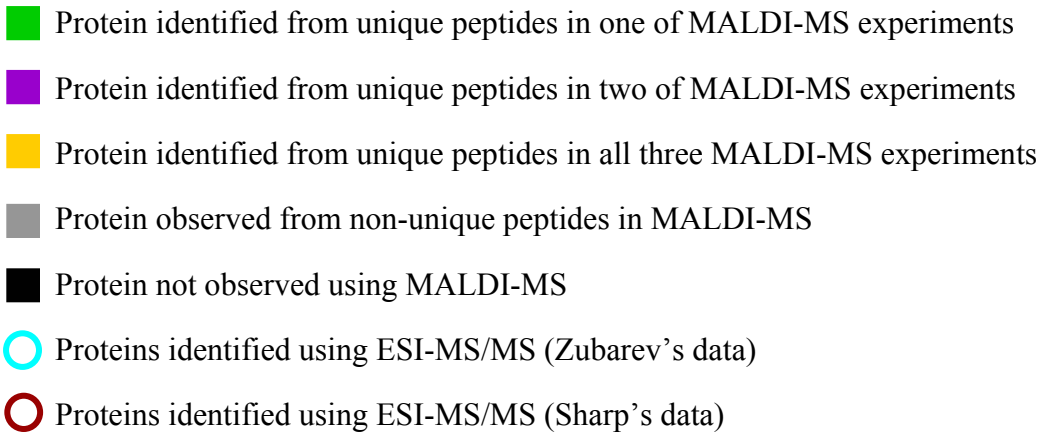


Figure 4.4 Map of protein assignments from triplicate analysis of a *M. maripaludis* proteome sample using MALDI-FTICR/MS with 5 ppm mass tolerance constrained by nitrogen stoichiometry in database searching (squares) and ESI-MS/MS analysis of the same sample using ESI-MS/MS with 10 ppm peptide mass 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.

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 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.

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	S.D.
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 subunit	**1.53**	**10**	**0.29**
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.

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

AN IMPROVED CALIBRATION METHOD FOR THE MALDI-FTICR ANALYSIS OF ^{15}N-METABOLICALLY LABELED PROTEOME DIGESTS USING A MASS DIFFERENCE APPROACH

Li Jing and I. Jonathan Amster. To be submitted to *J. Am. Soc. Mass Spectrom.*

ABSTRACT

High mass measurement accuracy of peptides in enzymatic digests is critical for confident protein identification and characterization in proteomics research. Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) can provide low or sub-ppm mass accuracy and ultrahigh resolving power. While for ESI-FTICR-MS, the mass accuracy is generally 1 ppm or better, with MALDI-FTICR-MS, the mass errors can vary from sub-ppm with internal calibration to over 100 ppm with conventional external calibration. A novel calibration method for ^{15}N-metabolically labeled peptides from a batch digest of a proteome is described which corrects for space charge induced frequency shifts in FTICR spectra without using an internal calibrant. This strategy utilizes the information from the mass difference between the ^{14}N/^{15}N peptide peak pairs to correct for space charge induced mass shifts after data collection. A procedure for performing the mass correction has been written into a computer program and has been successfully applied to HPLC-MALDI-FTICR/MS measurement of ^{15}N-metabolic labeled proteome. We have achieved an average measured mass error of 1.0 ppm and a standard deviation of 3.5 ppm for 900 peptides from 68 MALDI-FTICR mass spectra of the proteolytic digest of a proteome from *Methanococcus maripaludis*.

INTRODUCTION

The capability for accurate mass measurement is critical for confident protein identification and characterization in high-throughput proteomics. Sufficiently high mass measurement accuracy (MMA) can enable the identification of a protein from a single peptide that is unique within the mass error [1-6]. It is well-known that the number of protein or peptide

candidates drops rapidly with increasing MMA [1]. Also, the high MMA reduced the chances of false-positive matching in database searching.

Fourier transform ion cyclotron resonance (FTICR) mass spectrometry is a unique technique which is able to provide simultaneous high resolution, sensitivity, and accurate mass measurements. However, in FTICR mass spectrometry, the number of charges in the analyzer cell affects mass measurement due to space charge effects [7-9]. To take the advantage of high mass measurement accuracy with FTICR, it is necessary to limit systematic mass measurement errors resulting from space charge. Frequency shift from space charge can be a serious issue when the number of ions stored in the cell often varies from one experiment to another. Considerable effort have been made to reduce the effects of space charge, and a number of calibration methods have been proposed [6, 8, 10-15] to correct space charge effects to improve the accuracy of mass measurement. Presently, electrospray ionization (ESI) combined with FTICR/MS can achieve very high MMA by controlling the ion population using automatic gain control (AGC) [16, 17]. However, with matrix assisted laser desorption/ionization (MALDI) FTICR/MS measurements, AGC is not applicable due to the large shot-to-shot variation in ion intensity [18, 19].

It is known that the mass accuracy of a FTICR measurement depends critically on the method of mass calibration. External calibration has been used to account for frequency shift and to thereby improve the mass accuracy [20]. External calibration works best when the total ion population used in the calibration procedure is carefully matched to that of the analyte ions [9]. However, in proteomics analysis, liquid chromatography (LC) separation is often used before mass spectrometry to reduce the sample complexity, and the analyte ion population may vary widely in the analyzer cell during the course of a chromatographic run. For each spectrum, all

ions in the analyzer cell experience the same space charge induced frequency shift, resulting in a systematic mass error. Conventional internal calibration is based on mixing internal calibrants of known molecular weight with the analyte and then using the known masses to calibrate the mass measurements of unknown ones [16, 21-23]. Both calibrants and analyte are measured under identical condition, so the space charge effect is canceled during calibration. The mass accuracy obtained by internal calibration can be an order of magnitude greater than external calibration [21-25]. However, such internal calibrants may overlap with species of interest during MS scans and the implementation of standard internal calibrant can sometimes be tedious and time-consuming.

There are few calibration approaches that have been developed which avoid the use of internal calibrants. A method called Deconvolution of Coulombic Affect Linearity (DeCAL) developed by Bruce et al. [26] corrects the space charge effects in FTICR measurements. DeCAL is based on observation of multiple charged states of the same molecular species generated by ESI MS to improve MMA. This method requires at least two charge states of the same molecular species, which are generally not present in MALDI-FTICR mass spectra.

We recently described a novel stepwise-external calibration, in which a mass spectrum is first acquired at low trapping potential, with sub ppm MMA by external calibration [18]. 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 calibrants for the spectrum acquired at higher trapping potential. We have demonstrated that sub-ppm MMA can be achieved when SWIFT excitation is combined with stepwise-external calibration [19]. However, it should be noted that this calibration method requires two spectra for every sample, which doubles the effort of data collection and interpretation.

This paper presents a simplified, calibrant-free method for automatic calibration suitable for ^{15}N-metabolic labeled proteome digests by HPLC-MALDI-FTICR/MS. In the mass spectra, peptides appear as light and heavy pairs of peaks, and the nitrogen stoichiometry can be derived from the mass spacing between pairs of $^{14}N/^{15}N$ peptides [27-29]. Thus, in the mass spectra, the spacing between a heavy/light peak pair equals the number of nitrogen atoms present in the peptide times the mass difference between ^{14}N and ^{15}N, namely 0.9970 u. Space charge-induced frequency shift can be compensated by linearly shifting the measured frequency to exactly match the mass difference of light/heavy peak pair to the integer value of nitrogen atoms present in the peptide times 0.9970. The calibration algorithm optimizes frequency shift to minimize the mass error for a large number of light/heavy pairs in a mass spectrum, and then uses this frequency offset as the basis to recalculate all masses in the spectrum. This is done without knowledge of the exact molecular weight or ion intensity of the species. This post-acquisition calibration method is particularly attractive for MALDI experiments where large shot-to-shot variations in ion population typically occur.

EXPERIMENTAL

Sample Preparation

Differential Labeling ($^{14}N/^{15}N$), and sample preparation of proteome from *Methanococcus maripaludis* were performed as described previously [29], Briefly, equal amounts of labeled and unlabeled samples were mixed to form the cell mixtures. The cell mixtures were centrifuged at 10,000 x g for 30 min at 4 ^{o}C and followed by lysis with a French pressure cell. DNA was digested by DNase and incubated for 15 min at 37 ^{o}C. Cell debris was pelleted by centrifugation at 8000 x g for 30 min at 4 ^{o}C. Prior to digestion, proteins were

denatured by heating at 90 °C for 10 min. Disulfide bonds were reduced with tris (2-carboxythyl) phosphine (Pierce biotechnology, Rockford, IL). The mixture was digested overnight at 37 °C using trypsin (Promega, Madison, WI) at a 1:50 protease/protein ratio (by mass).

HPLC-MALDI-FTICR/MS

Separation of peptide mixture was carried out by nanoflow liquid chromatography using an UltiMate Plus HPLC system (Dionex LC packings, Sunnyvale, CA). Peptides were eluted with increasing acetonitrile (5-80% in 90 min) at approximate flow rate of 300 nL/min. The eluent was detected by UV absorption at 214 nm and was collected onto a stainless steel MALDI target at a 60-s interval using a Probot Micro Fraction collector (Dionex LC packings, Sunnyvale, CA). Samples were allowed to dry before adding 500 nL of 1 M 2,5-Dihydroxybenzoic acid (DHB) (Lancaster, Pelham, NH) as the MALDI matrix.

A 7-T FTICR mass spectrometer equipped with an intermediate pressure Scout 100 MALDI source (Bruker Daltonics Inc, Billerica, MA) was used to acquire all spectra in positive-ion mode. Conditions for operation of the FTICR/MS were similar to those reported previously [27] with minor modifications. The cell trapping potential was set as 1.10 V. The ions generated from 6 MALDI laser shots per scan and 16 scans were summed for each spectrum.

Calibration Methods

External mass calibration was established through Bruker XMASS 7.0.8 software (Bruker Daltonics, Billerica, MA) using a peptide mixture generated by tryptic digestion of bovine serum albumin (BSA) (Sigma, St. Louis, MO). The calibration function that was used to

convert ion cyclotron frequency to *m/z* values is presented in eq 1, where *A* and *B* are calibration constants, and *f* is the measured frequency.

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

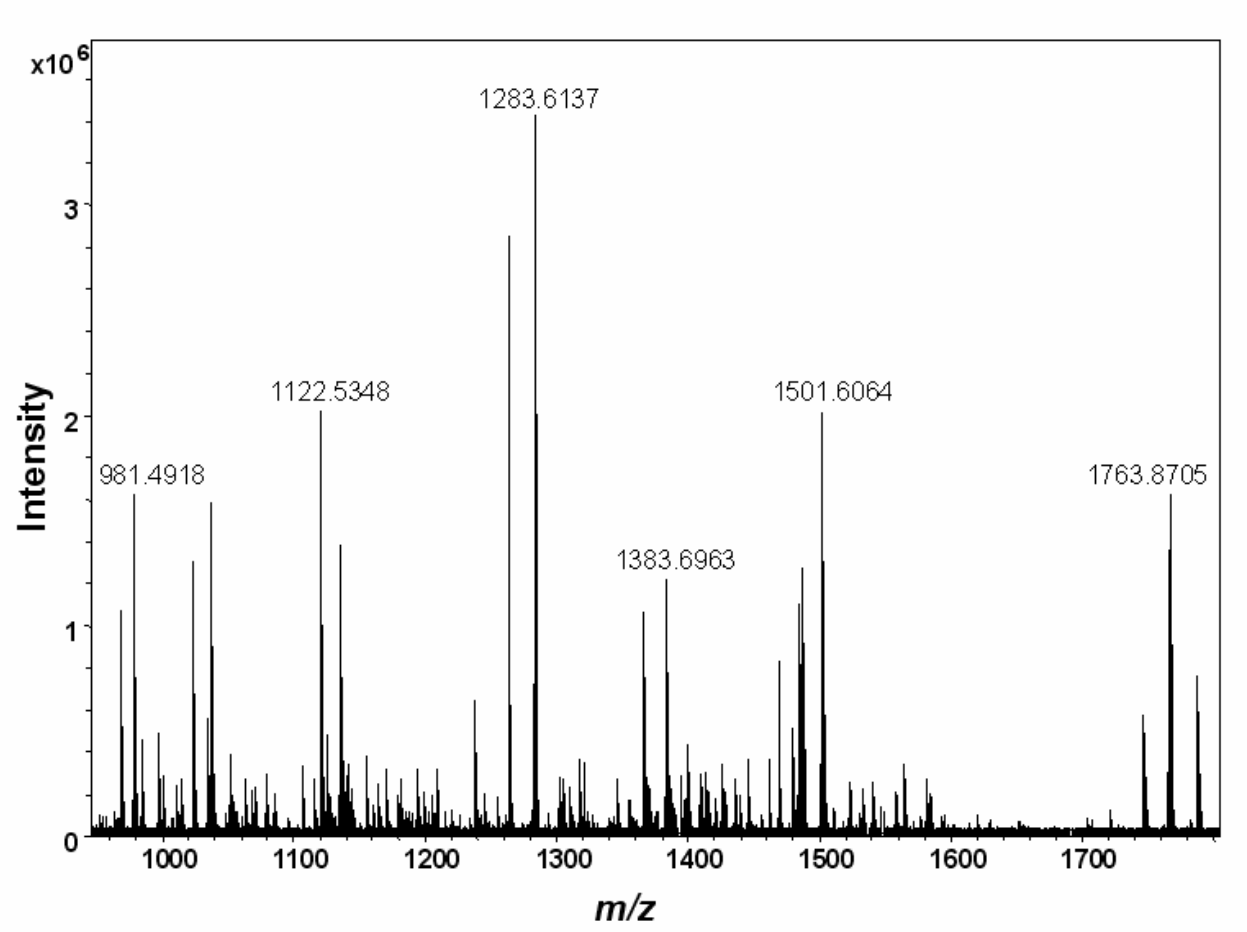


Figure 5.1 MALDI-FTICR mass spectrum of a HPLC fraction from the analysis of a tryptic digest of a proteome with ^{15}N-metabolical labeling.

The proposed calibration method examines the mass differences between ^{14}N and ^{15}N peak pairs in a single mass spectrum, and provides a frequency shift to all peaks in the mass spectrum to reduce the errors induced by space charge effects. In a mass spectrum, several light and heavy peak pairs are typically observed, as seen in Figure 5.1. The average number of light/heavy peak pairs is around 30 in a typical MALDI mass spectrum. Recently, we developed an algorithm to identify the $^{14}N/^{15}N$ peptide pairs and calculate peptide relative abundance ratios automatically in mass spectrum from ^{15}N metabolic labeling experiments [29]. The mass difference between a pair of peaks (^{14}N and ^{15}N peptides) is an integer number times 0.997 u. Space charge-induced frequency shifts can be compensated by shifting the observed frequency to match the mass difference of light/heavy pair to the exact number of nitrogen atoms times 0.997 u. A series of frequency shifts are applied to a pair of light/heavy peaks and the optimum frequency shift is produced when the minimum mismatch error is obtained between corresponding monoisotopic peaks from light/heavy pair. Then frequency shift for the spectrum is determined by calculating a weighted average of the frequency shifts measured for each light/heavy pair. The resultant frequency shift (Δf) is then used as the basis to recalculate all m/z species in the spectrum. The simple relationship between the *m/z* and frequency is defined by eq 2. Our algorithm uses the same calibration constants (A and B) as external calibration (eq1).

$$\frac{m}{z} = \frac{A}{f + \Delta f + B} \qquad (2)$$

For comparison purpose, stepwise-external calibration and post-processing internal calibration were performed. For stepwise-external calibration [18], an additional mass spectrum was acquired for each spot at low trapping potential (0.60 V). The mass values from the spectrum recorded at low trapping potential were used as calibrants to calibrate the mass spectrum recorded for the same sample at high potential (1.10 V) via eq 3. The calibration

constants, *A*, *B* and *C*, were determined by multi-linear fitting the measured frequencies and intensities of the peaks acquired at high potential to their corresponding mass values measured at low trapping potential. Multi-linear regression to obtain the constants for the calibration equations was preformed using software developed in our laboratory [19]

$$\left(\frac{m}{z}\right)_i = \frac{A}{f_i + B + C \cdot I_i} \qquad (3)$$

Post-processing internal calibration was also performed, using the peptides uniquely identified from MALDI-FTICR/MS analysis of the proteome sample as internal calibrants [6, 30]. The theoretical mass values of the peptides identified within10 ppm were used to calibrate the measured masses in the original externally calibrated spectrum using eq 3. Multi-linear regression was used to obtain the calibration constants, *A*, *B* and *C*, by fitting the measured frequencies and intensities of the identified peptides to their theoretical masses.

RESULTS AND DISCUSSION

Algorithm Description

The algorithm, N15Cal, has been incorporated into a Java program which acts as a correction process for space charge effects. N15Cal utilizes information that is derived from the mass difference for light/heavy peak pairs of the same peptides that are generally present in MALDI spectra of ^{15}N metabolic labeling data. As shown in Figure 5.2a, the exact mass difference between a light/heavy peak pair ($\Delta M_{calculated}$) should be the integer number of nitrogen atoms present in the peptide times the mass difference between ^{14}N and ^{15}N (0.9970 u) ($\Delta M_{theoretical}$). However, improper calibration due to space charge arise if the total ion population in the ICR cell during the experiment is different than the total ion population in the ICR cell during calibration. A difference between $\Delta M_{calculated}$ and $\Delta M_{theoretical}$ is observed. Here our

assumption is that the frequency shift is a constant value across the entire spectrum. By correcting the frequency shift, the recalculated mass difference of light/heavy peak should match the exact mass of nitrogen number in this peptide. It should also be applicable in cases where pairs of peaks having predictable mass differences.

Figure 5.2b illustrates the principle of N15Cal. The Y axis "error" is defined as the difference between $\Delta M_{calculated}$ and $\Delta M_{theoretical}$. By applying a series frequency shift, all the *m/z* values are recalculated by calibration eq 2. As can be seen, the error between the measured mass and an integer number of $^{14}N/^{15}N$ mass difference (0.9970 u) varies linearly with the applied frequency shift. A minimum error between $\Delta M_{calculated}$ and $\Delta M_{theoretical}$ is observed when the shift in frequency is equal to the frequency shift due to the space charge.

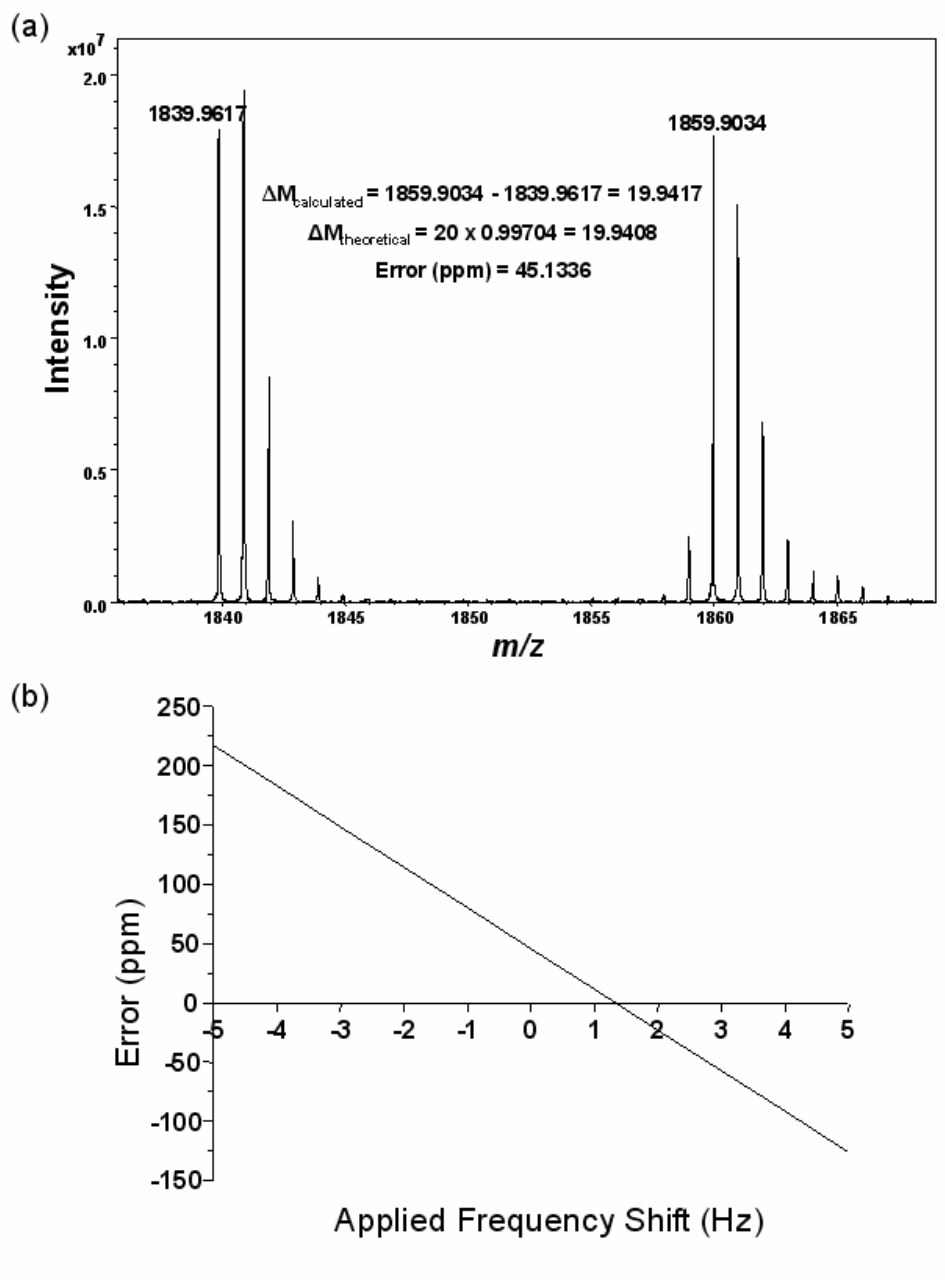


Figure 5.2 Schematic representation of N15Cal process. (a) Mass difference of light/heavy peak pair for the same peptide ($\Delta M_{calculated}$) should be same as the mass of nitrogen number in the peptide ($\Delta M_{theoretical}$). However, a mismatch of theses two value occurs because of the constant frequency offset. (b) Relationship between the mass error (the difference between $\Delta M_{calculated}$ and $\Delta M_{theoretical}$) and applied frequency shift. The calculated frequency shift to obtain the minimum mass error is 1.25 Hz.

The N15Cal algorithm calculates a corrective frequency shift for each light/heavy peak pair in a mass spectrum. A Q test is implemented in the algorithm to discard outliers with a 95% confidence level. After the Q test, an optimal frequency shift is determined by calculating the average of the frequency shifts for all peak pairs. The optimal frequency shift is then applied to all peaks in the spectrum and all masses are recalculated. N15Cal acts as a correction process for space charge effects by determining the optimal frequency shift for all light/heavy peak pairs in a single mass spectrum.

Evaluation of the Algorithm Performance for Proteome Analysis

To investigate the performance of the N15Cal, a ^{15}N-metabolic labeled proteome from *M. maripaludis* is studied. Unlabeled and labeled protein extracts were mixed in approximately equal amounts, and batch digestion yields a mixture of peptides pairs that differ in mass by number of nitrogen atoms present in the peptide. In this experiment, 68 fractions from the HPLC separation were collected directly onto a MALDI target, and then analyzed by MALDI-FTICR. Protein identifications were made by searching against the database using a mass tolerance of 10 ppm.

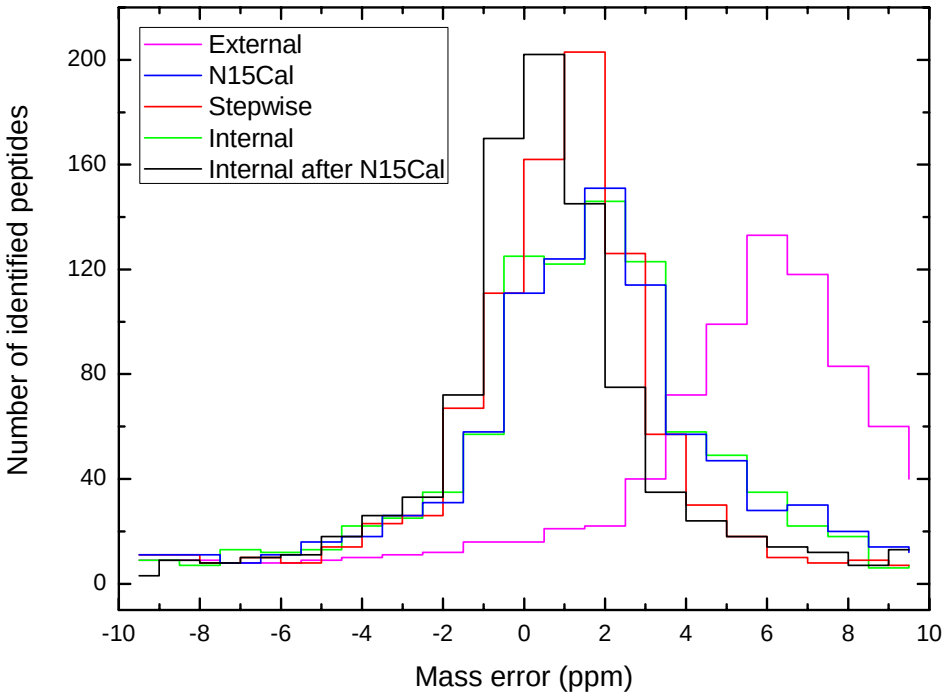


Figure 5.2 Mass measurement errors of the peptides identified from a *M. maripaludis* proteome sample using 10 ppm search tolerance by applying external calibration (pink), post-processing internal calibration (green), stepwise-external calibration (red), N15Cal calibration (blue) and internal calibration after application of N15Cal (black).

N15Cal was compared with other calibration methods by studying the mass error distribution of peptides identified within 10 ppm. Initially, the ^{15}N-labeled proteome sample was analyzed by HPLC-MALDI-FTICR/MS using external calibration with calibration eq 1. This externally calibrated dataset is then calibrated by three separate calibration methods. N15Cal was applied to this dataset to determine the frequency shift for each spectrum, and recalibrate all the

masses. Stepwise-external calibration mimics internal calibration, but the calibrants are the mass values obtained from a separately acquired spectrum using external calibration at low trapping potential [18, 19]. The post-processing internal calibration was performed by using the theoretical masses of identified peptides as calibrant points [6, 30].

To visualize the performance of N15Cal and compare N15Cal with other calibration methods, a plot of the mass measurement error distribution (shown in ppm) of the peptides identified by matching within 10 ppm is shown in Figure 5.3. Table 5.1 shows the number of identified peptides, average number of calibrationa points per spectrum, average mass measurement error (AVG) and standard deviation (SD) of mass measurement error for the calibration methods used in Figure 5.3. The center of the measured *m/z* error distribution of external calibration (pink line in Figure 5.3) is not equal to zero and it shows a broader error distribution than other calibration methods. The obvious systematic error is due to the analyte ion population being larger than that used for the prior calibration. Without sophisticated "automatic gain control", mismatch of ion population is very likely in MALDI-FTICR experiments. Fortunately, this systematic error induced by space charge is a constant offset which uniformly shifts all frequencies in the spectrum. N15Cal was then applied to the acquired data to correct space charge induced frequency shift without knowledge of the ion population. The application of the N15Cal algorithm significantly reduces the average mass measurement errors from 4.3 ± 4.1 ppm (AVG ± SD) to 1.0 ± 3.5 ppm (Table 5.1), which shows that N15Cal has virtually corrected the systematic error.

Table 5.1 Error analysis for the identified peptides using different calibration methods.

	External	N15Cal	Stepwise	Internal	Internal after N15Cal
Number of peptides	796	898	919	904	905
Average number of calibrants/spectrum	6	14	47	12	12
AVG (ppm)	4.3	1.0	0.7	0.9	0.5
SD (ppm)	4.1	3.5	3.1	3.4	3.0

The stepwise-external calibration yields an MMA of 0.7 ± 3.1 ppm (Table 5.1), which shows the best MMA compared to internal calibration and N15Cal. We attribute this to the fact that there are more calibrant points were used in stepwise-external calibration for each mass spectrum, as any peaks in the low trapping potential mass spectrum could be used as reference points for stepwise-external calibration. As shown in Table 5.1, the average calibrant points used in stepwise-external calibration is ~ 4-fold more than the points used in post-processing internal calibration and N15Cal. The mass accuracy that is obtained by using stepwise-external calibration depends on the spectrum acquired at low trapping potential, therefore the data acquisition and interpretation time is doubled. In addition, the mass accuracy that is obtained by

using stepwise-external calibration is based on the mass accuracy of the spectra acquired at low potential. In this experiment, the mass error of stepwise-external calibration has a average offset of 0.7 ppm, because small systematic error was observed in the calibrant points obtained from externally calibrated mass spectra acquired at low trapping potential.

The mass measurement error produced by post-processing internal calibration is 0.9 ± 3.1 ppm (Table5.1), which is slightly worse than stepwise-external calibration. Theoretically, the internally calibrated data could accurately remove the systematic bias. In this case, the calibrants are theoretical masses of peptides in the externally calibrated mass spectra that are uniquely identified within a 10 ppm mass tolerance, so the number of calibrant points is limited, and then the calibrant points may not span the mass range of analyte ions in all cases. In addition, the average mass error of external calibration is relatively large (4.3 ppm), so some peaks used as calibrant points may have been misassigned. To reduce the number of possible misassigned peaks used as internal calibrant points, the internal calibration method was also applied to the data after processing by N15Cal. As can be seen from Figure 5.3 (black line) and Table 5.1, the distribution of the identified peptides shows a central mass error of 0.5 ppm and SD is 3.0 ppm. The better MMA in this case results from using calibrant points identified by N15Cal, which provides more identified peptides and less misassigned peaks for use as internal calibrants.

In these studies involving ^{15}N metabolic labeled proteome, we carried out and compared several calibration methods. For mass error analysis of unique peptides, the SD error is highest for external calibration, having a value of 4.1 ppm, whereas the SD values for N15Cal, stepwise-external, post-processing internal and internal after N15Cal calibration methods are 3.5 ppm, 3.1 ppm, 3.4 ppm and 3.0 ppm, respectively. The resolving power (M/ΔM at fwhm) in all these spectra ranged from ~120 000 for low-mass ions at *m/z* 700 to ~ 22 000 at *m/z* 3000. The broad

error distribution may be a limitation of the resolving power, a result of the wide mass range examined in this proteomic study. As the *m/z* of a peak is determined by calculating the centroid of the peak, random errors occur because not all mass peaks are homogeneous. Lower resolution reduces the accuracy of the assignment of peaks. A standard means of improving mass accuracy in MALDI-FTICR analysis with a broad dynamic range is to (1) increase the magnetic field for FTICR, (2) increasing the memory size of the transient digitizer and (3) increasing the duration of the transient signal.

These results suggest substantial advantages of N15Cal for ^{15}N metabolic labeling proteomics analysis by MALDI-FTICR/MS. The same calibration coefficients from external calibration are used by this procedure, which corrects the mass measurement by applying correction to the measured cyclotron frequency to account for space charge induced frequency shifts.. The shift of cyclotron frequency to higher value leads to a lowering of the space charge effect. N15Cal shows comparable MMA to stepwise-external calibration and post-processing internal calibration, but avoids the substantial additional work for data acquisition and selection of calibrant points characteristic of these other methods. Overall, the application of N15Cal resulted in low ppm mass measurement errors, which increases the confidence for protein identification in ^{15}N-labeling proteomic measurements.

CONCLUSIONS

In applying of FTICR mass spectrometry to high throughput proteomics, it is important to achieve a high mass accuracy for the determination/characterization of complex mixtures. N15Cal is a simplified, calibrant-free method for automatic internal calibration suitable for ^{15}N-metabolic labeled proteome digests by HPLC-MALDI-FTICR/MS. N15Cal corrects the effects

of space charge and enables higher mass accuracy with the information of $^{14}N/^{15}N$ peak pair separation that is inherent in the spectra. This has been successfully applied to the HPLC-MALDI-FTICR/MS measurement of ^{15}N-metabolic labeled proteome from *Methanococcus maripaludis*. With this method, low ppm mass measurement errors are achieved without addition of internal calibrant or instrument modifications. This capability allows proteins to be identified unambiguously with a tighter mass tolerance, which yields greater confidence in search results. The approach may also be employed to FTICR mass spectrum containing pairs of peaks having predictable mass differences. In addition, the method should also be applicable to other mass spectrometry methods, such as orbitrap, where space charge frequency shift might also affect mass measurement accuracy.

REFERENCES

1 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.

2 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.

3 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.

4 Strittmatter, E. F.; Ferguson, P. L.; Tang, K.; 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.

5 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. *J. Proteome Res.* **2004**, *3*, 61-67.

6 Palmblad, M.; Bindschedler, 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.

7 Ledford, E. B.; Rempel, D. L.; Gross, M. L. Space Charge Effects in Fourier Transform Mass Spectrometry. I. Electrons. *Int. J. Mass Spectrom. Ion Processes* **1984**, *55*, 143-154.

8 Ledford, E. B.; Rempel, D. L.; Gross, M. L. Space Charge Effects in Fourier Transform Mass Spectrometry. II. Mass Calibration. *Anal. Chem.* **1984**, *56*, 2744-2748.

9 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.

10 Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villen, J.; Gygi, S. P. Optimization and Use of Peptide Mass Measurement Accuracy in Shotgun Proteomics. *Mol. Cell. Proteomics* **2006**, *5*, 1326-1337.

11 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.

12 Yanofsky, C. M.; Bell, A. W.; Lesimple, S.; Morales, F.; Lam, T. T.; Blakney, G. T.; Marshall, A. G.; Carrillo, B.; Lekpor, K.; Boismenu, D.; Kearney, R. E. Multicomponent Internal Recalibration of an LC-FTICR-MS Analysis Employing a Partially Characterized Complex Peptide Mixture: Systematic and Random Errors. *Anal. Chem.* **2005**, *77*, 7246-7254.

13 Tolmachev, A. V.; Monroe, M. E.; Jaitly, N.; Petyuk, V. A.; Adkins, J. N.; Smith, R. D. Mass Measurement Accuracy in Analyses of Highly Complex Mixtures Based Upon Multidimensional Recalibration. *Anal. Chem.* **2006**, *78*, 8374-8385.

14 Williams, D. K.; Chadwick, M. A.; Williams, T. I.; Muddiman, D. C. Calibration Laws Based on Multiple Linear Regression Applied to Matrix-Assisted Laser Desorption/Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Journal of Mass Spectrometry* **2008**, *43*, 1659-1663.

15 Zhang, J.; Ma, J.; Dou, L.; Wu, S.; Qian, X.; Xie, H.; Zhu, Y.; He, F. Mass Measurement Errors of Fourier-Transform Mass Spectrometry (FTMS): Distribution, Recalibration, and Application. *J. Proteome Res.* **2009**, *8*, 849-859.

16 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.

17 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.

18 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.

19 Jing, L.; Li, C.; 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.

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

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. *J. Am. Soc. Mass Spectrom.* **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. *Anal. Chem.* **2000**, *72*, 5886-5891.

24 Hofstadler, S. A.; Griffey, R. H.; Pasa-Tolic, L.; Smith, R. D. The Use of a Stable Internal Mass Standard for Accurate Mass Measurements of Oligonucleotide Fragment Ions Using Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Infrared Multiphoton Dissociation. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1400-1404.

25 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. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 99-106.

26 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.

27 Wong, R. L.; Amster, I. J. Combining Low and High Mass Ion Accumulation for Enhancing Shotgun Proteome Analysis by Accurate Mass Measurement. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 205-212.

28 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.

29 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*.

30 Mørtz, E.; Vorm, O.; Mann, M.; Roepstorff, P. Identification of Proteins in Polyacrylamide Gels by Mass Spectrometric Peptide Mapping Combined with Database Search. *Biol. Mass Spectrom.* **1994**, *23*, 249-261.

CHAPTER 6

CONCLUSIONS

The overall purpose of this work was to improve and develop methods for proteomic analysis by accurate mass measurements using FTICR-MS.

Chapter 2: Rapid and Automated Processing of MALDI-FTICR/MS Data for ^{15}N-Metabolic Labeling In a Shotgun Proteomics Analysis. We developed a simple but robust algorithm (N15Tool) that aids the process of ^{15}N metabolic labeling data resulting from LC-MS. N15Tool provides an automated method for identifying ^{14}N and ^{15}N peak pairs and calculating the ratios between the light and heavy peptides without prior knowledge of amino acid composition of the peptides. This automated program significantly reduces the burden of manually interpreting large scale LC-MS data sets from ^{15}N labeling experiments. N15Tool is able to be used with any ^{15}N metabolic labeling data from any instrument that can provide accurate mass measurement.

Chapter 3: Improved Mass Accuracy for Higher Mass Peptides by Using SWIFT Excitation for MALDI-FTICR Mass Spectrometry. We presented a comparison of chirp and SWIFT excitation for accurate mass measurement in MALDI-FTICR/MS using stepwise-external calibration. The utilization of a SWIFT excitation waveform reduced the mass errors significantly compared to chirp excitation, particularly for ions of higher mass-to-charge. The sub-ppm MMA can be achieved when SWIFT excitation is combined with stepwise-external calibration. 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.

Chapter 4: Shotgun Proteomic Analysis Using Accurate Mass Measurement and Nitrogen Stoichiometry — a HPLC-MALDI-FTICR/MS Approach. We reported the results of protein identification and quantitation using an accurate mass HPLC-MALDI-FTICR/MS approach combined with nitrogen stoichiometry. The protein identification was 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. We have also shown that protein identification can be improved by using stepwise-external calibration as it significantly improves the mass measurement accuracy. In the meanwhile, quantitative measurements of the changes in protein expression were made for the proteome sample analyzed.

Chapter 5: An Improved Calibration Method for the MALDI-FTICR Analysis of ^{15}N-Metabolically Labeled Proteome Digests Using a Mass Difference Approach. A simplified, calibrant-free calibration method, N15Cal, has been developed to correct for space charge-induced frequency shifts in FTICR spectra of ^{15}N-metabolic labeled proteome. N15Cal corrects the effects of space charge and to enable higher mass accuracy with the information of ^{14}N/^{15}N peak pairs that is inherent in the spectra. With this calibration method, low ppm mass measurement errors are achieved without addition of internal calibrant or instrumental modifications. This capability allows proteins to be identified unambiguously with tighter mass tolerance set, which yields greater confidence in search results.