University of Alberta

Development and Applications of Mass Spectrometric Methods for Proteome Analysis and Protein Sequence Characterization

by

Lu Chen

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

Fall 2013
Edmonton, Alberta

Abstract

Mass spectrometry (MS) is widely used in proteomic work for protein identification and characterization. A key to the success of MS for protein analysis is related to sample preparation. Current sample preparation methods do not meet all the needs of protein or proteome analysis. My thesis work is devoted to the development and application of new or improved protein sample preparation methods for MS analysis. These methods are focused on three aspects: microbore LC-UV for quantification of peptides generated from the proteins extracted from a small number of cells, 2-MEGA (N-terminal dimethylation after lysine guanidination) isotope labeling for proteome quantification, and microwave-assisted acid hydrolysis (MAAH) for protein sequence analysis. It is shown that microbore LC-UV is a robust technique for quantifying peptides down to about 40 ng, as well as for removing salts for LC-MS/MS analysis of peptides and this method can be used to gauge the sample integrity during the proteomic sample preparation in dealing with small number of cells. An automated 2-MEGA labeling method utilizing a commercial liquid handler was developed to minimize variability from sample handling during the labeling reaction for high throughput applications. To find biomarkers of breast cancer, the 2-MEGA labeling method was applied to the breast cancer tissues for relative proteomic comparison with the normal breast tissues. 119 proteins were differentially expressed in all three tumor samples and some of these proteins can potentially be verified and validated as biomarkers. To provide a detailed characterization of a protein of

interest including amino acid substitution and modifications, MAAH MS was further developed to improve its performance and applicability. Hydrolysis performed in an optimized procedure using a commercial microwave oven was found to be comparable with that in a household microwave oven. A method of characterizing terminal peptides of a protein based on HCl MAAH of proteins, LC fractionation of the hydrolysates and LC-ESI MS/MS analysis of the low molecular weight peptides, was developed. In addition, proteins separated by gel can be hydrolyzed by MAAH to analyze their protein sequences. Both electroelution of proteins from a gel and in-gel MAAH were studied for analyzing gel-separated proteins. In-gel MAAH provided higher sensitivity than the electroelution method.

Acknowledgements

First of all, I would like to sincerely thank my supervisor Professor Liang Li, for his guidance, encouragement, valuable advice and inspiring discussions in all these years of my PhD study. I have learnt how to find problems, how to solve problems, how to study and how to collaborate with others in his group and other groups. "Think, think and think" is another thing I learnt from him. I am certain that all these invaluable experiences and advices will benefit my future life and career.

I would like to thank my PhD committee members, Professors Charles A. Lucy, James Harynuk, Derrick Clive, and Professor David Schriemer of the University of Calgary for their assistance in my PhD study and thesis preparation.

I want to show my deep appreciation to all the collaborators I worked with. Thank Dr. Sambasivarao Damaraju and his technologist Jennifer Dufour for the breast cancer samples (Chapter 4) provided to me. Thank Gareth Lambkin for his assistance in cell culture and mCherry protein purification and Dr. Xiaofen Xu for her assistance in DNA Sequencing of mCherry (Chapter 6). Thank Dr. Andy Lo and Dr. Nan Wang for training me when I started the lab work and their guidance in LC-UV work (Chapter 3) and 2-MEGA work (Chapter 2). Thank Dr. Yanan Tang for collaborating with me to finish the thesis work in Chapter 2 and 3. Thank Ronghong Li and Professor Guohui Lin for the software program

developed for data processing. Thank Dr. Nan Zhang for her training and advice on MALDI-TOF.

I would like to thank all the members of Professor Li`s research group, especially D r. Xiaoxia Ye, Dr. Mingguo Xu, Dr. Andrea De Souza, Dr. Azeret Zuniga and Dr. Avalyn Lewis for their training, helpful discussions and advices. Thank Dan Xu for her assistance in the data processing. Thank other members, Dr. Peng Wang, Dr. Fang Wu, Margot Dawe, Difei Sun, Feifei Fu, Dr. Kevin Guo, Yiman Wu, Chiao-Li Tseng, Jun Peng, Ruokun Zhou, Jiamin Zheng, Tao Huan, Jared Curle, Zhendong Li and Tran Tran for the fruitful discussions and supports.

I would like to thank Dr. Randy Whittal, Dr. Angie Morales-Izquierdo, Jing Zheng and Bela Reiz from the Mass Spectrometry Facility as well as Wayne Moffat and Miranda Skjel from the Analytical and Instrumentation Laboratory in the Chemistry Department for their help in sample analysis and discussion about sample preparations.

I would like to also thank all the staff in glass shop, electronic shop, machine shop, IT group, general office of Chemistry Department for their assistances.

I would like to thank the Department of Chemistry and the University of Alberta to give me the opportunity to study in this famous university and the financial support provided.

I would like to specially thank my roommate, Zedu Huang and Li Xia, for their helps and assistance in the past years. It is really nice to be part of a small “family” with them.

At last, I would like to thank my dear parents, Houjian Chen and Zhenghua Li, for their love, encouragement, understanding and support in my life. As the only child in the family, I cannot take care of them while I am studying abroad in these years. I know they always miss me so much, but I can only come back home once a year. I feel so sorry for them. I would like to appreciate my husband, Menghui Qi for his love and waiting for me for such a long time in China. Not many people can wait for you for 6 years at the original place. I believe we will not be separated after I graduate from the University of Alberta and come back to China and the future will be filled with happiness.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2D	two dimensional
2-MEGA	dimethylation after guanidination
CHCA	α -cyanohydroxycinnamic acid
DHB	2, 5-dihydroxybenzoic acid
ESI	electrospray ionization
ETD	electron transfer dissociation
ECD	electron capture dissociation
GRAVY	grand average of hydropathicity
HCCA	α-cyanohydroxycinnamic acid
DC	direct current
AC	alternating current
RF	radio frequency
IEF	isoelectric focussing
LC	liquid chromatography
m/z	mass to charge ratio
MALDI	matrix-assisted laser desorption/ionization
MS	mass spectrometry
MS/MS	tandem mass spectrometry
TOF	time-of-flight
QTOF	quadrupole time-of-flight

RPLC	reversed-phase liquid chromatography
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TFA	trifluoroacetic acid
FA	formic acid
ACN	acetonitrile
BSA	bovine serum albumin
CNBr	cyanogen bromide
DTT	dithiothreitol
IAA	iodoacetamide
MAAH	microwave-assisted acid hydrolysis
NP-40	Nonidet-P40
PTM	posttranslational modification
PBS	phosphate-buffered saline
E. coli	*Escherchia coli*
SA	sinapinic acid
Tris	Tris(hydroxymethyl)aminomethane
m	milli-(10^{-3})
μ	micro-(10^{-6})
n	nano-(10^{-9})

Chapter 1

Introduction to Proteome Analysis by Mass Spectrometry

The Human Genome Project and other genome sequencing programs have been successful in sequencing the complete genomes of different species. As such the sequence of every protein encoded by a sequenced gene should be known in principle.[1] However, the actual protein sequence is usually different from the DNA-translated sequence due to post-translational modification (PTM). Thus proteome analysis is necessary and important to characterize the actual protein. A proteome includes all of the proteins involved in a given biological organism or system at a given time, and proteomics represents the large-scale analysis of the entire proteome, particularly structure and function.[2-4] Mass spectrometry (MS) is an indispensable tool for proteome analysis and has been widely used in the proteomics field. MS-based proteomics has been rapidly developed as a result of more and more complete gene sequence databases as well as technical and conceptual advances, especially those in protein ionization methods.[4] It is one of the most powerful tools for proteome analysis, including protein identification, quantification, and characterization. The fast development of proteomics is also due to the ongoing developments in mass spectrometry.

Sample preparation before MS analysis is critical in proteome analysis. It generally includes: protein digestion by enzyme or chemicals; protein/peptide separation or enrichment; isotopic labeling of protein/peptide for quantification; and sample cleanup. Various sample preparation methods have been developed

for different purposes. My thesis is mainly focused on methods of quantification and protein sequencing. Technologies related to my work and background information will be introduced in this chapter.

1.1 Mass Analysis

1.1.1 Ionization Methods

Only charged species can be guided into mass spectrometry and analyzed by mass to charge (m/z) ratios. As such ionization is very important for samples introduced into the mass spectrometer. Currently two ionization methods are widely used in the proteomic field: electrospray (ESI) and matrix-assisted laser desorption/ionization (MALDI).

1.1.1.1 Electrospray Ionization (ESI)

ESI is usually coupled to liquid chromatography (LC) to ionize analytes separated by LC. Figure 1.1 shows a schematic view of the ESI process. Charged droplets form at the capillary tip. Because of the high electric field at the capillary tip, the effluent continuously eluted from LC forms a Taylor cone, where a liquid filament forms due to the instability of the cone. As the downstream liquid filament becomes more and more unstable, the charged droplets eject from the capillary tip. These charged droplets gradually shrink to smaller charged droplets by solvent evaporation. When evaporation reaches to a certain extent, Coulombic repulsion forces would be sufficient to overcome surface tension, and an uneven

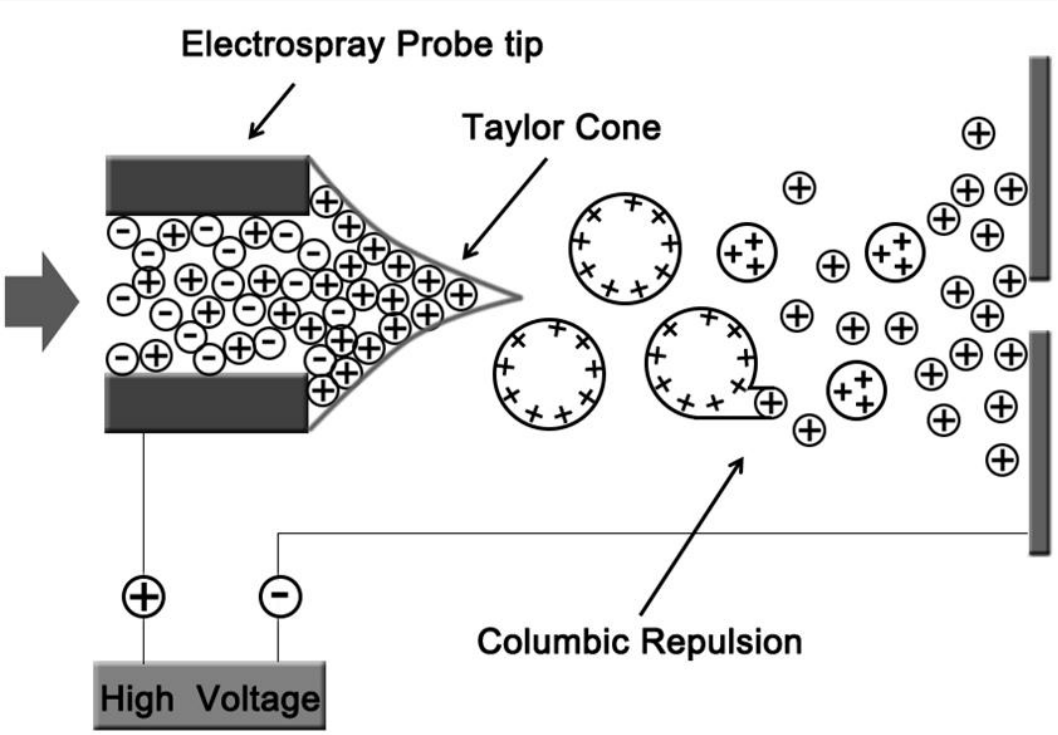


Figure 1.1 A schematic view of electrospray ionization process.

fission would happen. The charged droplets then split into a parent and offspring droplets. Afterwards the solvent evaporation as well as fission work nonstop to form extremely small droplets containing only one ion and gas ions are generated by subsequent evaporation (charge residue model). Alternatively, gas ions could be formed directly from droplets without further fission to extremely small droplets (ion evaporation model).[5-9]

1.1.1.2 Matrix-Assisted Laser Desorption/Ionization (MALDI)

Analytes are ionized with the assistance of a matrix in MALDI. Instead of utilizing LC separation in ESI, samples are deposited onto a solid MALDI plate with a matrix. The ionization process of MALDI is demonstrated in Figure 1.2. Co-crystallized matrices are ionized after absorbing laser energy through electronic or vibrational excitation. Sputtered analytes are desorbed from the plate with matrices and collide with the ionized matrix molecules to become protonated.[10-12] A variety of MALDI matrices have been developed. In general matrices are small organic molecules with labile protons which can be easily ionized by a laser beam and can trap analytes inside upon crystallization. α-Cyanohydroxycinnamic acid (CHCA/HCCA) and 2,5-dihydroxybenzoic acid (DHB) are two of the most commonly used MALDI matrices. The structures of the two are shown in Figure 1.3. Depending on different analytes, one matrix may give better signals than another. Usually the choice of the matrix is determined empirically.

1.1.2 MS Instrumentation

1.1.2.1 Quadrupole

Quadrupoles are constituted of four parallel cylindrical rods in a square configuration. A direct current (DC) and an alternating current/radio frequency (AC/RF) are applied to the two paired opposite rods to create a hyperbolic field. When DC and AC voltages ramp together at a particular ratio, ions of interest are transmitted one by one from low m/z to high m/z, while other ions are lost in

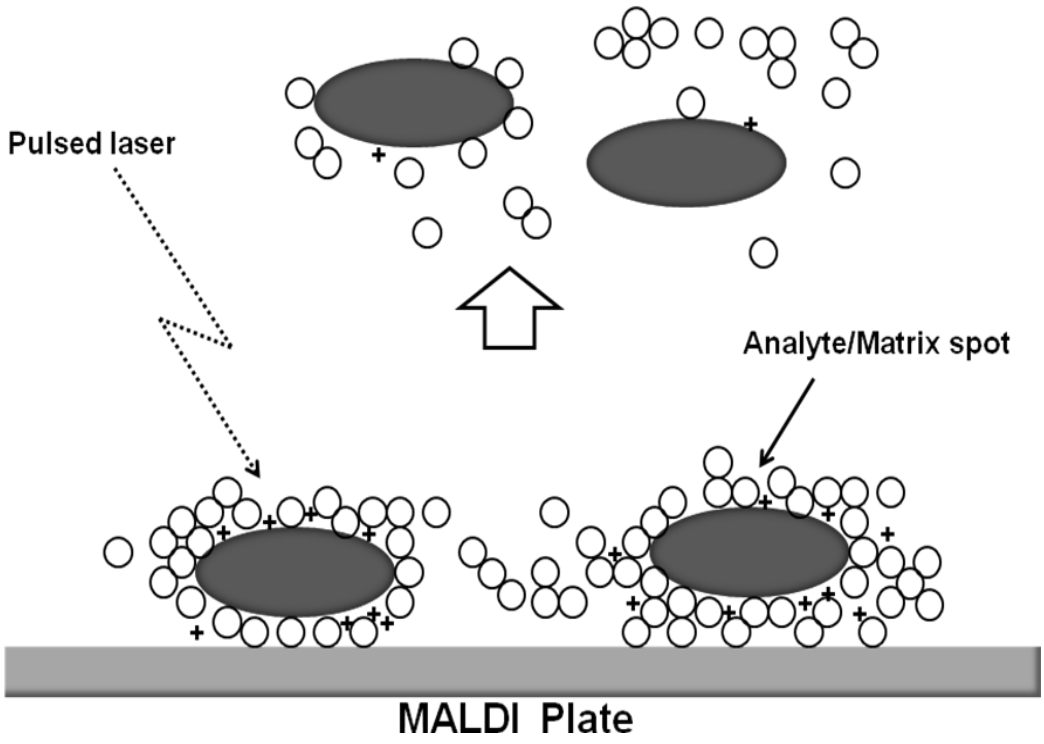


Figure 1.2 A schematic view of matrix-assisted laser ionization/desorption (MALDI) process.

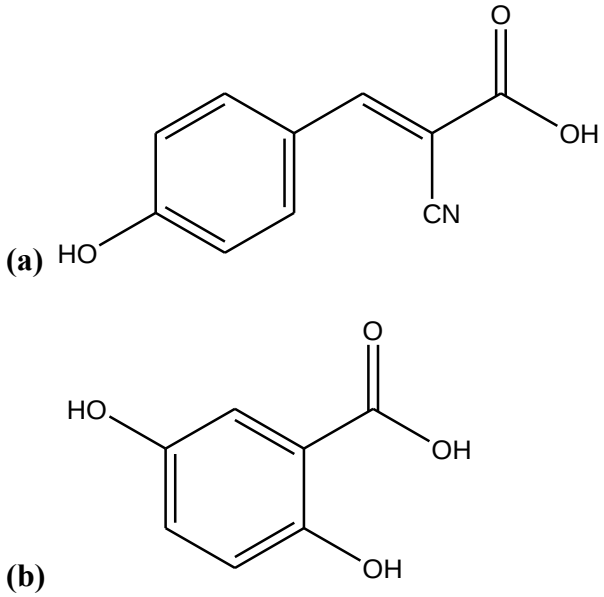


Figure 1.3 Structures of the two most commonly used MALDI matrixes: (a) α-cyanohydroxycinnamic acid (CHCA/HCCA); and (b) 2, 5-dihydroxybenzoic acid (DHB).

quadrupoles due to unstable trajectories. Quadrupoles can also act as a broad bandpass mass filter if the DC and AC voltages are adjusted to a certain value according to the mass range of ions passed. In an extreme case, all ions could pass the quadrupoles and be focused when only the AC voltage is applied to the quadrupoles. This is called "quadrupole ion focusing lens" or "quadrupole ion guide". The number of cylindrical rods can be increased to six or eight from four for better ion transfer, namely hexapole or octopole.[13-15]

1.1.2.2 Time-of-Flight (TOF)

TOF is a popular mass analyzer in mass spectrometry, readily coupled to GC/ LC and pulsed ionization methods such as MALDI. It is highly sensitive with all ions transmitted to the detector and its mass accuracy is as high as 5 ppm with good calibrations. The flight time of an ion in the field-free region is directly related to m/z of the ion. The larger the m/z of the ion, the longer the flight time. TOF mass spectrometer is usually equipped with a reflectron (or ion mirror) of several ring electrodes placed at the end of the drift tube (Figure 1.4). The voltage applied to the ring electrodes gradually increases when ions fly into the reflectron. The speed of the ions decreases to zero at a point, and then accelerates in the opposite direction. The ions with higher kinetic energy would penetrate deeper into the reflectron, and therefore spend longer time in the reflectron. In this way, the initial kinetic energy distribution is corrected and the resolution of the mass analyzer is improved.[16-18]

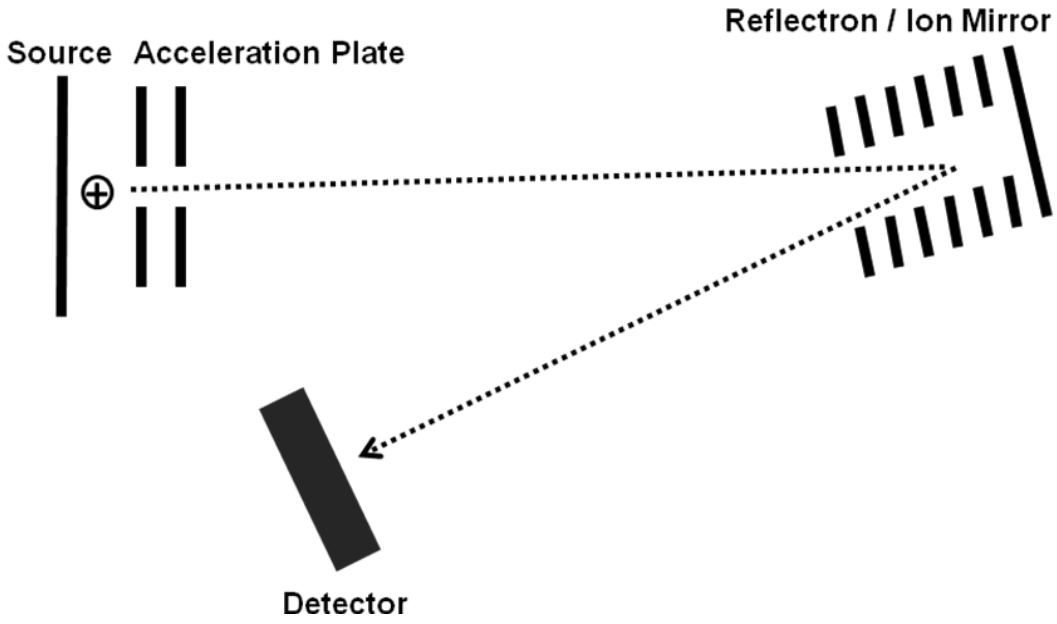


Figure 1.4 A schematic diagram of a reflectron TOF mass spectrometer.

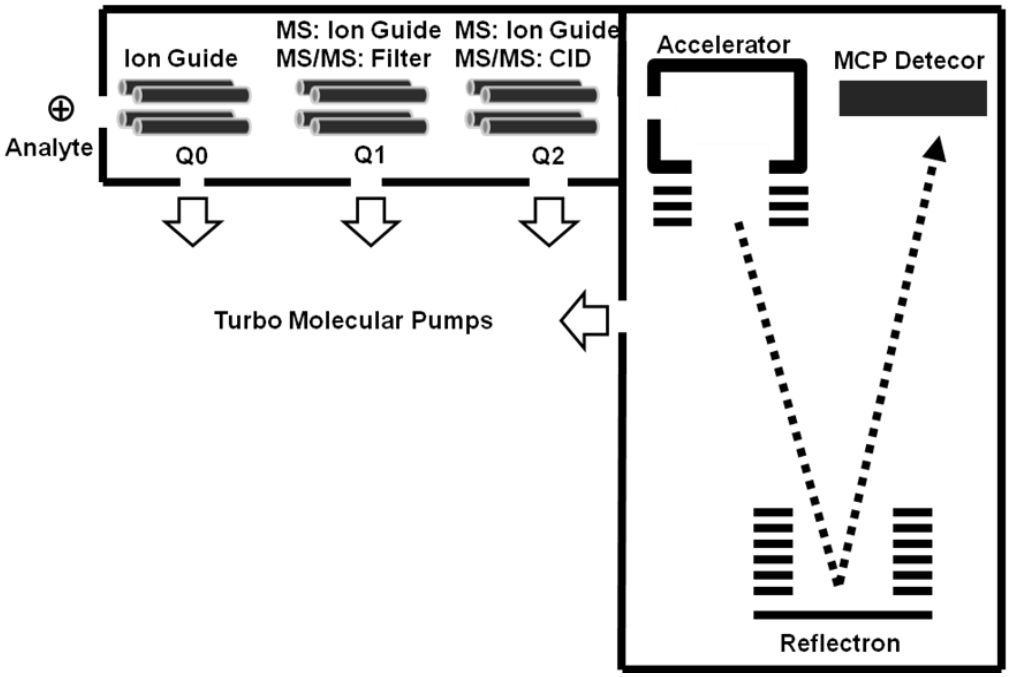


Figure 1.5 A schematic diagram of MALDI/ESI Q-TOF.

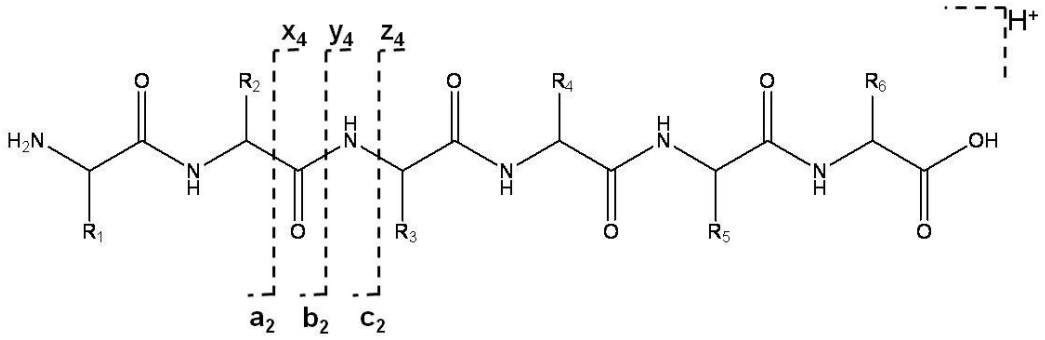
1.1.2.3 MALDI/ESI Q-TOF

The quadrupole time-of-flight tandem mass spectrometer (Q-TOF) is one of the most popular mass analyzers used in proteomics nowadays. When coupled to either MALDI or ESI, this instrument demonstrates powerful ability for proteome analysis. A schematic diagram of a Q-TOF is shown in Figure 1.5. A Q-TOF generally consists of three quadrupoles and a TOF mass analyzer. The first quadrupole (Q_0) is an RF-only ion guide, transmitting and focusing all of the ions. In MS mode, the second quadrupole (Q_1) and the third quadrupole (Q_2) act as RF-only ion guides, the same as Q_0. However, in MS/MS mode, Q_1 becomes a mass filter to select the parent ion and Q_2 is the collision cell where the parent ion fragments often collide with a gas such as N_2 or Ar.[19-21] Ions are then transmitted to the TOF analyzer in a 90° flight path, namely orthogonal TOF, correcting the kinetic energy distribution of the ions to increase the ion resolution.[20]

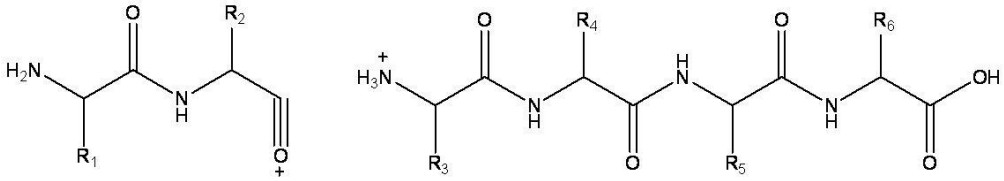
1.1.3 Protein/peptide Fragmentation

Generally, protein/peptide fragmentation happens along the amide backbone. Depending on the position where a protein/peptide breaks, fragment ions are named differently, as illustrated in Figure 1.6a.[22,23] Only fragments carrying charges are detected. N-terminal fragment ions are classified as a, b, and c ions, while C-terminal ions as x, y and z ions. The subscript indicates the number of residues in the fragment. Thus, fragments formed by cleavages before the amide bond are a and x ions, whereas breakage of the amide bonds form b and y ions. The b and y ions are the dominate species in low energy (10 – 100 eV) collision-induced dissociation (CID) or collision-activated dissociation (CAD) spectra.

(a)

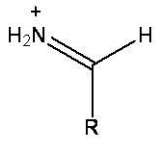


(b)



b_2 y_4

(c)



Immonium Ion

Figure 1.6 (a) Fragmentation pattern of a peptide ion; (b) the structure of fragmented b_2 and y_4 ions of a peptide ion example; and (c) the general structure of an immonium ion.

Many mass spectrometers adopt low energy CID to generate tandem mass spectra, like triple quadrupole, ion trap, and Q-TOF.[24-27] An example of b and y ions from a hexapeptide is demonstrated in Figure 1.6b. Fragments c and z ions are generated by cleavages after the amide bond, the most common in electron transfer dissociation (ETD), electron capture dissociation (ECD), and MALDI in-source decay (ISD).[28-33] In addition to fragmentation along the amide backbone, sometimes an internal fragment called the immonium ion can be formed by a combination of a type and y type cleavage (shown in Figure 1.6c). The immonium ion is an analogue of a single amino acid (without the carboxylic acid), and it is labeled with the one letter code of its corresponding amino acid.

1.1.4 Database Search

In a database search, both the precursor m/z and the product m/z of the tandem mass spectra are compared with the theoretic m/z generated by computer using the known protein sequences in the database as well as protein/peptide fragmentation patterns and/or protein digestion methods. Various algorithms have been developed to score the peptide by different search engines such as MASCOT[34], X!Tandem[35], and SEQUEST[36]. In Mascot, the higher the score, the more confident the search result would be. The identity threshold for a peptide/protein hit depends on the chosen confidence level, the mass tolerance window, and the selected database. Generally, the highest score protein/peptide passing the identity threshold is considered to be the correct match. Certain search parameters can be defined by users in the database search, such as mass accuracy of the precursor and product ions. Changes from either post translational

modifications (PTMs) or chemical labelings can also be defined and incorporated into the search to determine the actual location of the modifications.

1.2 Methods for Proteome Identification

Protein identification is to generate the protein ID in a given sample, also called proteome profiling. MS based protein identification methods can be categorized as either a bottom-up (shotgun) method or a top-down method.

1.2.1 Bottom-up (Shotgun) Method

The workflow of shotgun proteomics is shown in Figure 1.7. Proteins extracted from tissues, cell lines, body fluids or other biological sources are digested into peptides by enzymes or chemicals. The peptide mixtures generated are separated to various fractions by liquid chromatography (LC), and then analyzed on the mass spectrometer. Collected spectra are searched against the database to identify the peptides and proteins. For high molecular mass, fragmentation sufficient for protein identification becomes difficult. Therefore digesting a large protein to peptides and identifying the protein through lower molecular weight peptides is necessary for increasing proteome coverage.

The most commonly used enzyme for protein digestion is trypsin, a highly specific enzyme cutting at the C-terminus of lysine or arginine unless the next residue is proline.[37-39] Because lysine and arginine are abundant in protein sequences, tryptic digested peptides are usually within the mass range of 500 to

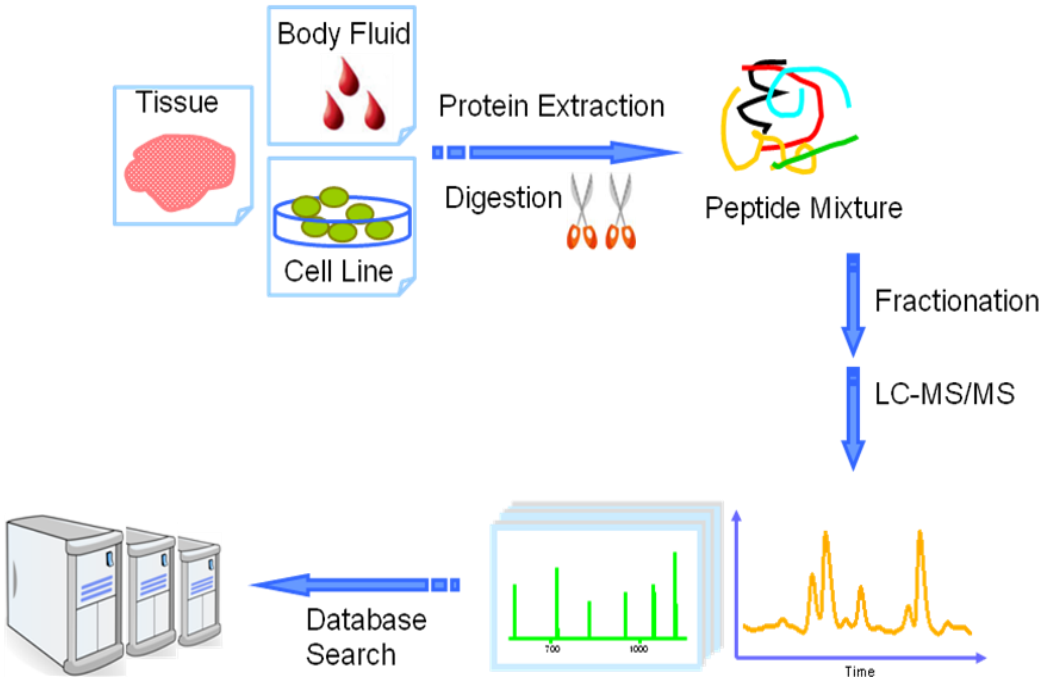


Figure 1.7 The workflow of shotgun proteomics.

3000 dalton with an average length of ~14 amino acids, facilitating MS analysis and MS/MS fragmentation.[40] Other enzymes such as chymotrypsin,[41] Lys-C,[42] Lys-N,[43] and Glu-C[44] are used in protein digestion as well. Chemical methods with different specificities are also widely used for protein digestions, and are complimentary to existing enzymatic methods. For example, cyanogen bromide (CNBr) cleaves at the C-terminal of methionine[45]; diluted solutions of hydrochloric acid (HCl),[46] formic acid (FA)[47-49] or acetic acid (AcOH)[50] have been reported to cleave at the C-terminal of aspartyl (Asp) residues.

The shotgun proteomics allows high data throughput, easy automation, and high protein detection sensitivity, as well as the sequencing capability of large proteins for proteome analysis.[51] However, it is time-consuming (i.e. digestion, liquid chromatography, desalting), the N- (or C-) terminal information is often lacking, and post-translational modifications (PTMs) are hardly detected.[52]

1.2.2 Top-down Method

An alternative strategy for proteome analysis, introduced by McLafferty *et al*., is the "top-down" method, with identifies proteins using accurate mass measurement and/or tandem mass spectrometry.[53] The "top-down" approach, utilizing MALDI in-source decay (ISD),[33] ESI electron capture dissociation (ECD) or electron transfer dissociation (ETD)[28-31] has been proved to be efficient in sequencing mid-size proteins and identifying PTMs,[54-56] but this method is limited in the numbers of identified proteins from a given organism due to a limited dynamic range, as well as a lack of sophisticated bioinformatic tools for data analysis.[53] The top-down method can only analyze relatively simple samples, so protein separation by SDS-PAGE or liquid chromatography before MS analysis is necessary. Recently, Neil Kelleher's group has used isoelectric focusing (IEF) combined with gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) to effectively fractionate nuclear and cytosolic extracts of HeLa S3 cells.[57] Followed by top-down proteome analysis, the method revealed a greater than 20-fold increase in the identification of intact proteins over any previous work in mammalian cells. This has been the most comprehensive implementation of top-down mass spectrometry so far.

1.2.3 Middle-down Method

A new variant method, so called 'middle-down', focuses on analyzing large peptide fragments (>3 kDa). As such it combines some benefits of both bottom-up and top-down approaches (e.g. identifying more peptides and multiple PTMs at the same time).[58] Limited proteolysis producing large peptide fragments in middle-down method can be obtained by rapid digestions with enzymes or chemicals in a short period of time (e.g. a few minutes).[59,60] A complex peptide mixture generated by bottom-up method contains peptides far more than even the most sofisticated instrument can analyze, whereas middle-down method produces fewer, but larger peptides, so the sample complexity can be reduced without compromising the information contained. The large peptides in the middle-down method (3000−20000 Da) have showed improved performance in separation, ionization and fragmentation in different instruments.[40] Compared with the bottom-up method, the middle-down method offers more confident peptide identifications with increased protein sequence coverage and identification of PTMs.[61]

1.3 Protein Quantification Methods

Quantification of proteins is useful in many fields, such as time-course studies or biomarker discovery.[62] Colorimetric assays (e.g. Biuret assay[63], Lowry assay[64], Bicinchoninic acid assay[65], and Bradford assay[66]) and spectroscopic methods are traditional ways for protein quantification. Since three aromatic acids (tryptophan,

tyrosine, and phenylalanine) give native fluorescence as well as UV-Vis absorbance, fluorescence and/or UV-Vis detection[67-69] have been successfully carried out for the quantification[70,71] of proteins. In addition to colorimetric assays and spectroscopic methods, MS-based quantitative proteomics has been developed and applied in many biological studies[72]. MS-based quantification methods can be categorized into two classes: label-free and label-based methods, depending on whether isotopes are incorporated into samples.

Label-free method quantifies different peptides/proteins by signal intensities or the number matched spectra of peptides/proteins.[73] This method is easy to operate without labeling and compatible with any type of sample. However, the disadvantages are that errors can arise from sample processing and MS analysis, as well as from interfering substances such as detergents and the abundant background proteins.[74] Thus the quantification is far less accurate than the label-based method.

Isotope labeling combined with MS analysis is widely used for quantification of proteins.[75-77] Quantification by isotopic labeling is accurate, has a generally low limit of quantification (LOQ) as a result of highly sensitive mass spectrometry analysis, and it is possible to identify and quantify every single peptide or protein. The cost resides in labor-intensive labeling steps and long analysis times. The stable isotope tags can be incorporated into the samples through *in vivo* metabolic labeling (e.g. SILAC: stable isotope labeling by amino acids in the cell culture), or *in vitro* by chemical or enzymatic means (e.g. trypsin digestion in oxygen-18 water).[62] As a universal quantification method, chemical labeling can be applied

to any protein sample. Various chemical labeling methods, either isobaric or non-isobaric, have been developed in the past few years.[75-78]

Most chemical labelings target the reactive functional groups of a given protein, like the primary amine of the N-terminus, the side chain amine of lysine, and the thiol group of cysteine.[79] One of the most sophisticated non-isobaric methods, the isotope-coded affinity tag (ICAT) approach developed by Gygi *et al.*, involves derivatizing cysteine residues specifically with a reagent containing zero or eight deuterium isotopes and introducing a biotin group for affinity purification and a linker cleavable for subsequent MS analysis.[75] ICAT is only applicable for cysteine-containing proteins/peptides. As such it leaves non-cysteine containing proteins alone and thereby reduces the sample complexity.

2-MEGA labeling reaction[80] has been used in our lab for peptide labeling. 2-MEGA stands for "N-terminal dimethylation after lysine guanidination". The reaction scheme is illustrated in Figure 1.8. Briefly, O-methylisourea selectively guanidinates lysines within peptides, blocking these residues from further reaction. In the presence of fomaldehyde and borohydride, the N termini of peptides are dimethylated. With common formaldehyde, the mass of the peptide increases ~28Da. While using heavy formaldehyde, the mass of the peptide increases ~34Da, 6Da heavier than the light labeled peptide. Compared to other methods, 2-MEGA is an indiscriminate peptide labeling method and inexpensive. It produces peptides with a fixed mass shift and increases the labeling percentage of lysine-containing peptides.[76]

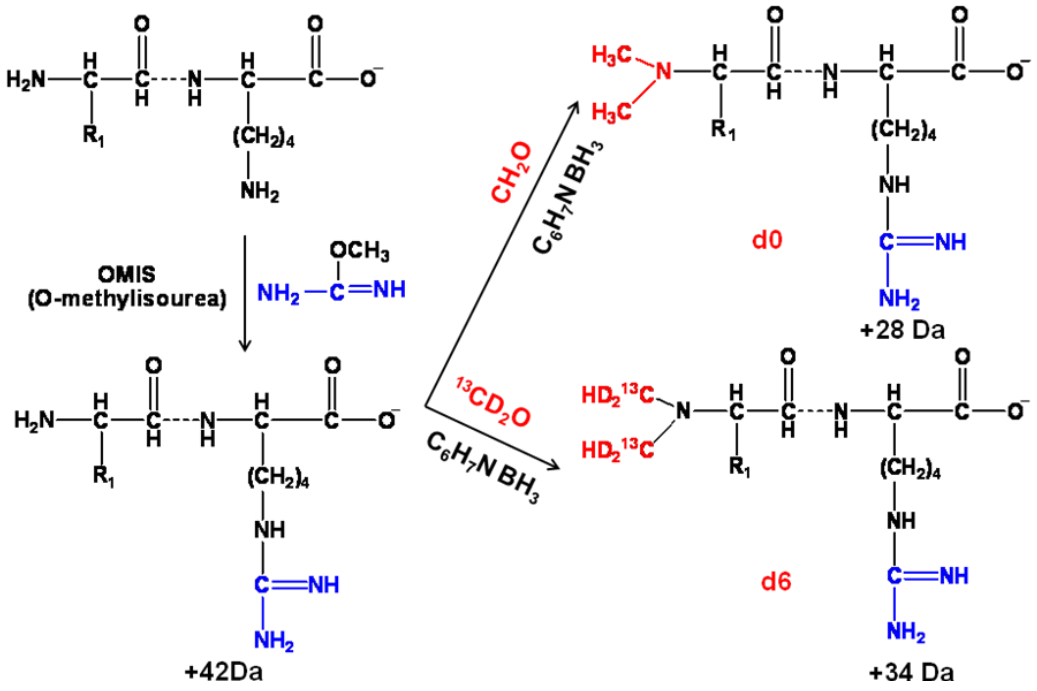


Figure 1.8 Reaction scheme of 2-MEGA labeling reaction.

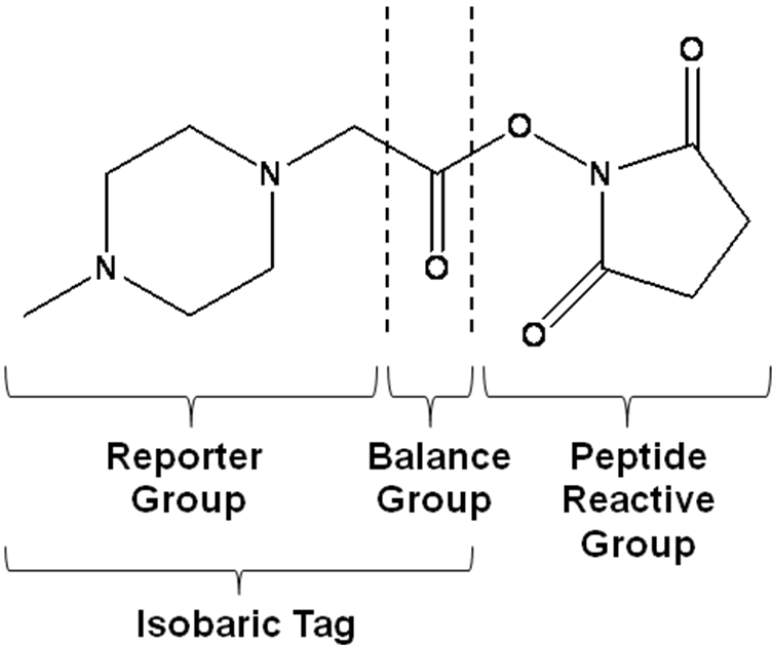


Figure 1.9 Structure of isobaric labeling reagent iTRAQ.

Isobaric stable isotope labeling uses isobaric N-hydroxy succinimide (NHS)-activated derivatives to modify primary amino groups in peptides/proteins. Each reagent possesses a reporter group for quantification and a mass balance group for mass difference adjustment.[81] In most chemical labeling techniques, relative quantification is achieved by chromatogram peak area comparisons of peptides labeled by different isotopes. An isobaric mass tagging first introduced by Thompson and co-workers is different in that quantifications are performed at the MS/MS stage by the reporter ion cleaved in peptide fragmentations.[73] The chromatogram of isobaric labeling is much simpler than non-isobaric ones especially for multiplex quantification. This is because labeled peptides co-elute in liquid chromatography, which reduces overlaps of different peptides and enhances the signals of co-eluted labeled peptide pairs. However, different labeled peptide pairs must be separated to avoid interferences by other peptides of similar masses and fragmentation must be sufficient to generate the reporter ion and peptide backbone fragment ion at the same time, otherwise the identity and quantification ratio of labeled peptide pairs cannot be simultaneously determined in tandem mass spectrum. Two commercial isobaric labeling reagents are available: iTRAQ (isobaric tags for relative and absolute quantification) and TMT (tandem mass tags).[77,78] iTRAQ is the most popular reagent for proteome comparison, with parallel comparisons of eight samples at the most. The structure of iTRAQ is shown in Figure 1.9.

1.4 Protein Sequencing Methods in Mass Spectrometry

A protein sequence can be deduced from the corresponding DNA sequence in principal. However, the actual protein sequence is usually different from what is translated directly from the DNA sequence. Different translation products are generated by changed coding genes through RNA splicing, shuffling, and other processing mechanisms.[82] Post-translational modifications (PTMs) of amino acid residues [e.g. phosphorylation of side-chain hydroxyl groups (serine, threonine and tyrosine) [83], acetylation at protein N-terminus[84], methylation of arginine/lysine[85], and proteolytic processing by exopetidases/endopeptidases] are also the differences in protein sequences.[86] Protein terminal truncation is common for converting a protein into its functional form but artifacts in the purification process can sometimes induce truncation of the protein termini.[87] Protein sequences are important for functional studies, essential for a better understanding of what is going on in biological systems. As such protein sequence analysis is vital in many areas of research and applications. For example, protein-based drugs are produced in large scale in industry, and the production process needs to be closely monitored to ensure the correct sequence is expressed.

Edman degradation has been a gold standard method for protein N-terminal sequencing since 1967.[88] This method is accurate and efficient for up to 50 – 60 residues, with a reduced efficiency after that. It is only applicable to proteins with intact (unblocked or unmodified) N-termini [89] and the process is rather time-consuming. For protein C-terminal sequencing, the Schlack-Kump degradation with (iso)thiocyanate[90,91] is a reaction similar to Edman degradation. The

disadvantages of this method are low efficiency and reproducibility; and only less than ten C-terminal residues could be obtained.[92] In comparison, MS-based methods are fast and sensitive, and can also identify protein modifications. Currently, MS approach for protein sequencing is in development. For example, studies have shown that proteins can be digested by exopeptidases like carboxypeptidase Y/B (CPY/CPB)[93,94] and leucine aminopeptidase M (LAP)[95]; subsequent analysis of the cleaved peptides differing by one amino acid by matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF)[93,94] or tandem mass spectrometry[92,96] revealed the sequence of these proteins. Chait *et al.*[97] generated a "protein sequencing ladder" using phenylisothiocyanate (PITC, the reagent used in Edman degradation) and phenylsiocyanate (PIC, a stable terminator not to be cleaved from the peptide backbone) instead of exopeptidases. Protein sequences were read out by MALDI/TOF analysis of the peptide ladder mixtures.

Alternatively, top-down proteomics provides another way for protein sequencing. MALDI in-source decay (ISD),[33] ESI electron capture dissociation (ECD) and electron transfer dissociation (ETD)[28-31] are efficient in identifying terminal peptide ladders of mid-size proteins. The protein sequences can be obtained from fragments of peptide ladders.

Recently, the newly developed microwave-assisted acid hydrolysis (MAAH) MS has been applied as an effective method for protein sequencing. Within several minutes, the proteins are digested into peptides with the help of acid and microwave irradiation, and data analysis is very simple. This method was first

introduced by Zhong *et al.*[98] and they demonstrated that after one minute microwave irradiation exposure, the purified protein was hydrolyzed to predominantly terminal peptides in the presence of hydrochloric acid (HCl). Subsequent MALDI-TOF analysis revealed both N- and C-terminal series of polypeptide ladders. In comparision, without the assistance of the microwave irradiation, acid hydrolysis of the same protein took 15 h to generate intense terminal peptide peaks at room temperature. The complete protein sequence as well as the PTMs was then obtained by analyzing the polypeptide ladders. Only the most abundant ions, the terminal peptides were detected because of the relatively low dynamic range of MALDI-TOF and ion suppression in MALDI.[31] This technique could rapidly characterize both termini of the protein (even blocked N-termini) at the same time as well as PTM identifications and the results are reproducible. However, this method cannot completely replace the Edman degradation, for it is difficult to identify truncated N-termini which are easily recognized by the Edman method and it is only applicable to sequence small proteins (MW<2 kDa).[31]

In addition to ladder-based protein sequencing methods, *de novo* sequencing[29,99,100] of peptides by automated interpretation of MS/MS spectra without the assistance of protein sequence database is currently a hot topic in development. By combining all the peptide sequences identified, it is possible to piece together the whole protein sequence. Diverse proteases can be utilized together to generate different sets of peptides. Since these peptides would overlap with each other, two sequences could be stitched by an overlap of three amino

acids [a motif of three amino acids is in principal unique for a middle-size protein (20–60 kDa)].[101]

1.5 Scope of the Thesis

In Chapter 2 and 4, I show the automation of the 2-MEGA labeling method for protein quantification and application of 2-MEGA labeling for quantification of proteins in breast cancer tissues to discovery potential biomarkers, respectively. In Chapter 3, I describe a LC-UV method to quickly quantify proteins in a small number of cells for checking sample integrity. In Chapter 5, I demonstrate a series of problems encountered with purifying electroeluted proteins and digesting these proteins by HCl MAAH for MS analysis. In Chapter 6 to 8, I describe improvements of the HCl MAAH method for protein sequencing and the development of an in-gel HCl MAAH method. Conclusions and related future work are included in the last chapter of this thesis.

1.6 Literature Cited

(1) Aebersold, R.; Goodlett, D. R. *Chem Rev* **2001**, *101*, 269.
(2) Daoud, S. S.; Totowa, N.J.: Humana Press: 2008.
(3) Langley, S. R.; Dwyer, J.; Drozdov, I.; Yin, X.; Mayr, M. *Cardiovascular Research* **2013**, *97*, 612.
(4) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198.
(5) Kebarle, P.; Tang, L. *Anal Chem* **1993**, *65*, 972A.
(6) Dole, M.; Mack, L. L.; Hines, R. L. *J. Chem. Phys.* **1968**, *49*, 2240.
(7) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287.
(8) Banerjee, S.; Mazumdar, S. *International Journal of Analytical Chemistry* **2012**, *2012*, 282574.

(9) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. *Anal Chem* **2013**, *85*, 2.
(10) Sunner, J.; Kulatunga, R.; Kebarle, P. *Anal Chem* **1986**, *58*, 1312.
(11) Knochenmuss, R.; Dubois, F.; Dale, M.; Zenobi, R. *Rapid Communications in Mass Spectrometry* **1996**, *10*, 871.
(12) Lehmann, E.; Knochenmuss, R.; Zenobi, R. *Rapid Communications in Mass Spectrometry* **1997**, *11*, 1483.
(13) Campana, J. E. *International Journal of Mass Spectrometry and Ion Processes* **1980**, *33*, 101.
(14) Muntean, F. *International Journal of Mass Spectrometry and Ion Processes* **1995**, *151*, 197.
(15) Campana, J. E.; Jurs, P. C. *International Journal of Mass Spectrometry and Ion Processes* **1980**, *33*, 119.
(16) Cornish, T. J.; Cotter, R. J. *Rapid Communications in Mass Spectrometry : RCM* **1993**, *7*, 1037.
(17) Boesl, U.; Weinkauf, R.; Schlag, E. W. *International Journal of Mass Spectrometry and Ion Processes* **1992**, *112*, 121.
(18) Cornish, T. J.; Cotter, R. J. *Rapid Communications in Mass Spectrometry : RCM* **1994**, *8*, 781.
(19) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *J Mass Spectrom* **2001**, *36*, 849.
(20) Lacorte, S.; Fernandez-Alba, A. R. *Mass Spectrom Rev* **2006**, *25*, 866.
(21) Bateman, R. H.; Carruthers, R.; Hoyes, J. B.; Jones, C.; Langridge, J. I.; Millar, A.; Vissers, J. P. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 792.
(22) Roepstorff, P.; Fohlman, J. *Biol Mass Spectrom* **1984**, *11*, 601.
(23) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Anal Chem* **1987**, *59*, 2621.
(24) Sobott, F.; Watt, S. J.; Smith, J.; Edelmann, M. J.; Kramer, H. B.; Kessler, B. M. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 1652.
(25) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. *Mol Cell Proteomics* **2011**, *10*, M111 009910.
(26) Chowdhury, S. M.; Munske, G. R.; Ronald, R. C.; Bruce, J. E. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 493.
(27) Kim, M. S.; Zhong, J.; Kandasamy, K.; Delanghe, B.; Pandey, A. *Proteomics* **2011**, *11*, 2568.
(28) Sun, R. X.; Dong, M. Q.; Chi, H.; Yang, B.; Xiu, L. Y.; Wang, L. H.; Fu, Y.; He, S. M. *Prog Biochem Biophys* **2010**, *37*, 94.
(29) Standing, K. G. *Curr Opin Struc Biol* **2003**, *13*, 595.
(30) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *P Natl Acad Sci USA* **2004**, *101*, 9528.
(31) Lill, J. R.; Ingle, E. S.; Liu, P. S.; Pham, V.; Sandoval, W. N. *Mass Spectrom Rev* **2007**, *26*, 657.
(32) Kocher, T.; Engstrom, A.; Zubarev, R. A. *Anal Chem* **2005**, *77*, 172.
(33) Resemann, A.; Wunderlich, D.; Rothbauer, U.; Warscheid, B.; Leonhardt, H.; Fuchser, J.; Kuhlmann, K.; Suckau, D. *Anal Chem* **2010**, *82*, 3283.

(34) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551.
(35) Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20*, 1466.
(36) Eng, J. K.; McCormack, A. L.; Yates, J. R., III *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976.
(37) Song, Z.; Zhang, Q.; Zhang, Y.; Qin, W.; Qian, X. *Chinese Journal of Chromatography* **2012**, *30*, 549.
(38) Sleczka, B. G.; D'Arienzo, C. J.; Tymiak, A. A.; Olah, T. V. *Bioanalysis* **2012**, *4*, 29.
(39) Wu, S. T.; Ouyang, Z.; Olah, T. V.; Jemal, M. *Rapid Communications in Mass Spectrometry : RCM* **2011**, *25*, 281.
(40) Switzar, L.; Giera, M.; Niessen, W. M. *J Proteome Res* **2013**, *12*, 1067.
(41) Rupley, J. A.; Scheraga, H. A. *Biochimica et Biophysica Acta* **1960**, *44*, 191.
(42) Glatter, T.; Ludwig, C.; Ahrne, E.; Aebersold, R.; Heck, A. J.; Schmidt, A. *J Proteome Res* **2012**, *11*, 5145.
(43) Taouatas, N.; Mohammed, S.; Heck, A. J. *Methods Mol Biol* **2011**, *753*, 157.
(44) Kalli, A.; Hakansson, K. *J Proteome Res* **2008**, *7*, 2834.
(45) Meyer, B.; Papasotiriou, D. G.; Karas, M. *Amino Acids* **2011**, *41*, 291.
(46) Lo, L. H.; Huang, T. L.; Shiea, J. *Rapid communications in mass spectrometry : RCM* **2009**, *23*, 589.
(47) Hua, L.; Low, T. Y.; Sze, S. K. *Proteomics* **2006**, *6*, 586.
(48) Inglis, A. S. *Methods in enzymology* **1983**, *91*, 324.
(49) Li, A.; Sowder, R. C.; Henderson, L. E.; Moore, S. P.; Garfinkel, D. J.; Fisher, R. J. *Anal Chem* **2001**, *73*, 5395.
(50) Swatkoski, S.; Gutierrez, P.; Ginter, J.; Petrov, A.; Dinman, J. D.; Edwards, N.; Fenselau, C. *J Proteome Res* **2007**, *6*, 4525.
(51) Nesvizhskii, A. I. *Methods Mol Biol* **2007**, *367*, 87.
(52) Hardouin, J. *Mass Spectrom Rev* **2007**, *26*, 672.
(53) VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L., Jr. *J Proteome Res* **2002**, *1*, 239.
(54) Lanucara, F.; Eyers, C. E. *Mass Spectrom Rev* **2013**, *32*, 27.
(55) Zhang, H.; Ge, Y. *Circulation. Cardiovascular genetics* **2011**, *4*, 711.
(56) Cui, W.; Rohrs, H. W.; Gross, M. L. *The Analyst* **2011**, *136*, 3854.
(57) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254.
(58) Siuti, N.; Kelleher, N. L. *Nature Methods* **2007**, *4*, 817.
(59) Tan, Y. J.; Wang, W. H.; Zheng, Y.; Dong, J.; Stefano, G.; Brandizzi, F.; Garavito, R. M.; Reid, G. E.; Bruening, M. L. *Anal Chem* **2012**, *84*, 8357.
(60) Wu, C.; Tran, J. C.; Zamdborg, L.; Durbin, K. R.; Li, M.; Ahlf, D. R.; Early, B. P.; Thomas, P. M.; Sweedler, J. V.; Kelleher, N. L. *Nature Methods* **2012**, *9*, 822.

(61) Wu, S. L.; Kim, J.; Hancock, W. S.; Karger, B. *J Proteome Res* **2005**, *4*, 1155.
(62) Wong, J. W.; Cagney, G. *Methods Mol Biol* **2010**, *604*, 273.
(63) Tichy, M.; Friedecky, B.; Budina, M.; Maisnar, V.; Buchler, T.; Holeckova, M.; Gotzmannova, D.; Palicka, V. *Clinical Chemistry and Laboratory Medicine : CCLM / FESCC* **2009**, *47*, 235.
(64) Upreti, G. C.; Wang, Y.; Finn, A.; Sharrock, A.; Feisst, N.; Davy, M.; Jordan, R. B. *BioTechniques* **2012**, *52*, 159.
(65) Bainor, A.; Chang, L.; McQuade, T. J.; Webb, B.; Gestwicki, J. E. *Anal Biochem* **2011**, *410*, 310.
(66) Ku, H. K.; Lim, H. M.; Oh, K. H.; Yang, H. J.; Jeong, J. S.; Kim, S. K. *Anal Biochem* **2013**, *434*, 178.
(67) Lee, T. T.; Yeung, E. S. *Journal of Chromatography* **1992**, *595*, 319.
(68) Sluszny, C.; He, Y.; Yeung, E. S. *Electrophoresis* **2005**, *26*, 4197.
(69) Heath, T. G.; Giordani, A. B. *Journal of Chromatography* **1993**, *638*, 9.
(70) Saraswat, S.; Snyder, B.; Isailovic, D. *Journal of Chromatography B,* **2012**, *902*, 70.
(71) Russell, J. D.; Hilger, R. T.; Ladror, D. T.; Tervo, M. A.; Scalf, M.; Shortreed, M. R.; Coon, J. J.; Smith, L. M. *Anal Chem* **2011**, *83*, 2187.
(72) Zhang, G.; Ueberheide, B. M.; Waldemarson, S.; Myung, S.; Molloy, K.; Eriksson, J.; Chait, B. T.; Neubert, T. A.; Fenyo, D. *Methods Mol Biol* **2010**, *673*, 211.
(73) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Anal Bioanal Chem* **2007**, *389*, 1017.
(74) Ong, S. E.; Mann, M. *Nature Chemical Biology* **2005**, *1*, 252.
(75) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat Biotechnol* **1999**, *17*, 994.
(76) Ji, C.; Lo, A.; Marcus, S.; Li, L. *Journal of Proteome Research* **2006**, *5*, 2567.
(77) Pichler, P.; Kocher, T.; Holzmann, J.; Mohring, T.; Ammerer, G.; Mechtler, K. *Anal Chem* **2011**, *83*, 1469.
(78) Pichler, P.; Kocher, T.; Holzmann, J.; Mazanek, M.; Taus, T.; Ammerer, G.; Mechtler, K. *Anal Chem* **2010**, *82*, 6549.
(79) Julka, S.; Regnier, F. *Journal of Proteome Research* **2004**, *3*, 350.
(80) Ji, C.; Guo, N.; Li, L. *Journal of Proteome Research* **2005**, *4*, 2099.
(81) Hung, C. W.; Tholey, A. *Anal Chem* **2012**, *84*, 161.
(82) Yamaguchi, M.; Nakayama, D.; Shima, K.; Kuyama, H.; Ando, E.; Okamura, T. A.; Ueyama, N.; Nakazawa, T.; Norioka, S.; Nishimura, O.; Tsunasawa, S. *Rapid Communications in Mass Spectrometry* **2008**, *22*, 3313.
(83) Vargas-Suarez, M.; Castro-Sanchez, A.; Toledo-Ortiz, G.; Gonzalez de la Vara, L. E.; Garcia, E.; Loza-Tavera, H. *Biochimie* **2013**, *95*, 400.
(84) Lu, J. Y.; Lin, Y. Y.; Zhu, H.; Chuang, L. M.; Boeke, J. D. *Aging* **2011**, *3*, 911.
(85) Weiss, D.; Stockmann, C.; Schrodter, K.; Rudack, C. *Cell Oncol (Dordr)* **2013**.

(86) Song, H.; Hecimovic, S.; Goate, A.; Hsu, F. F.; Bao, S.; Vidavsky, I.; Ramanadham, S.; Turk, J. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 1780.
(87) Xiang, B. S.; Macisaac, S.; Lardizabal, K.; Li, B. *Rapid Communications in Mass Spectrometry* **2010**, *24*, 3447.
(88) Edman, P.; Begg, G. *Eur J Biochem* **1967**, *1*, 80.
(89) Stevenson, T. I.; Loo, J. A.; Greis, K. D. *Anal Biochem* **1998**, *262*, 99.
(90) Stark, G. R. *Biochemistry* **1968**, *7*, 1796.
(91) Boyd, V. L.; Bozzini, M.; Guga, P. J.; Defranco, R. J.; Yuan, P. M.; Loudon, G. M.; Nguyen, D. *J Org Chem* **1995**, *60*, 2581.
(92) Gao, Y. F.; Wang, H. X. *Chinese J Anal Chem* **2007**, *35*, 1820.
(93) Schar, M.; Bornsen, K. O.; Gassmann, E. *Rapid Communications in Mass Spectrometry* **1991**, *5*, 319.
(94) Patterson, D. H.; Tarr, G. E.; Regnier, F. E.; Martin, S. A. *Anal Chem* **1995**, *67*, 3971.
(95) Doucette, A.; Li, L. *Proteomics* **2001**, *1*, 987.
(96) Samyn, B.; Sergeant, K.; Castanheira, P.; Faro, C.; Van Beeumen, J. *Nature Methods* **2005**, *2*, 193.
(97) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. *Science* **1993**, *262*, 89.
(98) Zhong, H. Y.; Zhang, Y.; Wen, Z. H.; Li, L. *Nat Biotechnol* **2004**, *22*, 1291.
(99) Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *J Comput Biol* **1999**, *6*, 327.
(100) Hughes, C.; Ma, B.; Lajoie, G. A. *Methods Mol Biol* **2010**, *604*, 105.
(101) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. *Proteomics* **2010**, *10*, 634.

Chapter 2

Automation of Dimethylation-after-Guanidination (2-MEGA) Labeling Chemistry for High Throughput Mass Spectrometry-Based Shotgun Proteomics*

2.1 Introduction

Mass spectrometry (MS) based proteomics approaches have experienced a marked shift from simply identifying proteins to quantifying thousands of peptides from complex matrices to generate detailed quantitative information about proteome changes. The ability of MS to both identify and quantify thousands of components in a single experiment positions MS uniquely within the repertoire of researchers interested in a variety of biological processes and phenomena. With careful experimental design, quantitative information about alterations in a proteome resulting from a given perturbation or organism state can be obtained for a number of cellular processes, such as phosphorylation,[1] acetylation,[2,3] glycosylation[4] and protein production/degradation.[5,6] Although MS-based approaches can provide substantial amounts of data, increasing demands on the overall productivity of MS-based workflows for

*This work was done under the supervision of Dr. Andy Lo. A form of this chapter has been accepted for publication as: A. Lo, Y.N. Tang, L. Chen, and L. Li, 2013, "Automation of Dimethylation-after-Guanidination (2-MEGA) Labeling Chemistry for High Throughput Mass Spectrometry-Based Shotgun Proteomics", Analytica Chimica Acta, in press. Y.N. Tang and L. Chen contributed equally to this work.

analysis remains, particularly for large scale systems biology studies.[7,8] With recent advances in LC-MS instrumentation for rapid data generation, sample preparation methods that are inexpensive, reproducible and robust will become increasingly important for many quantitative proteomics applications.[9-11]

While a range of approaches for MS-based proteome quantification have been developed, they can be broadly categorized as either label-based or label-free, depending on whether or not isotopes are introduced for quantification. Regardless of the strategy employed, both methods offer their own advantages. Label-free methods typically use additional information from identified peptides across multiple runs, such as ion current intensity or the frequency of MS/MS sequencing, to determine relative changes between various samples.[12] Label-based methods utilize relative signal intensities from isotopically-encoded references.[13] Within the realm of label-based methods, various metabolic and chemical isotope incorporation methods exist alongside targeted approaches using standard addition of synthetically prepared isotopically labeled peptides.[13] The introduction of isotopes by metabolic or chemical derivatization methods is an additional experimental procedure that can be a source of variation in the observed ratio between samples in shotgun methods using LC-MS.

Chemical derivatization approaches are universally applicable to samples regardless of origin.[9,10] However, considerable care must be taken to reduce variation in the experimental steps before samples are combined for analysis. Any variation prior to sample mixing introduces changes not reflective of genuine differences between the samples. Derivatization schemes typically react protein

digests using isotopically-encoded variants of the same reagent. Differences in the isotopic purity of the labels can be corrected during data processing. However, variations in label incorporation can be nearly impossible to correct, since it is difficult to accurately estimate the conversion efficiency. While complete label incorporation is ideal, incomplete labeling can be tolerated so long as consistent reaction performance is achieved.

In this chapter, we describe our efforts to develop an inexpensive isotope labeling method for quantitative proteome analysis based on the optimization and automation of a previously reported isotope labeling chemistry. We investigated the robustness of this labeling method to handle proteomic samples containing a variety of surfactants and buffers. Using the 2-MEGA protocol[14] (dimethylation after guanidination), an automated differential isotopic labeling method utilizing a commercial liquid handler is described to minimize variability from sample handling during the labeling reaction for high throughput applications. The 2-MEGA protocol produces peptides with a fixed mass shift when used for labeling experiments and has been previously shown to increase the percentage of lysine containing peptides observed.[15] Furthermore, the comparatively low cost of the isotopically labeled reagent allows this automated method to be used for processing multiple samples. In this chapter, the reaction conditions were optimized for labeling of simple protein mixtures and complex tryptic digests of *E. coli*. Both front-end sample preparation methods and post-labeling workup are discussed. Potential side reactions, functional sample concentration ranges, and method limitations are also considered.

2.2 Experimental

2.2.1 Chemicals and Reagents

LC-MS grade water, methanol, acetonitrile, and ProteaseMAX™ were obtained from Fisher Scientific (Edmonton, AB). *O*-methylisourea hemisulfate, formaldehyde, sodium cyanoborohydride, sodium triacetoxyborohydride, pyridine-borane complex in tetrahydrofuran (THF), borane-THF complex, 2-picoline borane, urea, CellyticTM™ M cell lysis buffer, LC-MS grade formic acid and trifluoroacetic acid were obtained from Sigma Aldrich (Oakville, ON). Rapigest™ was purchased from Waters (Milford, MA). Anionic acid labile surfactant (AALS) was obtained from Canadian Life Sciences (Peterborough, ON). Total Protein Extraction Kit was purchased from Biochain Institute (Hayward, CA).

2.2.2 Protein Sample Preparation

E. coli K12 digest was prepared by culturing cells until OD600, lysing cells using an Emulsiflex homogenizer, and precipitating the proteins using acetone (1:5, v/v) at -80 ℃ overnight. The proteins were re-solubilized in 0.1% sodium dodecyl sulphate (SDS) and protein concentration was determined by BCA assay. Samples were reduced with dithiothreitol, alkylated with iodoacetamide and digested using a 50:1 ratio (protein:enzyme) of trypsin from swine in 50 mM NH_4HCO_3. After digestion, samples were acidified with 10% formic acid to pH 2 and SDS was removed from the peptide digest by strong cation exchange chromatography with a polysulfoethyl A column (2.1 mm i.d. x 250 mm with particle size of 5 μm

diameter and 300 Å pores). Desalting and peptide quantification were performed using an LC-UV method as previously described.[16] Digests were dried down in a vacuum centrifuge and reconstituted in the appropriate buffers for labeling optimization (200 mM NH_4HCO_3 or 200 mM KH_2PO_4).

2.2.3 Labeling Optimization

The manual 2-MEGA labeling method was previously described [80] and was used as the basis for labeling optimization to facilitate automation. Figure 2.1A shows the overall reaction scheme with the desired reaction product. Figure 2.2 shows the apparatus used in the automated 2-MEGA labeling. A Gilson 215 liquid handler with standard racks was used. The only modification was a homebuilt aluminum heating block with a thermocouple for temperature control with a temperature controller (Barnart Scientific) (see Inset in Figure 2.2). Initial experiments used tryptic horse myoglobin digest and *E. coli* digest was used for validation of general tryptic mixtures. In brief, peptide solutions (0.5 µg/µL) were adjusted to pH 11 using 2 M NaOH. *O*-methylisourea hemisulfate solution (~3M) was prepared in a 1:1 mixture (v/v) of 2 M NaOH and 1 M Na_2CO_3 (pH 12). The guanidination reaction was allowed to proceed before adjustment to pH 7 using 6 M HCl and further adjustment to pH 5 using 1 M acetate buffer. Formaldehyde (4% in H_2O, (w/w)) was added and followed by subsequent addition of 2-picoline borane (1 M in methanol). The dimethylation reaction was allowed to complete before adjusting to pH <2 using 10% TFA. Samples were desalted and quantified by LC-UV prior to mass spectrometric analysis.[16]

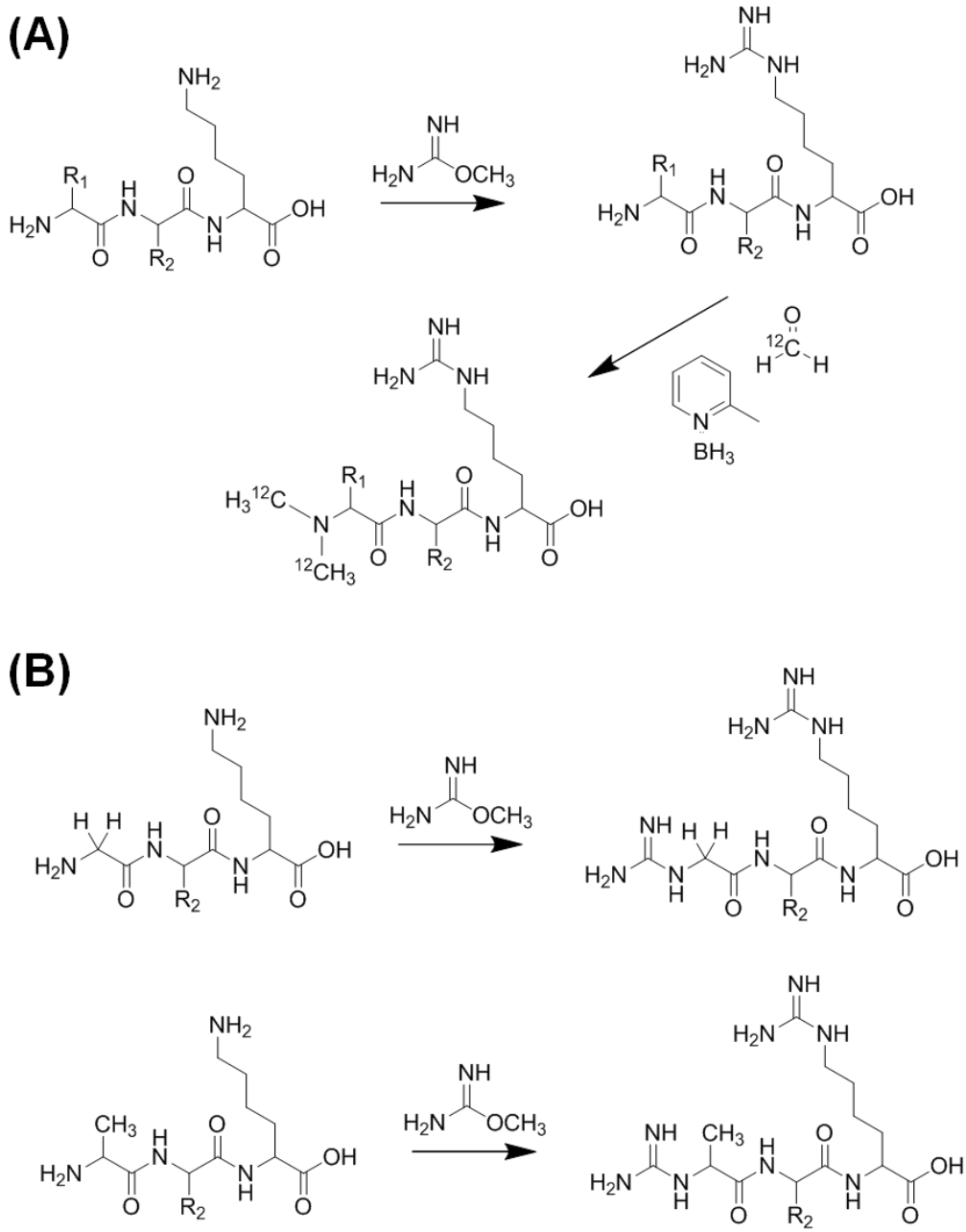


Figure 2.1 (A) 2-MEGA reaction scheme. For method optimization, only $^{12}CH_2O$ was used for testing. For quantitative shotgun proteome analysis, both $^{13}CD_2O$ and $^{12}CH_2O$ can be used for differential isotope labeling of two comparative samples. (B) Structures of major products from side reactions observed in 2-MEGA if the reaction conditions are not fully optimized.

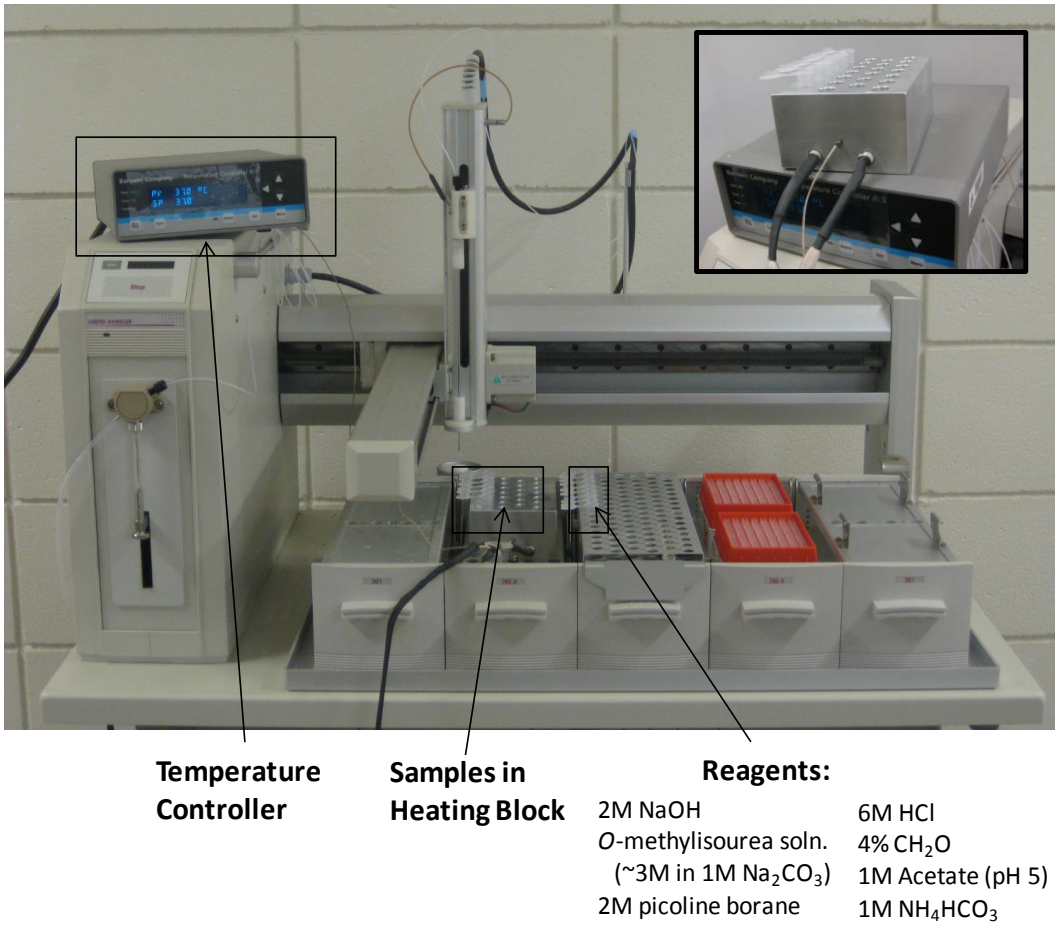


Figure 2.2 Reagents and a liquid handler for automated 2-MEGA labeling. A commercially available Gilson 215 liquid handler was used for the labeling work. The only modification was a homemade aluminum heating block used to heat the samples (see inset).

2.2.4 Mass Spectrometry and Data Analysis

Samples were separated on a 300 µm i.d. x 150 mm Discovery C_{18} column (Waters, Milford, MA) using a Waters nanoAcquity LC followed by analysis on a Waters ESI-QTOF Premier mass spectrometer. Solvent A used was 0.1% formic acid in water, and Solvent B was 0.1% formic acid in ACN. 5 µL of peptides were separated at 35 ℃ by a 120 min gradient (2-6% Solvent B for 2 min, 6-25% Solvent B for 95 min, 30-50% Solvent B for 10 min, 50-90% Solvent B for 10 min, 90-5% Solvent B for 5 min) after column equilibration at 2% Solvent B for 20 min and electrosprayed into the mass spectrometer fitted with a nanoLockSpray source at a flow rate of 300 nL/min. Peak lists were processed by ProteinLynx and searched using MASCOT (enzyme: trypsin; missed cleavages: 2; fixed modifications: Carbamidomethylation (C); MS tolerance: 30 ppm; MS/MS tolerance: 0.2 Da). Variable modifications for searches included both expected modifications from labeling (Guanidinyl (K, $+CN_2H_2$); Dimethylation (N-term; $+C_2H_4$)) as well as known side reaction products (Guanidinyl_NTerm (N-term, $+CN_2H_2$); Dimethylation_K (K; $+C_2H_4$)). A modified instrument-type setting using standard ESI-QTOF fragmentations further allowing a-ions was used. A database containing only *E. coli* K12 proteins was used for database searching (4337 sequences). Calculations to determine the extent and efficiency of labeling were taken as the number of correct identifications divided by the total number of identifications. Incorrect modifications were classified as peptides that had unlabeled groups, N-terminal guanidination, or dimethylation at lysine.

2.3 Results and Dscussion

2.3.1 Automated Liquid Handling

To allow for parallel processing of samples with minimal analyst intervention, a commercially available liquid handler was used to automate the liquid handling processes involved in 2-MEGA labeling (Figure 2.2). Dispensing on the liquid handler was performed with a single dispensing head that exchanged pipette tips between solution additions; it was anticipated that the described method should be applicable for multi-head liquid handling systems. The only non-standard modification used was a homemade thermocouple-controlled aluminum heating block used to accommodate standard 2-mL microcentrifuge vials and capable of heating to 95 ℃ for labeling optimization (see Figure 2.2, inset). However, it was found that heating to 37 ℃ was sufficient and that sample mixing and agitation during the reaction were not required. The minimal equipment setup was selected to allow the method to be easily adopted to address the specific demands of alternative dispensing configurations.

While many commercially available isotope labeling kits use expensive reagents, the 2-MEGA labeling method uses inexpensive isotope reagents that are suitable for processing multiple samples. 2-MEGA is a two-step labeling procedure using successive selective covalent modifications of the side chain amine of lysine, followed by reaction at the primary amine at the N-termini of peptides (see Figure 2.1A). By limiting introduction of the isotopically coded groups exclusively to free peptide N-termini (i.e., $^{13}CD_2O$ vs. CH_2O), a fixed mass shift is observed for all correctly modified peptides.[14,17] Conversion of the lysine residues to

homoarginine increases the relative proportion of lysine containing peptides identified from LC-MS experiments,[15] likely by increasing the ESI response of lysine containing peptides.[18] While not demonstrated in this paper, this fixed mass shift is independent of the number of lysines contained in the peptide, which may be useful when an enzyme without lysine-based specificity is used (e.g., Asp-N). Since the differentially labeled peak pair has a fixed mass shift, commercially available data processing software can be used for peak picking and peak ratio calculation. However, as the side chain amine and peptide N-termini are similar, careful and complete modification of the side chain amines is required before proceeding to the second reaction. We have optimized the labeling method by considering reagent concentrations, pH, temperature and reaction time, as described below.

2.3.2 Guanidination

In 2-MEGA, the first reaction is the conversion of the primary amine side chain of lysine into homoarginine, while leaving the N-terminus unreacted.[19,20] Although the N-terminus and lysine chain amine are both primary amines, differences in reactivity are observed due to differences in the local steric environment and the presence of the amide group beta to the N-terminus of a peptide. The difference in their chemical behavior is evident from differences in their pKa values: peptide N-termini have a pKa around 8, whereas the lysine side chain is around 10.5. Since reaction of the lysine side chain requires that the amine is deprotonated, pH values around 10, 10.5, 11, 11.5, and 12 for the guanidination reaction were considered with pH 11.5 found to be optimal. For pH values of higher than 12, an

increased proportion of guanidination at peptide N-termini was observed. At pH of lower than 10.5, the reaction was sluggish and would not reach completion even after two hours at temperatures as high as 65 ℃.

Two different buffer systems were considered for the digestion and initial reaction step. Ammonium bicarbonate (pKa ~9) was adjusted with NaOH to the bicarbonate/carbonate buffer pair (pKa ~12) for guanidination. Similarly, sodium phosphate was also considered since the $H_2PO_4^-/HPO_4^{2-}$ pair buffers around the desired range for digestion (pKa ~8) and the HPO_4^{2-}/PO_4^{3-} pair can be used for guanidination (pKa ~11). Guanidination in the presence of phosphate often gave incomplete yields, even when increased concentrations of *O*-methylisourea or elevated temperatures between 45 to 75 ℃ were used. The exact rationale for why this occurred is unclear, but may be due to the interaction of the phosphate groups with the primary amines or charged *O*-methylisourea cation in solution. In order for the guanidination reaction to reach completion, a significant molar excess of *O*-methylisourea hemisulfate is required. Given the sample solution buffer concentration (200 mM) and the nearly equal volume of *O*-methylisourea hemisulfate solution (~3 M) added, the overall buffer pH is controlled by the reagent solution. When unacceptable results during the guanidination reaction were observed, adjusting the reagent solution pH using 2 M NaOH was found to be the most effective method for controlling the reaction outcome.

Even under the optimized conditions, approximately 1-2% of the peptides observed will have guanidination at the peptide N-terminus, primarily on glycine and alanine N-terminal peptides (Figure 2.1B). An example is shown in Figure

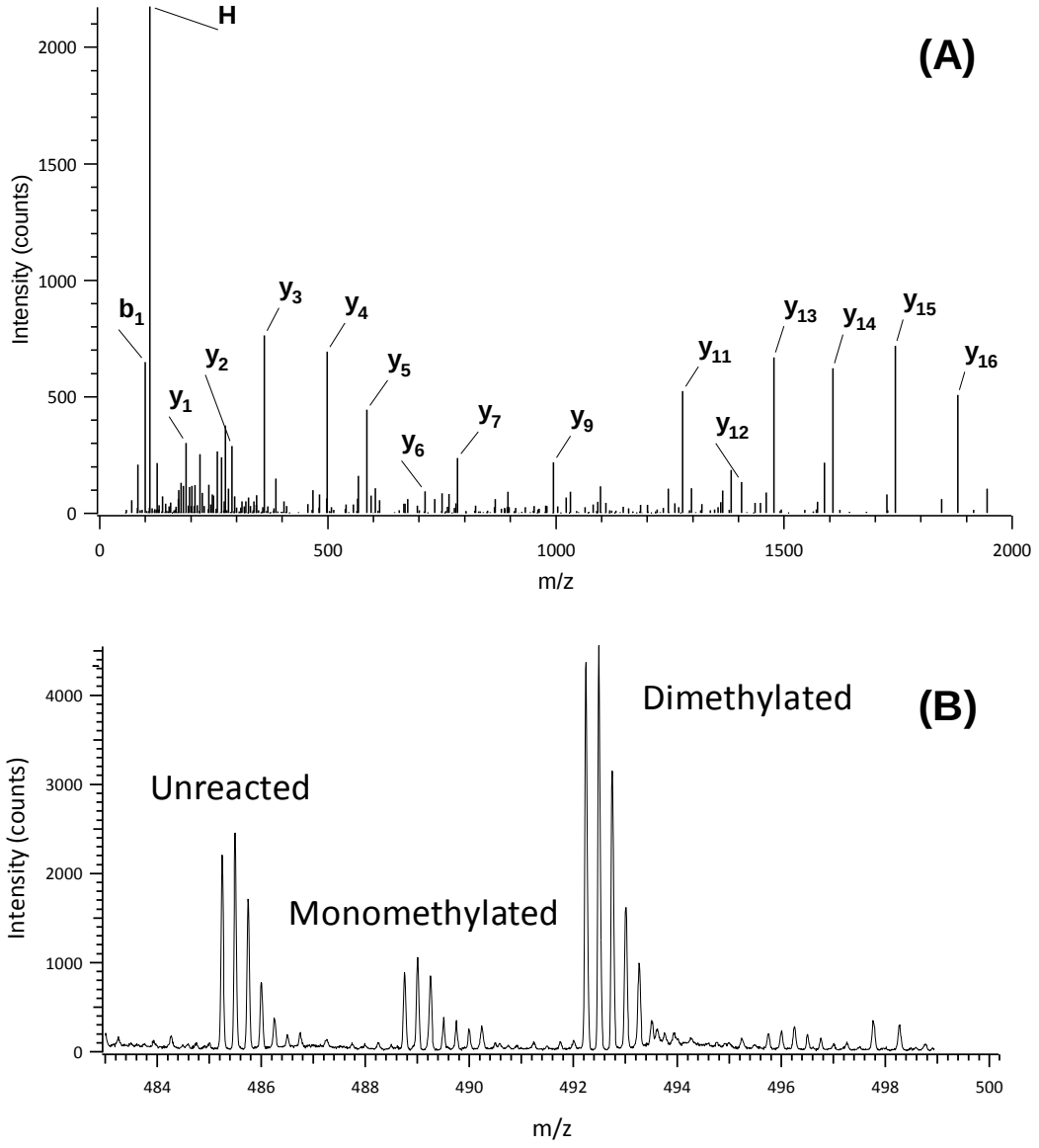


Figure 2.3 (A) MS/MS spectrum of GHHEAELKPLAQSHATK. Guanidination on the N-terminus is confirmed from the b_1 ion as well as the y-ion ladder. (B) MS spectrum showing incomplete dimethylation. When insufficient reagents are used for dimethylation, the monomethylated peak is usually weaker than the unreacted peak or the dimethylated peak.

2.3A for GHHEAELKPLAQSHATK from myoglobin. The precursor ion mass of the peptide suggests three guanidiation groups attached to the peptide. The MS/MS spectrum provides assignment of one of modifications to the N-terminus, due to the location of the b_1 ion. The nearly completely y-ion ladder also suggests modification at the N-terminus. The strong peak in the low mass region is the immonium ion of histidine. It was found that excessive reagents (threefold increase of concentration) lead to a slight increase of guanidination of the N-terminus (~5% of total identified peptides). Although most shotgun proteomics protocols reduce disulfide bonds and alkylate using thiol active reagents, such as dithiothreitol or N-ethylmaleimide, prior to digestion, they will become methylated quantitatively ($+CH_2$, +14 Da) if free cysteines are present in the sample.[21] We did not find any other adverse effects. Despite the high concentration, *O*-methylisourea is hydrolyzed into methanol and urea after standing in water.

2.3.3 Dimethylation

The dimethylation reaction was found to be robust and only required limited modification from previously reported protocols.[22-26] Side products were not found and insufficient reagent addition resulted primarily in properly labeled and fully unlabeled peptides. When insufficient reagents were used, it was noted that the intensity of monomethylated peptides was often less than unlabeled and dimethylated peptides; an example is shown in Figure 2.3B. In automating 2-MEGA, the primary objective was substitution of the toxic sodium cyanoborohydride used for the reduction of the imine formed from the

condensation of formaldehyde with the free N-termini of peptides and remaining unreacted lysine side chains. Sodium cyanoborohydride is particularly useful for the reductive methylation of peptides due to its reasonably strong reducing potential and high stability under aqueous conditions.[27] The commercial liquid handler apparatus used was open to the laboratory atmosphere and slow outgassing of hydrogen cyanide, even under basic conditions, remained a key safety consideration. Thus, alternative reducing agents were considered to overcome this issue.

The initial reducing agents tested (triacetoxyborohyride,[28] borane-THF, and borane-pyridine complex[29]) are water sensitive, but have been previously used for reductive aminations. Since there is no report on the hydrolysis half-lives of these compounds, they were evaluated for their feasibility in aqueous solutions. Even with several molar equivalents of reducing agent, the borane-THF and borane-pyridine complex resulted in non-quantitative conversion (~90%) and sodium triacetoxyborohydride had nearly no conversion (<10%). 2-Picoline borane complex[30] was a useful alternative since it is relatively air stable, non-toxic, and can be prepared in methanol at usable concentrations (>2 M). The main drawback encountered in using 2-picoline borane over sodium cyanoborohydride is its limited solubility in aqueous solutions. Upon standing in aqueous solution for a short period of time (~30 min), a precipitate is formed which needs to be removed by centrifugation before the sample is subjected to additional downstream processing. Since the precipitation is complete in the time required to finish the reaction, the samples can be centrifuged once without concern that additional precipitate will form. Since the dimethylation reaction proceeded quantitatively

under a wide range of pH values (4-8), pH optimization was not required. Various buffers, such as ammonium bicarbonate (pH 8), triethylammonium bicarbonate (pH 8) and sodium acetate/acetate (pH 5) were all suitable for the dimethylation reaction.

2.3.4 Effect of Sample Mass

Tryptic digests between 2 to 200 µg per sample vial at a concentration of 0.5 µg/µL were successfully labeled with >95% complete labeling, with the optimal reaction efficiency around 20 to 150 µg (>97%). This range is sufficient for most shotgun proteomics applications. The liquid handler was imprecise for the delivery of volumes less than 2 µL, which ultimately limited the lower limit that could be reached. Concentration ranges between 0.1-2 µg/µL were also tested and also give similar performance characteristics. For each sample, the mass of *O*-methylisourea and formaldehyde/2-picoline borane complex added was adjusted for the total peptide mass.

Since the stepwise reaction scheme necessitates that the conversion of lysine groups is complete before addition of the reductive methylation reagents, a reasonably close estimate of the guanidination reagent mass is required. Initially, an LC-UV peptide quantification method was used to quantify the peptide mass in a digest to determine the optimal reagent mass for guanidination. Ideally, it would be preferable to go directly from digestion to the labeling step without an intermediate quantification step. Assuming that a protein concentration was determined prior to trypsin digestion, such as by the Bradford or BCA assay, we investigated how deviations from the ideal reagent mass would affect labeling

Table 2.1 Number of labeled peptides and the rate of correct 2-MEGA labeling in the duplicate analysis of *E. coli* proteomic digests prepared with or without CelLytic M™ and TM buffers.

	CelLytic M™ buffer		CelLytic M™ buffer with acetone precipitation		TM buffer		TM buffer with acetone precipitation	
	Run #1	Run #2	Run #1	Run #2	Run #1	Run #2	Run #1	Run #2
Number of correctly labeled peptides	975	966	1029	1010	1144	1143	1109	1107
Total number of labeled peptides	1037	1024	1094	1067	1204	1204	1180	1177
Rate of correct labeling	94.0%	94.3%	94.1%	94.7%	95.0%	94.9%	94.0%	94.1%

efficiency. For a 50 μg sample, reagent masses corresponding to 20 to 500% of the ideal were tested. To maintain over 95% labeling efficiency, reagent ranges from 25% to 200% of the ideal were required. Thus, given the relative accuracy of these protein quantification methods, it should be feasible to go directly from

estimation of the protein concentration to the finished labeled peptide products with over 95% correct labeling.

2.3.5 Compatibility with Cell Lysis Buffers

An ideal labeling method should be applicable to a wide range of protein samples that contain a variety of surfactants and other chemicals. To test the optimized labeling chemistry with a variety of front end sample preparation methods, different commonly used cell lysis and protein solubilization methods were considered. Since surfactants and buffer components are not always removed prior to sample workup, they were evaluated for their potential to interfere with the automated 2-MEGA labeling protocol. It should be noted that labeling strategies targeting the amine functionalities of proteins or peptides are among, if not, the most common for proteomics applications.[10,13,31] Thus the findings here should be generally applicable to other similar methods such as the commercial iTRAQ™ and TMT™ reagents.

CelLytic™ M and TM buffers are two commonly used cell lysis buffers, although the identities of the components present in either buffer are unknown. To investigate potential interferences, *E. coli* digests were lysed with CelLytic™ M or TM buffer, with or without acetone protein precipitation to remove buffer components, and labeled with the automated method to gauge the labeling efficiency. Table 2.1 summarizes the number of correctly labeled peptides, the total number of peptides detected, and percentage of correct 2-MEGA labeling under each condition. More than 1000 peptides were identified with each lysis buffer, regardless of whether or not the cell lysis buffer was removed, with a

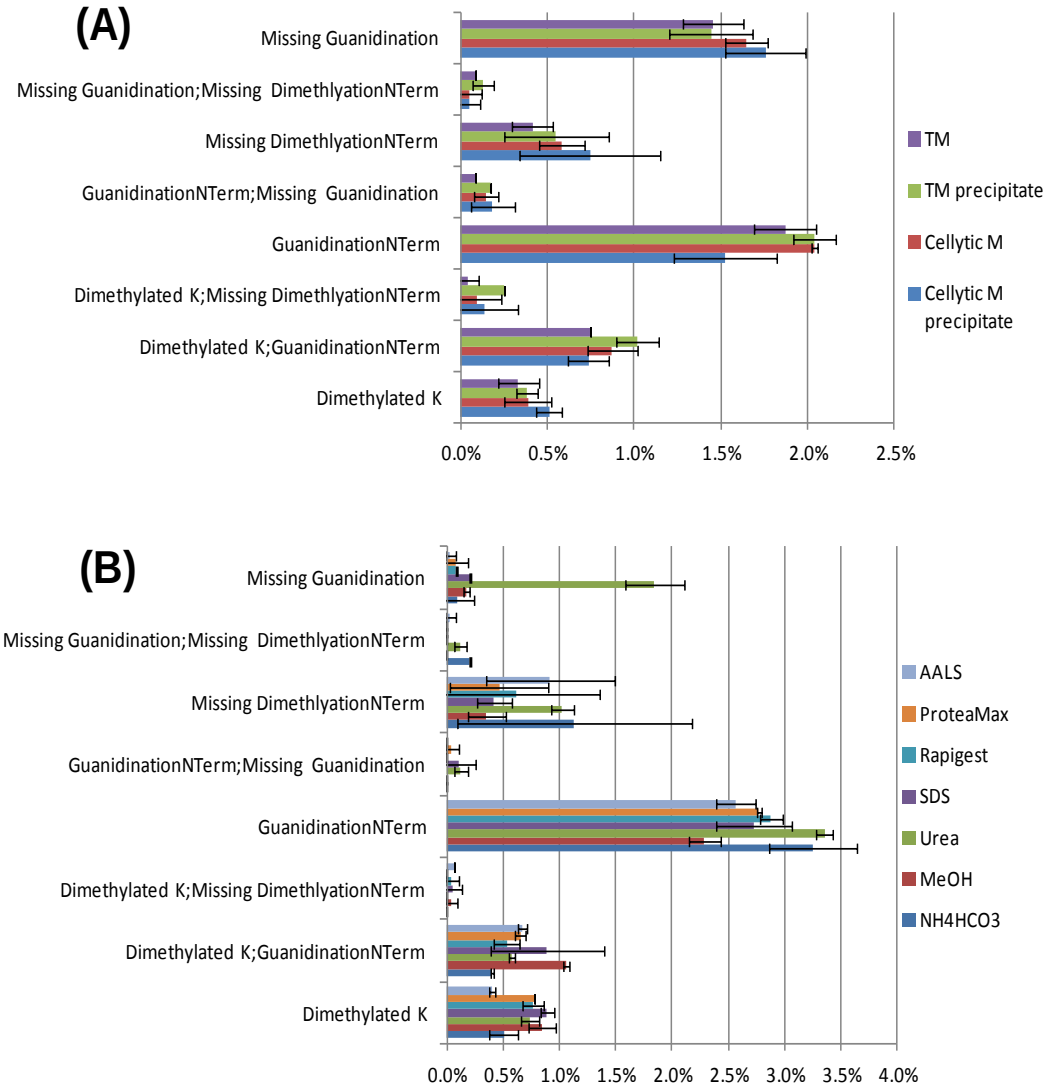


Figure 2.4 Distribution of frequently occurred side reactions when 2-MEGA labeling was performed in samples prepared by using (A) different lysis buffers and (B) different protein solubilization reagents.

labeling efficiency over 94%. All four labeling conditions examined had similar correct labeling rates and side reaction profiles. Figure 2.4A shows the percentage of frequently occurred side reactions under each reaction condition. The most frequent side reaction was guanidination on the peptide N-termini (1.5 to 2.0% in all the identified peptides), followed by missing guanidination on the lysine side chain (K) (~1.5% of total peptides). Other side reactions, such as missing dimethylation on N-termini, were less than 1% of the total identifications. These results indicate that the automated 2-MEGA labeling reaction is compatible with the CelLytic^{TM} M and TM buffers.

2.3.6 Compatibility with Protein Solubilization Reagents

Various buffers, organic solutions, and surfactants can be used to aid protein solubilization.[32-35] Notably, recently commercialized cleavable detergents assist with protein solubilization and are compatible with downstream mass spectrometric analysis. While the structures of some cleavable surfactants are available in the research literature (Rapigest^{TM} [36,37]) or the associated commercial literature (ProteaseMAX^{TM}, sodium 3-((1-(furan-2-yl)undecyloxy)carbonylamino)propane-1-sulfonate), others structures are not disclosed (AALS). Cleavable surfactants often use acetals [134] and carbamates as the linker functionality between the hydrophobic and hydrophilic portions of the detergent. The cleavage protocol described for the reported acetal-containing surfactants is typically treatment with acid (pH < 2), which is similar to the suggested treatment for cleavage of AALS. Since the cleavage products may yield functional groups that interfere with the reaction by potentially consuming

Table 2.2 Number of labeled peptides and the rate of correct 2-MEGA labeling in the duplicate analysis of *E. coli* proteomic digests prepared using different protein solubilization reagents.

Reagent	Number of correctly labeled peptides	Total number of labeled peptides	Rate of correctly labeled peptides
NH_4HCO_3	960	1008	95.2%
	903	965	93.6%
MeOH	1003	1052	95.3%
	1234	1297	95.1%
Urea	1089	1183	92.1%
	1167	1263	92.4%
SDS	915	965	94.8%
	893	944	94.6%
Rapigest	1090	1140	95.6%
	1250	1322	94.6%
ProteaseMax	1232	1290	95.5%
	1103	1162	94.9%
AALS	1492	1561	95.6%
	1363	1434	95.1%

reagents, such as the amine group produced from the hydrolysis of the carbamate linker in ProteaseMAX™, they were evaluated for their potential effects.

The compatibility of seven widely used protein solubilization reagents with the 2-MEGA labeling were studied. *E. coli* cells were lysed with TM buffer and proteins were precipitated with acetone. The protein pellets were then dissolved in 100 mM NH_4HCO_3 (control), 60% MeOH, 6 M Urea, 0.6% SDS, 0.6% Rapigest, 0.15% ProteaseMax or 0.6% AALS; the concentration used in each case is representative of those commonly used for dissolving proteins. The samples were then labeled with the 2-MEGA protocol and analyzed by LC/MS. Table 2.2 shows the number of peptides identified and the percentage of desired 2-MEGA labeling for each reagent. AALS had the greatest number of peptide identifications (~1500), while SDS gave the lowest number (~950). It is noted that strong cation exchange was used to remove SDS prior to LC/MS analysis, which likely caused loss of some peptide content.[38] The other reagents (MeOH, Urea, Rapigest and ProteaseMax) gave broadly similar peptide identification numbers and were all slightly higher than the control group, NH_4HCO_3.

All of the buffers, aside from urea, have similar labeling rates (94 to 95%), indicating general compatibility with 2-MEGA. The labeling efficiency of urea was slightly lower at 92%. To explain the discrepancy, the side reactions found were categorized. Figure 2.4B illustrates that guanidination on peptide N-termini remains the most common side reaction (over 2.5% of the total peptides detected). Due to the potential of carbamylation from urea (i.e., N-terminal amine or side chain of lysine[39-41]), the MS/MS data was searched with carbamyl (N-terminal)

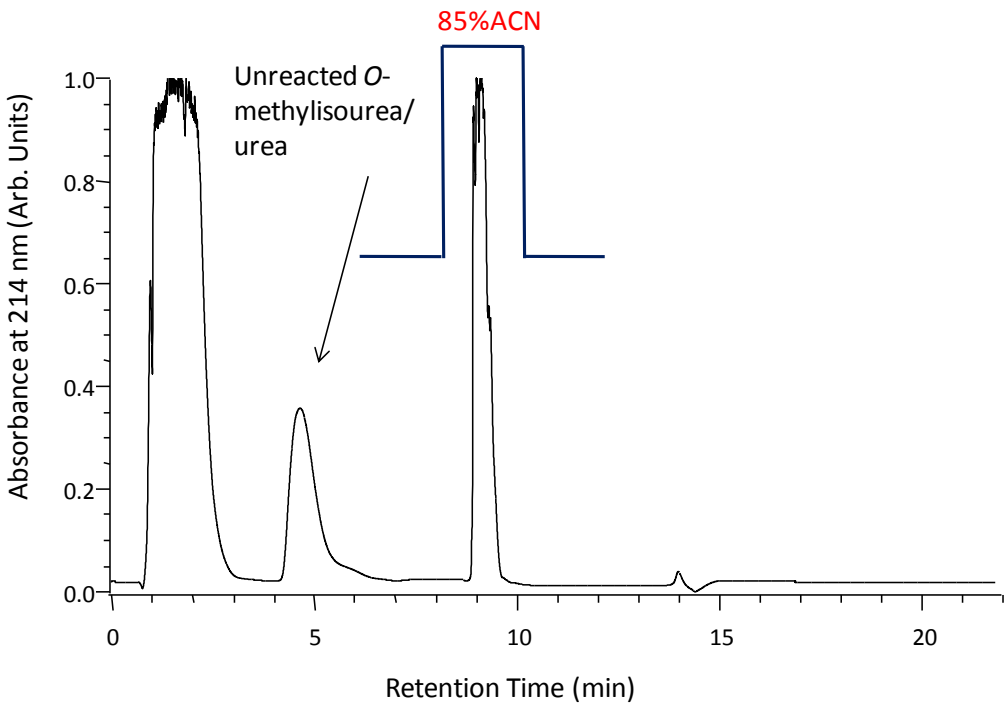


Figure 2.5 Desalting UV chromatogram of a 2-MEGA labeled sample. The previously published desalting and peptide quantification method[115] elutes peptides around t_R = 5 min. However, the presence of *O*-methylisourea leads to a broad, weakly retained peak at t_R = 4.8 min, which requires a longer hold time of the equilibration mobile phase at the beginning, leading to peptide elution at t_R = 9 min.

and carbamyl (K) as variable modifications to examine the frequency of carbamylation. For the urea samples, carbamylation was detected for over 2% of the total identified peptides, but was less than 0.5% for the other sample solutions. Thus, carbamylation appears to be the main cause of lowering the rate of correct 2-MEGA labeling in the urea samples. The use of basic conditions at higher

temperatures during guanidination is known to elevate the formation of cyanate, which reacts with amines to form carbamyl groups. After carbamylation, the amine is effectively blocked, which causes a reduction in the labeling efficiency.

For all sample preparations, the peptide samples were acidified with trifluoroacetic acid, desalted, and quantified using an RPLC-UV prior to LC-MS analysis. Desalting was necessary, due to the relatively high concentration of salts and buffers used in the labeling scheme. The typical hold with the equilibration mobile phase was extended from five to eight minutes in order to remove the weakly retained salts, which appeared as a strong tailing peak in the chromatogram, as shown in Figure 2.5. The first major peak at t_R = ~2 min in Figure 2.5 is from salt, buffer components, and other unretained compounds in the sample. The second peak, which is typically tailing, at t_R = ~5 min is from unreacted *O*-methylisourea and urea, which is the hydrolysis product of *O*-methylisourea. The final major peak at t_R = 9 min is the desalted peptide sample and was collected by a fraction collector, dried down in a vacuum centrifuge, and analyzed by LC-MS.

2.4 Conclusions

We have developed and optimized an automated dimethylation after guanidination (2-MEGA) labeling process for quantitative proteomics. A commercially available liquid handler was used for sample preparation and compatibility with various front end protein preparation methods was

demonstrated. Under several buffer and surfactant conditions, over 94% of tryptic peptides were correctly labeled; in the case of urea-containing buffers, 92% correct labeling was obtained. The studies on the reproducibility and accuracy of the 2-MEGA method will be reported in a future publication. Future work will include applications of the described labeling method for the preparation and analysis of protein samples for quantitative proteomics.

2.5 Literature Cited

(1) Lemeer, S.; Jopling, C.; Gouw, J.; Mohammed, S.; Heck, A. J. R.; Slijper, M.; den Hertog, J. *Molecular and Cellular Proteomics* **2008**, *7*, 2176.
(2) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. *Science* **2009**, *325*, 834.
(3) Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; Verdin, E.; Kahn, C. R.; MacCoss, M. J.; Gibson, B. W. *Molecular & Cellular Proteomics* **2012**, *11*, 202.
(4) Shakey, Q.; Bates, B.; Wu, J. *Analytical Chemistry* **2010**, *82*, 7722.
(5) Jayapal, K. P.; Sui, S.; Philp, R. J.; Kok, Y.-J.; Yap, M. G. S.; Griffin, T. J.; Hu, W.-S. *Journal of Proteome Research* **2010**, *9*, 2087.
(6) Park, S.-S.; Wu, W. W.; Zhou, Y.; Shen, R.-F.; Martin, B.; Maudsley, S. *Journal of Proteomics*, *75*, 3720.
(7) Gehlenborg, N.; O'Donoghue, S. I.; Baliga, N. S.; Goesmann, A.; Hibbs, M. A.; Kitano, H.; Kohlbacher, O.; Neuweger, H.; Schneider, R.; Tenenbaum, D.; Gavin, A.-C. *Nat Methods* **2010**, *7*, S56.
(8) Gstaiger, M.; Aebersold, R. *Nature Reviews Genetics* **2009**, *10*, 617.
(9) Wu, Q.; Yuan, H. M.; Zhang, L. H.; Zhang, Y. K. *Anal. Chim. Acta* **2012**, *731*, 1.
(10) Rodriguez-Suarez, E.; Whetton, A. D. *Mass Spectrom. Rev.* **2013**, *32*, 1.
(11) Zhang, J.; Zhang, L.; Wang, H.-Y.; Guo, Y.-L. *Chin. J. Org. Chem.* **2011**, *31*, 2043.
(12) Dicker, L.; Lin, X. H.; Ivanov, A. R. *Molecular and Cellular Proteomics* **2010**, *9*, 2704.
(13) Gevaert, K.; Impens, F.; Ghesquiere, B.; Van Damme, P.; Lambrechts, A.; Vandekerckhove, J. *Proteomics* **2008**, *8*, 4873.
(14) Ji, C.; Guo, N.; Li, L. *Journal of Proteome Research* **2005**, *4*, 2099.
(15) Ji, C.; Lo, A.; Marcus, S.; Li, L. *Journal of Proteome Research* **2006**, *5*, 2567.
(16) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Analytical Chemistry* **2009**, *81*, 1049.

(17) Yang, S. J.; Nie, A. Y.; Zhang, L.; Yan, G. Q.; Yao, J.; Xie, L. Q.; Lu, H. J.; Yang, P. Y. *J. Proteomics* **2012**, *75*, 5797.
(18) Brancia, F. L.; Openshaw, M. E.; Kumashiro, S. *Rapid Communications in Mass Spectrometry* **2002**, *16*, 2255.
(19) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2070.
(20) Beardsley, R. L.; Reilly, J. P. *Analytical Chemistry* **2002**, *74*, 1884.
(21) Banks, T. E.; Shafer, J. A. *Biochemistry* **1970**, *9*, 3343.
(22) Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. *Analytical Chemistry* **2003**, *75*, 6843.
(23) Ji, C.; Li, L. *Journal of Proteome Research* **2005**, *4*, 734.
(24) Melanson, J. E.; Avery, S. L.; Pinto, D. M. *Proteomics* **2006**, *6*, 4466.
(25) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. *Nature Protocols* **2009**, *4*, 484.
(26) Fu, Q.; Li, L. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 859.
(27) Borch, R. F.; Bernstein, M. D.; Durst, H. D. *Journal of the American Chemical Society* **1971**, *93*, 2897.
(28) Abdel-Magid, A. F.; Carson, K. G.; Harris, B. D.; Maryanoff, C. A.; Shah, R. D. *Journal of Organic Chemistry* **1996**, *61*, 3849.
(29) Bomann, M. D.; Guch, I. C.; DiMare, M. *Journal of Organic Chemistry* **1995**, *60*, 5995.
(30) Sato, S.; Sakamoto, T.; Miyazawa, E.; Kikugawa, Y. *Tetrahedron* **2004**, *60*, 7899.
(31) Julka, S.; Regnier, F. *Journal of Proteome Research* **2004**, *3*, 350.
(32) Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D. *J Proteome Res* **2002**, *1*, 351.
(33) Goshe, M. B.; Blonder, J.; Smith, R. D. *J Proteome Res* **2003**, *2*, 153.
(34) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat Biotechnol* **2001**, *19*, 946.
(35) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat Biotechnol* **2001**, *19*, 242.
(36) Yu, Y.-Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C. *Analytical Chemistry* **2003**, *75*, 6023.
(37) Yu, Y.-Q.; Gilar, M.; Gebler, J. C. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 711.
(38) Wu, F.; Sun, D. F.; Wang, N.; Gong, Y.; Li, L. *Anal. Chim. Acta* **2011**, *698*, 36.
(39) Angel, P. M.; Orlando, R. *Rapid Commun Mass Spectrom* **2007**, *21*, 1623.
(40) McCarthy, J.; Hopwood, F.; Oxley, D.; Laver, M.; Castagna, A.; Righetti, P. G.; Williams, K.; Herbert, B. *J Proteome Res* **2003**, *2*, 239.
(41) Ye, X.; Li, L. *Anal Chem* **2012**, *84*, 6181.

Chapter 3

Quantification of Total Peptide Amount by an Optimized LC-UV Method for Assessing Sample Integrity during Proteome Sample Preparation*

3.1 Introduction

Shotgun proteome analysis based on liquid chromatography mass spectrometry (LC-MS) has become a powerful tool for proteome profiling of cells, tissue and biofluids. In proteomics applications, hundreds of micrograms of proteins are often used as the starting material. However, the sample amount can be very limited in some important areas of applications, such as in dealing with single cells,[1-5] circulating cancerous cells captured from a blood sample of a patient with early sign of tumor in a specific organ,[6-8] stem cells isolated from a large population of other types of cells[9,10] and primary cells procured from tissues.[11,12] To ensure maximum proteome coverage for these samples, both the sample preparation procedure and MS analysis method need to be optimized. Prior to LC-MS analysis, the shotgun workflow requires protein extraction from cells, followed by protein digestion. This sample preparation process can potentially

---

*This work was done under the supervision of Dr. Nan Wang. A form of this chapter has been submitted for publication as: N. Wang, Y.N. Tang, L. Chen, and L. Li, "Microbore Liquid Chromatography UV Detection for Quantification of Total Peptide Amount and Its Application for Assessing Sample Quality in Shotgun Proteome Analysis of Few Cells", Journal of Chromatography A, submitted. N. Wang, Y.N. Tang and L. Chen contributed equally to this work.

lose some proteins. For example, protein or peptide adsorption to containers may lose a fraction of the original sample. In working with a large quantity of samples, this sample loss may not affect the proteome coverage. However, for handling a limited amount of proteins, sample loss can reduce the proteome coverage significantly.

There are a variety of sample preparation methods that have been reported for handling a relatively large number of cells.[13-15] These methods and others can be evaluated and further developed for handling few cells with an objective of increasing the efficiencies of cell lysis, protein extraction and digestion while minimizing sample loss.[5,16-19] However, during the course of developing or applying a sample handling protocol for proteome analysis of few cells, it is desirable to use a simple method to quantify the total amount of peptides generated. The total amount can be used as a bench mark for method comparison. In addition, if the quantification method is non-destructive, the peptides can be collected and the amount can be used to assess the sample quality to determine whether the sample was properly prepared for MS analysis and, if so, what the optimal conditions of the LC-MS settings should be. For example, a shortened gradient time is favored in analyzing smaller amounts of peptides.[16]

Determination of the total protein amount extracted from the cells can be done using a commercially available kit, such as the Bradford or bicinchoninic acid (BCA) assay, although these are distractive methods.[20] However, knowing the protein amount does not allow us to evaluate the entire workflow of the shotgun method where the final product of analysis is the peptides, not proteins. On the

other hand, measuring the peptide amounts in nanograms or sub-microgram range is not trivial. In this chapter, we report a microbore column (1-mm diameter) LC-UV method with a step gradient elution to rapidly quantify the total peptide amount in proteomic digest samples while removing salts and other reagent impurities that may interfere with LC-MS analysis. This method was illustrated to be useful for quantifying peptides and assessing sample quality in the shotgun proteome analysis of 250, 500 and 1000 MCF-7 breast cancer cells.

3.2 Experimental

3.2.1 Chemicals and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless stated otherwise. Acetonitrile (ACN) and HPLC grade water were purchased from Fisher Scientific Canada (Edmonton, AB, Canada).

3.2.2 Cell Culture and Cell Sorting by Flow Cytometer

The MCF-7 breast cancer cells (ATCC® number: HTB-22™) were cultured in 15 cm diameter plates at 37 ℃ in Gibco® Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum. The cells were harvested by scraping from the plates into the PBS^{++} buffer (0.68 mM $CaCl_2$, 0.5 mM $MgCl_2$, 1.4 mM KH_2PO_4, 4.3 mM Na_2HPO_4, 2.7 mM KCl, and 137 mM NaCl) and centrifugation at 100 *g* for 8 min at 4 ℃. The harvested cells were then fluorescently stained by incubating with a Fluorescein Isothiocyanate (FITC)-conjugated mouse anti–human HEA (Human Epithelium Antigen) antibody

(Miltenyi Biotec number: 130-080-301) in a 1:100 (v:v) ratio on ice for 15 min. The stained MCF-7 cells were introduced into the flow cytometer for counting, according to the cell size and their fluorescence response. Then 250, 500, 1000, 2500 or 5000 MCF-7 cells were collected into 0.6 mL low retention microcentrifuge vials.

3.2.3 Cell Lysis and In-solution Digestion

The process used for cell lysis and in-solution digestion of few cells was similar to that reported previously.[16] Briefly, the cells in each vial were mixed with 5 to 10 µl Nonidet-P40 (NP40) lysis buffer (1%) and sonicated in ice-water ultrasonic bath for 5 min. The protein solutions were reduced with dithiothreitol (DTT) and alkylated with iodoacetamide (IAA). Acetone (pre-cooled to -80^{o}C) was added gradually (with intermittent vortexing) to the protein extract to a final concentration of 80% (v/v). The solution was then incubated at -80°C for 4 hr and centrifuged at 14 000 rpm for 10 min. The supernatant was decanted. The pellet was carefully washed once using cold acetone to ensure the efficient removal of NP40 detergent. After the residual acetone was evaporated, 50 mM ammonium bicarbonate was used to sufficiently redissolve the pellet in the vial. Trypsin digestion was then carried out in a final enzyme concentration of 8 ng/µL (5 to 20 µL) at 37°C for 4 hr.

3.2.4 Preparation of Calibration Standards

A four-protein mixture solution containing equal moles of myoglobin (16.7 kDa), cytochrome C (11.6 kDa), lysozyme (14.3 kDa) and ß-casein (23.6 kDa) was

prepared by dissolving intended amount of protein standards in 50 mM ammonium bicarbonate. After reduction with 20 mM DTT and alkylation with the same volume of 40 mM IAA, the protein mixture was then digested by trypsin at a final enzyme concentration of 8 ng/µl at 37 ^{o}C for 8 hr.

3.2.5 RPLC for Peptide Quantification and Desalting

The tryptic peptides were desalted and quantified on an Agilent 1100 HPLC system (Agilent, Mississauga, ON) with a 1 mm × 50 mm Polaris C18 A column with 3 µm particles and 300 Å pores (Varian, Palo Alto, CA, USA). The flow rate used was 100 µL/min. Peptides were eluded from columns at room temperature by a step gradient: flushing column with 97.5% mobile phase A (0.1% TFA in water) for 5 min and then 85% of mobile phase B (0.1% TFA in acetonitrile) for 5 min to completely elute the peptide fractions at room temperature, followed by 15 min re-equilibration with 97.5% mobile phase A. The UV absorbance of eluted peptides was detected at 214 nm.

3.2.6 MS Analysis

After desalting and quantification, the digests were analyzed using a quadrupole time-of-flight (QTOF) Premier mass spectrometer equipped with a nanoACQUITY Ultra Performance LC (UPLC) system (Waters, Milford, MA). Briefly, the desalted digests were concentrated on SpeedVac and reconstituted with 0.1% formic acid. Then 5 µL of the digest solution was injected onto a 75 µm × 100 mm Atlantis dC18 column. Solvent A consisted of 0.1% formic acid in water, and Solvent B consisted of 0.1% formic acid in ACN. Peptides were separated using their optimal lengths of solvent gradients ranging from 90 min to

150 min at 35 ℃ and electrosprayed into the mass spectrometer fitted with a nanoLockSpray source at a flow rate of 250 nL/min. One MS scan was acquired from m/z 350-1600 for 0.8 s, followed by 4 MS/MS scans from m/z 50-1900 for 0.8 s each. A mixture of leucine enkephalin and (Glu1)-fibrinopeptide B used as mass calibrants (i.e., lock-mass), was infused at a flow rate of 250 nL/min, and a 1 s MS scan was acquired every 1 min throughout the run.

3.2.7 Protein Database Search

Raw LC-ESI data were lock-mass corrected, de-isotoped, and converted to peak list files using ProteinLynx Global Server 2.2.5. Peptide sequences were identified via automated database searching of peak list files using the MASCOT search program (version 1.8). Database searching was restricted to *Homo sapiens* (human) in the SWISSPROT database with following search parameters: enzyme: trypsin; missed cleavages: 1; peptide tolerance: 30 ppm; MS/MS tolerance: 0.2 Da; peptide charge: 1+, 2+, and 3+; fixed modification: Carbamidomethyl (C); variable modifications: acetyl (Protein), oxidation (M), pyro-Glu (N-term Q) and pyro-Glu (N-term E). All the identified peptides with scores lower than the MASCOT threshold score for identity at a confidence level of 95% were then removed from the protein list, as well as the redundant peptides for different protein identities.

Because of the small data set generated from the proteome analysis of a few cells, accurate analysis of the false discovery rate (FDR) is difficult. The commonly used target-decoy search strategy is best suited for analyzing a large data set.[21] To ensure data quality, we have manually verified many of the matched MS/MS

spectra with peptide sequences.[16] Specifically, peptide matches with a matching score of less than 10 points above the MASCOT threshold score for identity were manually analyzed. The peptide match was considered as positive identification if the fragment ions contained more than five isotopically resolved y-, b-, or a-ions and the major fragment ion peaks with high intensity (i.e., peak intensity of >30% in a normalized spectrum). Most of the high intensity fragment ions (i.e., top 5) must also belong to y-, b-, or a-ions, not internal fragment ions. Peptide matches which failed to meet these criteria were removed from the protein lists. Typically, this manual verification process eliminated about 3% of the low score matches. A protein was considered to be identified even if a single peptide match was found, due to the high quality MS/MS spectra acquired in QTOF.

3.3 Results and Discussion

3.3.1 Calibration Curve of Microbore LC-UV

To mimic the diverse peptide composition seen in a proteomic digest while easily controlling the peptide amount, a digest of a four-protein mixture was used as the peptide standard.[22] A stock solution of the digest was linearly diluted and injected into the microbore LC-UV to establish a calibration curve for peptide quantification. A step gradient, as described in the experimental section, was used to rapidly elute peptides from the RPLC column to generate one integral peak (see Figure 3.1). The peak area could be related to the amount of peptides injected. With a step gradient, a system peak from the blank injection was usually observed

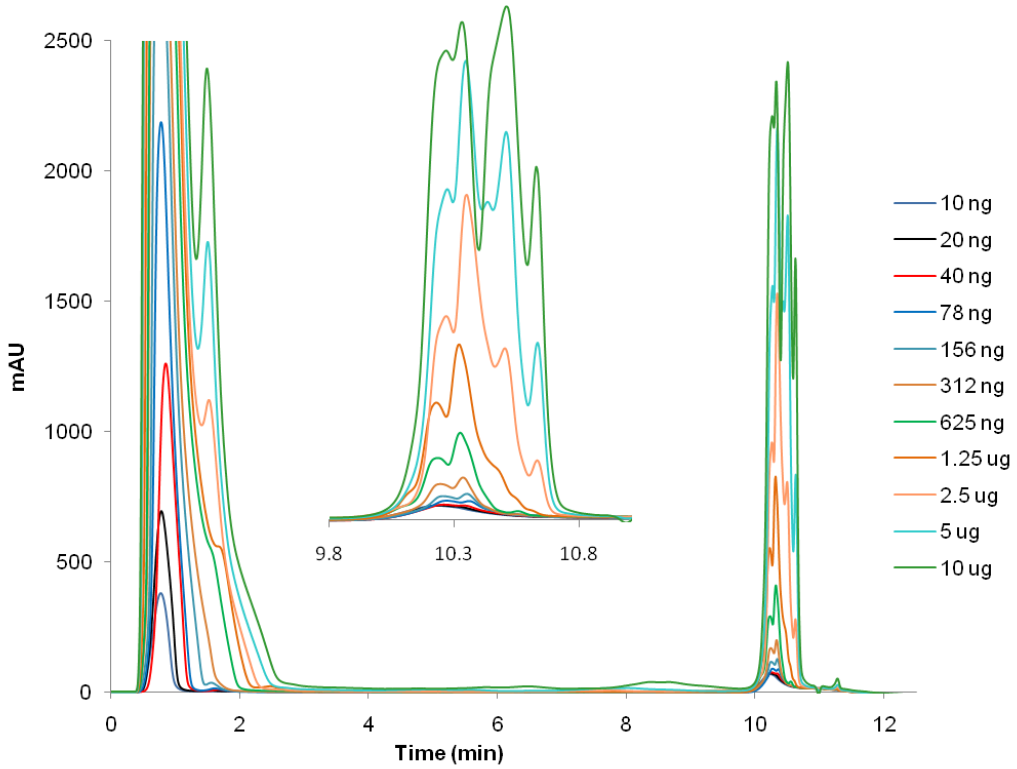


Figure 3.1 Step-gradient LC-UV chromatograms of tryptic digests of a mixture of four protein standards with different amounts of sample injection. A 1.0×50 mm C18 column was used with UV detection at 214 nm. The front peaks were from the salts and other impurities not retained well on the column and the peaks at around 10-11 min were from the eluted peptides.

at 214 nm. Fortunately, the system peak area was relatively constant. Thus the peak area from the peptides could be determined by integrating the overall peak area from the UV chromatogram and subtracting the system peak contribution. To obtain a reproducible system a reproducible system peak from the microbore

column, careful optimization of the flow rates and solvent conditions was carried out. For example, it was found that a much sharper system peak was obtained when the flow rate of 100 μL/min was used, compared to the flow rate of 75 or 50 μL/min. Figure 3.1 shows the chromatograms generated from the injection of 0.039, 0.078, 0.156, 0.3125, 0.625, 1.25, 2.5 and 5 μg of the four-protein tryptic digest. The relative standard deviations in signal responses at different sample loading amounts on the same day were <5%, while the day-to-day relative standard deviations were <10%, indicating that the method can generate reproducible results.

Figure 3.2A shows the integrated peptide peak area as a function of the peptide amount injected. A non-linear transition region between 0.3 μg and 0.6 μg is noticeable. Apparently there are two linear calibration ranges. The calibration equation was $y = 4520x - 48$, with a very good correlation ($R^2 = 0.9991$) by linear regression between 0.039 μg and 0.3125 μg (shown in Figure 3.2B) and $y = 8197x - 551$, with a good correlation ($R^2 = 0.9958$) by linear regression between 0.625 μg and 5 μg (shown in Figure 3.2C). The intercept of both curves has a small negative value, indicating that the system peak area in the sample runs is smaller than that from a blank run. The observation of two distinct linear regions may be related to the appearance of the late eluting peaks in the overall elution profiles as the sample loading amount increases. As the inset of Figure 3.1 shows, the elution profile from the 625 ng injection shows extra peaks at the retention time of greater than 10.5 min. These peaks increase as the injection amount increases above 625 ng. They are likely from the more hydrophobic or larger peptides that retain stronger on RPLC compared to the earlier elution peptides.

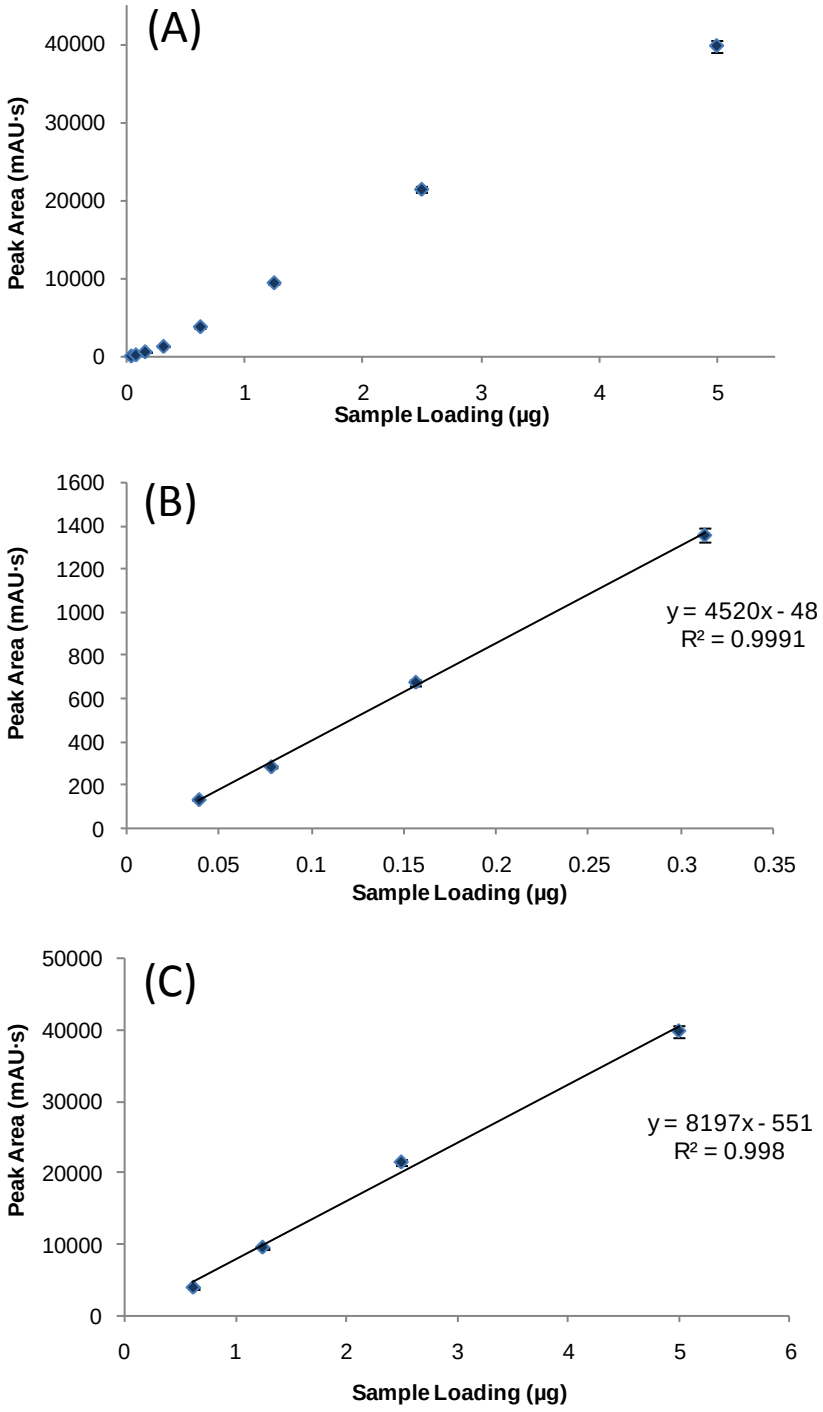


Figure 3.2 (A) Total peak area of the UV absorbance from the eluted peptides as a function of the peptide amount injected. (B) Linear calibration region between 39 ng and 312 ng for the four-protein-digest. (C) Linear calibration region between 625 ng and 5 µg.

Another possible reason of having two linear regions is related to the system peak which might be different from that of a blank injection and might change as the sample injection amount increases. Thus, the subtraction of a constant system peak area during the calculation of the peptide peak area may not be correct.

We have examined the possibility of expanding the linear range using derivative integration. Derivative integration, which calculates and displays chromatograms by d(absorbance)/d(time) vs. time, can place more emphasize on the contribution from analytes and minimize the influence of system peak, since the change in analytes' absorbance with time is much larger than that for the background.[23] There are several methods to do integration on derivative chromatogram, such as using peak height or maximum to minimum. We used the peak height to do integration, as we found some peak was increasing proportionally to the sample amount (see Figure 3.3A). The derivative integration shows very good linearity from 39 ng to 5 μg (see Figure 3.3B). The calibration curve was $y = 5784x + 432$, with a correlation $R^2=0.9992$. From the comparison of the calibration equations obtained from the peak area integration and derivative integration, the derivative integration method also increases the detection sensitivity.

However, derivative integration was more sensitive to the effect of the peak shapes than normal peak area integration. While it provides better linear range and sensitivity, for a real sample, it may present some problems as derivative integration requires constant derivative chromatogram, which is, unfortunately, very sensitive to changes in sample complexity and contaminants. If the sample components do not co-elute and thus show splitting

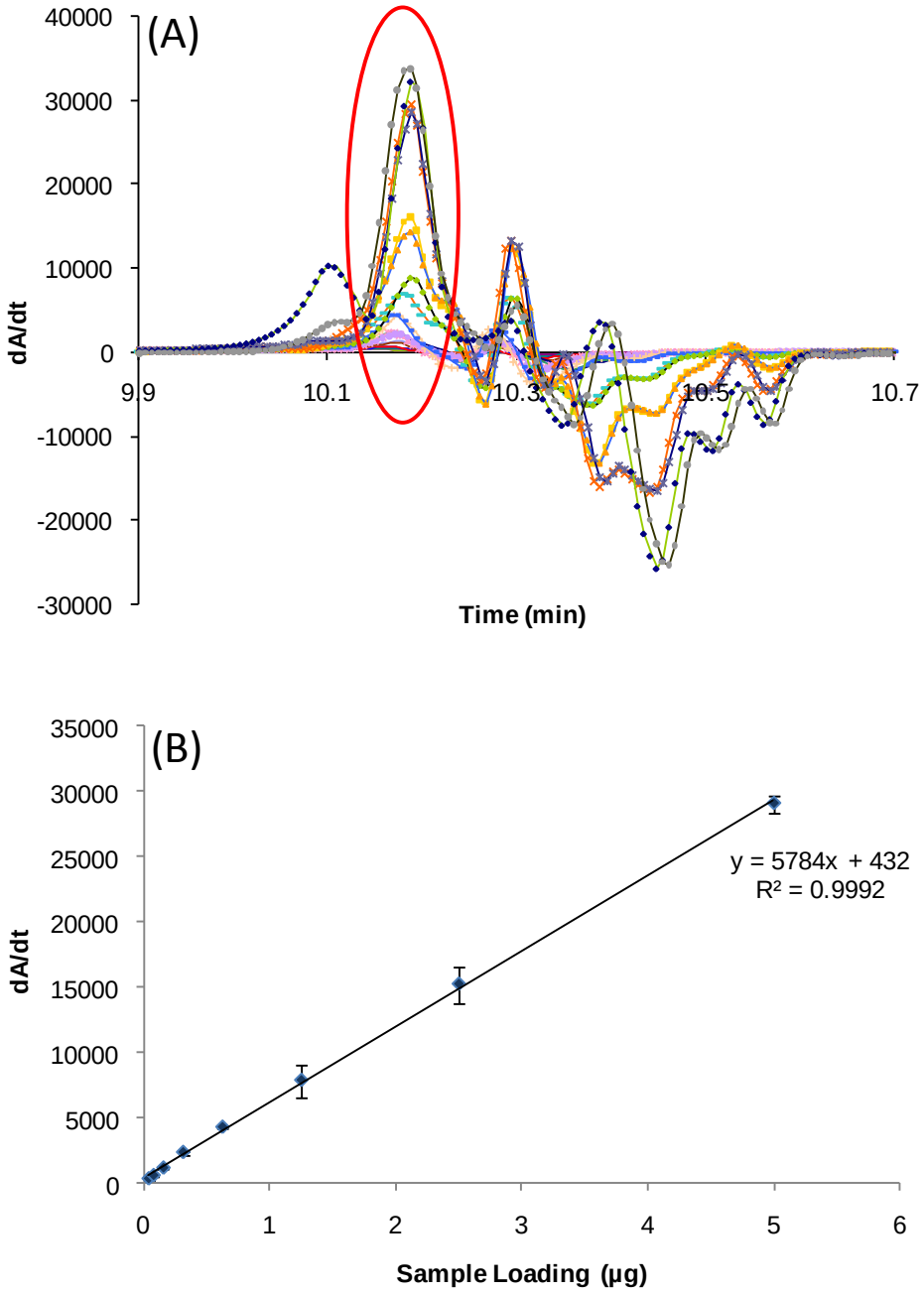


Figure 3.3 (A) Derivative chromatograms of the four-protein-digest with different amounts of peptides injected. The intensity of the circled peptide peak was proportionally related to the sample loading amount, as shown in the calibration curve (B).

chromatographic peaks, its derivative chromatogram would change from a co-elution profile. And the peak used for calibration may not be able to be used for integration. To give an example, for complex protein samples, such as whole cell lysates, their derivative chromatograms were different from the calibration derivative chromatograms produced by the four-protein-digest standard (data not shown). Giving this consideration, we feel that it is difficult to apply the derivative integration method to a complex sample. Therefore, in our subsequent experiment of MCF-7 cells, we used normal peak area for integration. However, we would like to note that, if one is dealing with a relative simple digest (e.g., a few affinity-captured protein biomarkers from a small number of cells), derivative integration is worth exploring as a means of expanding the linear range and detection sensitivity.

The above results indicates that peptide peak area integration from microbore LC-UV can be used for the quantification of total peptide amount down to about 40 ng and up to about 5 μg. It should also be noted that, compared to the use of a 4.6-mm column where a linear calibration was obtained between 0.5 and 20 μg of peptides (R^2=0.9999),[16] the use of a microbore column is more sensitive and better suited for analyzing small amounts of peptides encountered in few cell proteome analysis (see below).

3.3.2 Calibration Curve from Diluted Digests of 30,000 Cells

The overall absorptivity of the peptides from the four-protein-digest standard is similar to that from the complex tryptic digests of cell lysates.[22] Thus, the calibration curves shown in panels B and C of Figure 3.2 can be used for

Table 3.1 Summary of the peptide amounts determined from the dilution of a tryptic digest of a cell lysate prepared using 30000 MCF-7 cells as the starting material.

Theoretical Cell Number	Absorbance (mAu s)	Peptide Amount (μg)
78	46.8 ± 44.0	0.021 ± 0.010
156	50.5 ± 16.3	0.022 ± 0.004
312	258.1 ± 4.8	0.068 ± 0.001
625	763.7 ± 42.0	0.18 ± 0.01
1250	1729.4 ± 25.3	0.39 ± 0.01
2500	3641.6 ± 287.0	0.82 ± 0.06
5000	8131.7 ± 291.8	1.81 ± 0.06

quantification of the total peptide amount produced from a cell lysate. In order to establish a calibration curve of a proteomic digest, we carried out a linear dilution experiment from a tryptic digest of a cell extract. In this case, 6 tubes with each containing 5000 MCF-7 cells were pooled and the pooled cells (30,000) were lysed, reduced by DTT, alkylated by IAA and trypsin digested. Using a large number of cells as the starting material reduces the negative effect of sample loss on the overall peptide amount. The sample was aliquoted to generate the 5000-cell-equivalent digests which was diluted at ½ rate to the equivalents of 2500, 1250, 625, 313, 156 and 78 cells. The diluted digests were quantified using microbore LC-UV. The total peptide amounts of each sample were calculated based on the calibration curve built by the four-protein mixture digest and the quantification results are shown in Table 3.1. Based on the calculation, the

peptide amounts of 78- and 156-cell-equalvent digests were lower than the limit of quantification (LOQ). Thus, the amounts determined for these two samples cannot be trusted, which was also proven by their large values of relative standard derivation obtained.

The calibration curve for 313- to 5000-cell-equivalent digests has an equation of $y = 4555x - 248$, which is statistically the same as that of the four-protein-digest calibration curve (i.e., $y = 4520x - 48$) at 95% confidence level. This is not surprising considering that we were using the four-protein digest to calculate the peptide amount of the 5000-cell-equalvent digest. However, this equation or calibration curve allows us to determine the peptide amount in a real digest of a small number of cells (see below) to gauge if there is any sample loss or interfering species introduced during the sample preparation of few cells. In other words, if we assume that a real digest of 5000 cells (not a diluted digest of a much larger number of cells) is prepared perfectly (i.e., all cells are lysed, all proteins extracted and digested, and all peptides introduced onto LC-UV with no sample loss), the peptide amount of a 5000-cell digest should be 1.81 μg. For a peptide sample prepared from 313 cells, the peptide amount should be about 113 ng. The 40-ng quantification limit should be corresponding to about 110 cells.

3.3.3 Calibration Curve from Digests of Different Numbers of Cells

To study the relation between the peptide amount present in a digest and the actual number of cells used as the starting material for proteome analysis, tryptic digests of different numbers of MCF-7 cells collected by flow cytometry (5000, 2500, 1000, 500, 250 and 100 cells) were analyzed by microbore LC-UV.

Table 3.2 Summary of the peptide amounts determined from the tryptic digests of the cell lysates prepared using different numbers of MCF-7 cells as the starting material.

Cell Number	Absorbance (mAu s)	Peptide Amount (μg)
100	244.0 ± 17.5	0.12 ± 0.01
250	870.9 ± 2.3	0.28 ± 0.01
500	938.2 ± 27.6	0.29 ± 0.01
1000	2211.2 ± 171.9	0.61 ± 0.04
2500	2847.4 ± 14.1	0.76 ± 0.01
5000	7477.2	1.70

Figure 3.4B shows the quantification results based on the use of the 5000-equalvent-cell-dilution calibration curve (Figure 3.4A). Table 3.2 lists the amount of peptides determined for each sample. For the 5000-cell sample, the peptide amount was found to be 1.70 μg, which is close to 1.81±0.06 μg found in the 5000-cell-equalvent sample. Likewise, for the 2500-cell sample, the peptide amount was 0.76±0.01 μg, which is also close to 0.82±0.06 μg found in the 2500-cell-equalvent sample. The slightly reduced amount of peptides found in the actual cell samples, compared to the diluted cell-equalvent samples, likely reflects the fact that sample loss was encountered when processing the 5000 and 2500 cells.

However, for the 1000-cell sample, the peptide amount was found to be 0.61±0.04 μg, which is much more than that expected from the 1000-cell-equalvent sample (i.e., 0.39±0.01 μg). For the 500-cell sample, the peptide amount found was also

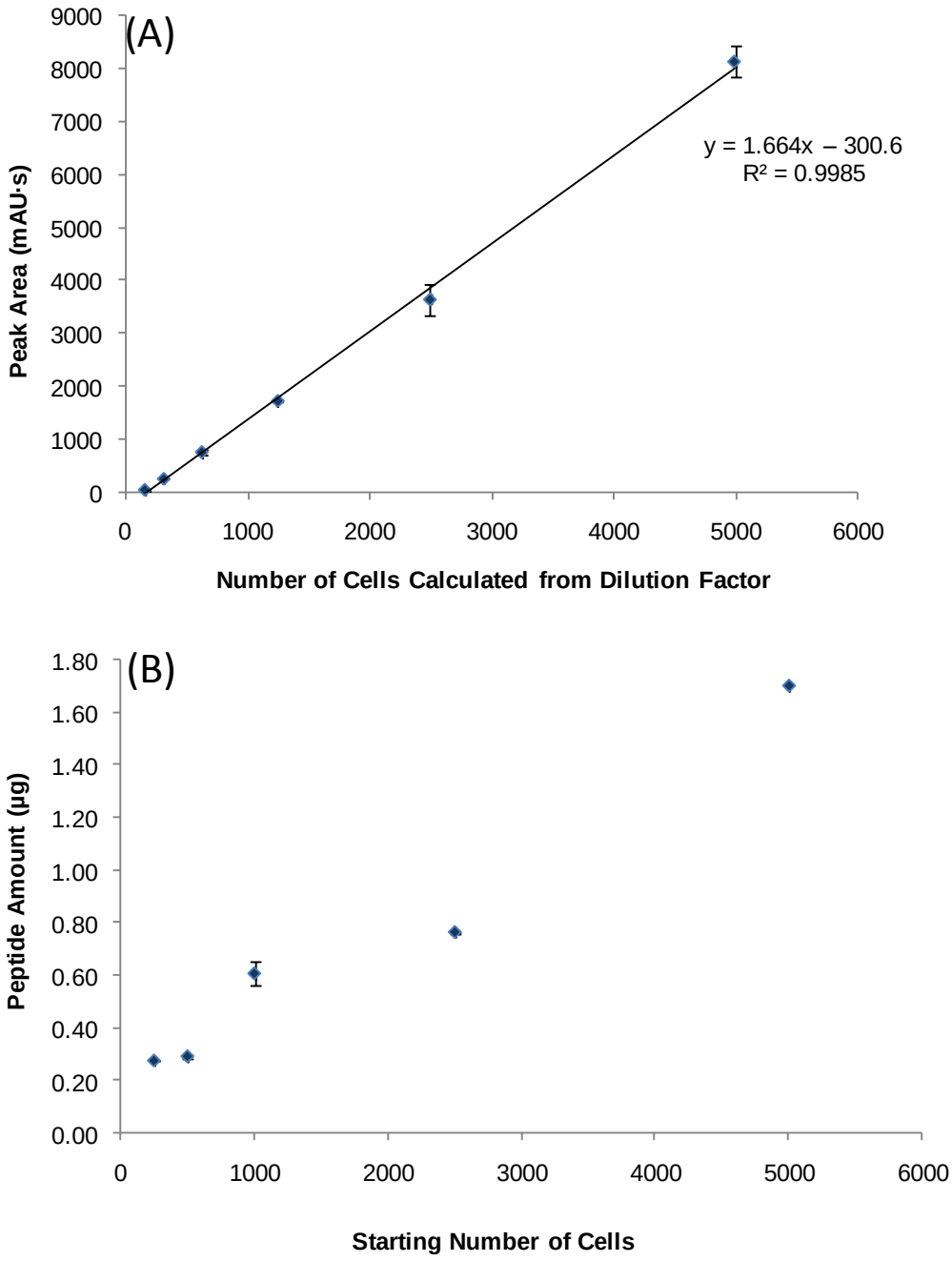


Figure 3.4 (A) Total peak area of the UV absorbance of the eluted peptides from the diluted tryptic digests of a cell lysate prepared using 30000 MCF-7 breast cancer cells as the starting material. The number of cells in the x-axis was calculated by the number of the starting cells multiplying by the dilution factor. (B) Peptide amounts determined from the total peak areas of the UV absorbance of the eluted peptides using 250, 500, 1000, 2500 and 5000 cells as the starting materials.

higher than expected. As indicated earlier, the microbore LC-UV method shouldbe able to quantify the peptides down to 40 ng or about 110 cells. However, for the 250-cell sample, the amount was found to be 0.28±0.01 μg. Even for the 100-cell sample, 0.12±0.01 μg of peptides were found. In fact, for blank samples, similar amounts of peptides could be found. These results were initially somewhat surprising. However, after LC-MS analysis of these peptide samples (see below), it was apparent that peptides from protein contaminants, particularly keratins, were present in these samples. When a large number of cells (e.g., >2500 cells) were processed to generate a peptide sample, the amount contribution from the protein contaminants to the overall peptide quantity in the digest was negligible. However, in working with a smaller number of cells (e.g., <1000 cells), the digest of protein contaminants can be the major constituents of the peptide sample. The implications of these findings on proteome analysis of few cells are discussed below.

3.3.4 Proteins Identification from a Small Number of Cells

Based on the peptide amounts of 1000, 500 and 250 MCF-7 cell lysates, their separation and analysis on LC-ESI MS/MS were optimized. The effect of gradient speed on the number of peptides identified by LC-ESI MS/MS has been investigated previously.[16] Generally speaking, the optimum gradient time increases as the number of cells in a sample increases. Specifically, for the 250- and 500-cell samples, a 90-min gradient was used. The gradient time was increased to 150 min for the 1000-cell samples.

Table 3.3 Numbers of unique peptides and proteins identified (in duplicate) from blanks and tryptic digests of cell lysates prepared using different numbers of MCF-7 cells as the starting material.

	Blank_1 *	Blank_2 *	250_1	250_2	500_1	500_2	1000_1	1000_2
# of peptides	41(36)	46(40)	368	258	410	387	836	784
# of proteins	11(15)	7(10)	126	81	154	122	282	256

*x (y) where x is the number of peptides or proteins identified by using the human proteome database, while y is from the mammal proteome database.

Table 3.3 summarizes the number of peptides and proteins identified from 1000, 500 and 250 MCF-7 cells. In each group, replicate experiments were carried out. The number of peptides and proteins identified increases as the cell number increases. The number change is not linearly proportional to the cell number. For example, an average of 398±12 peptides or 138±16 proteins were identified from the 500-cell samples, while 313±55 peptides or 103±23 proteins were identified from the 250-cell samples. Although the cell number decreases by 2-fold, the number of peptides and proteins identified decreases by about 1.2- and 1.3-fold,

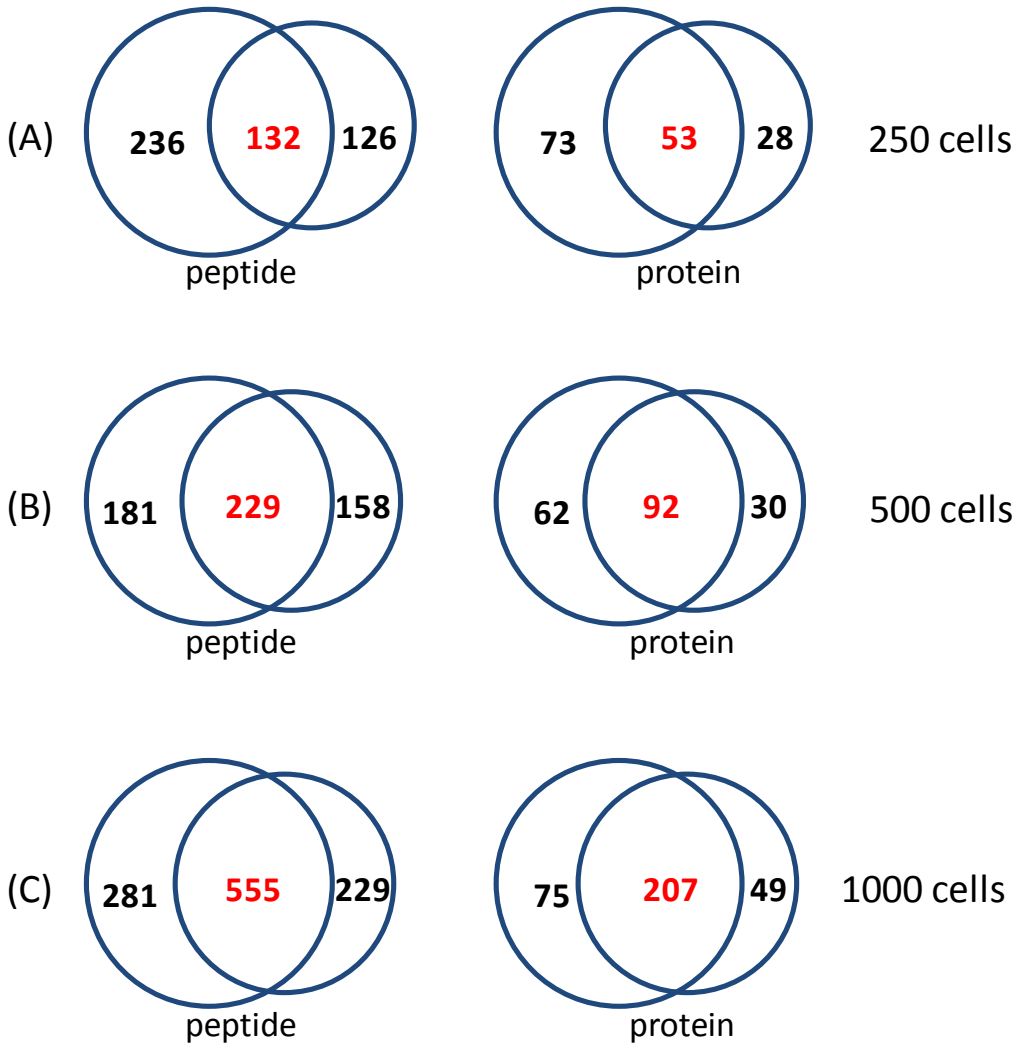


Figure 3.5 Venn Diagrams of the number of overlapped peptides and proteins in the (A) 250 cell, (B) 500 cell, and (c) 1000 cell duplicate samples.

respectively. However, the peptide/protein ratio is kept ~3 for the 1000-, 500- and 250-cell samples. The Venn Diagrams of the number of overlapped peptides and proteins in the 250, 500 and 1000 cell duplicate samples are shown in Figure 3.5. The ratio of the overlapped peptides and proteins are not high, mainly due to the random loss of the samples in the sample preparation and MS analysis process.

The blank samples (i.e., LC-MS grade water, instead of cell solution) were prepared by using the same procedure with the addition of the same reagents as those in analyzing MCF-7 cells. Several peptides were identified from the blanks, but most of them were from various forms of human keratins. Two peptides from a carryover of a previous sample run were found in replicate 1 of the blank, but disappeared in replicate 2. To identify other non-human proteins possibly present in the blank, we searched the MS/MS data against the mammal proteome database. Three and four peptides from pig trypsin were found in the two blanks, respectively. This is not surprising considering that, as the protein concentration reduces, trypsin autolysis becomes more severe. It will be interesting to see if immobilized trypsin can reduce the trypsin autolysis when working with few cells [24-26]. In addition, three non-human keratins were also detected from each blank run. One peptide with high match score was from bovine pancreatic trypsin inhibitor; the source of this protein is unknown. By comparison of the number of peptides identified, many more keratin peptides than trypsin peptides were detected. These results suggest that keratin contamination can be a major issue in working with a small amount of proteomic samples. The sources of these protein contaminants were unknown. However, they were likely from the reagents, as great care was taken to avoid any dust being introduced into the samples during the sample preparation process. The analysis of the 100-cell samples generated the similar results to those from the blank samples. Both the low concentrations of the cellular digest peptides and the presence of background peptides contribute to the failure of identifying any peptides from the MCF-7 cellular proteins.

3.3.5 Microbore LC-UV for Few Cell Proteomics

This work illustrates that microbore LC-UV with a step-gradient can be used as a simple means of determining the total peptide amount in the range from 40 ng to 5 μg, while simultaneously removing salts for shotgun proteome analysis. The low limit of quantification is corresponding to about 110 cells. It should be possible to further improve the detection sensitivity of the method by using an ever smaller diameter column. However, the microbore system should be already adequate for current development of sample handling protocols and applications of the optimized protocols to analyze the proteome of hundreds of cells. This work also points out the need of developing sample preparation methods that will not be readily susceptible to protein contaminations originated from reagents, environment and other sources encountered during the cell sorting, cell lysis, protein extraction and protein digestion steps. While only a few tryptic peptides were observed in few cell analysis, enabling methods for reducing trypsin autolysis and improving digestion efficiencies[24-26] are certainly worth exploring within the context of handling few cells.

3.4 Conclusions

A microbore LC-UV method has been developed for the quantification of the total peptides present in a proteomic digest. This method can be used to quantify 40 ng to 5 μg of peptides while removing salts, which is particularly suitable for shotgun proteome analysis of hundreds of cells. Because the method is non-distractive, the peptide sample collected from LC-UV can be introduced into LC-

MS for identification. Thus, the method offers a means of gauging the peptide sample quality prior to MS analysis and should be useful in few-cell proteomics.

3.5 Literature Cited

(1) Li, L.; Golding, R. E.; Whittal, R. M. *J. Am. Chem. Soc.* **1996**, *118*, 11662.
(2) Harwood, M. M.; Christians, E. S.; Fazal, M. A.; Dovichi, N. J. *J Chromatogr A* **2006**, *1130*, 190.
(3) Li, L. J.; Garden, R. W.; Sweedler, J. V. *Trends Biotechnol.* **2000**, *18*, 151.
(4) Chen, R. B.; Li, L. J. *Anal. Bioanal. Chem.* **2010**, *397*, 3185.
(5) Sun, L. L.; Zhu, G. J.; Li, Y. H.; Yang, P.; Dovichi, N. J. *Anal. Chem.* **2012**, *84*, 8715.
(6) Scatena, R.; Bottoni, P.; Giardina, B. *Biochim. Biophys. Acta-Rev. Cancer* **2013**, *1835*, 129.
(7) Schilling, D.; Todenhofer, T.; Hennenlotter, J.; Schwentner, C.; Fehm, T.; Stenzl, A. *Nat. Rev. Urol.* **2012**, *9*, 448.
(8) Maheswaran, S.; Haber, D. A. *Curr. Opin. Genet. Dev.* **2010**, *20*, 96.
(9) He, J.; Liu, Y.; Lubman, D. M. *Curr. Med. Chem.* **2012**, *19*, 6050.
(10) Gundry, R. L.; Boheler, K. R.; Van Eyk, J. E.; Wollscheid, B. *Proteom. Clin. Appl.* **2008**, *2*, 892.
(11) Thakur, D.; Rejtar, T.; Wang, D. D.; Bones, J.; Cha, S.; Clodfelder-Miller, B.; Richardson, E.; Binns, S.; Dahiya, S.; Sgroi, D.; Karger, B. L. *J. Chromatogr. A* **2011**, *1218*, 8168.
(12) Liu, N. Q.; Braakman, R. B. H.; Stingl, C.; Luider, T. M.; Martens, J. W. M.; Foekens, J. A.; Umar, A. *J. Mammary Gland Biol. Neoplasia* **2012**, *17*, 155.
(13) Switzar, L.; Giera, M.; Niessen, W. M. A. *J. Proteome Res.* **2013**, *12*, 1067.
(14) Shevchenko, G.; Musunuri, S.; Wetterhall, M.; Bergquist, J. *J. Proteome Res.* **2012**, *11*, 2441.
(15) Dai, L.; Li, C.; Lubman, D. M. *Development of Micro-scale Sample Preparation and Prefractionation Methods in LC-MS-Based Proteomic Studies*; Springer: Dordrecht, 2011.
(16) Wang, N.; Xu, M. G.; Wang, P.; Li, L. *Anal. Chem.* **2010**, *82*, 2262.
(17) Maurer, M.; Muller, A. C.; Wagner, C.; Huber, M. L.; Rudashevskaya, E. L.; Wagner, S. N.; Bennett, K. L. *J. Proteome Res.* **2013**, *12*, 1040.
(18) Braakman, R. B. H.; Tilanus-Linthorst, M. M. A.; Liu, N. Q.; Stingl, C.; Dekker, L. J. M.; Luider, T. M.; Martens, J. W. M.; Foekens, J. A.; Umar, A. *J. Proteomics* **2012**, *75*, 2844.
(19) Masuda, T.; Sugiyama, N.; Tomita, M.; Ishihama, Y. *Anal. Chem.* **2011**, *83*, 7698.
(20) Krieg, R. C.; Dong, Y.; Schwamborn, K.; Knuechel, R. *J. Biochem. Biophys. Methods* **2005**, *65*, 13.
(21) Elias, J. E.; Gygi, S. P. *Nat. Methods* **2007**, *4*, 207.
(22) Wang, N.; Xie, C. H.; Young, J. B.; Li, L. *Anal. Chem.* **2009**, *81*, 1049.

(23) Garcia, I.; Ortiz, M. C.; Sarabia, L.; Aldama, J. M. *Anal. Chim. Acta* **2007**, *587*, 222.
(24) Sun, L. L.; Li, Y. H.; Yang, P.; Zhu, G. J.; Dovichi, N. J. *J. Chromatogr. A* **2012**, *1220*, 68.
(25) Jiang, H.; Yuan, H. M.; Liang, Y.; Xia, S. M.; Zhao, Q.; Wu, Q.; Zhang, L. H.; Liang, Z.; Zhang, Y. K. *J. Chromatogr. A* **2012**, *1246*, 111.
(26) Chen, Y. Z.; Wu, M. H.; Wang, K. Y.; Chen, B.; Yao, S. Z.; Zou, H. F.; Nie, L. H. *J. Chromatogr. A* **2011**, *1218*, 7982.

Chapter 4

Quantitative Proteome Analysis of Breast Cancer Tissue Samples for Identification of Putative Cancer Biomarkers

4.1 Introduction

Cancer is a heterogeneous disease with changing levels and activities of important cellular proteins, e.g. oncogenes and tumor suppressors.[1] Cancer is a leading cause of death around the world. Although progress has been made in detection and therapy, cancer still remains a major public health challenge. Among all the cancers, breast cancer is the most common form of cancer in women. The worldwide mortality rate of breast cancer has increased 1.8% annually over the past three decades.[2]

A biomarker is defined as "an objectively measured characteristic that can be evaluated as an indicator of normal biologic processes, pathogenic processes, or therapeutic responses".[3] It is useful in early diagnosis, patient stratification and treatment administration.[4] Breast cancer is classified into different types such as Luminal A, Luminal B and HER2 by the status of well-known breast cancer biomarkers, i.e. estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER-2). In addition, the tumors of triple-negative breast cancer (TNBC) are ER negative, PR negative and HER-2 negative. This type of breast cancer can be divided into two subtypes: normal-like (or

unclassified) subtype and basal-like subtype, characterized by epidermal growth factor receptor (EGFR), and/or cytokeratin 5/6 (CK5/6).[5] Improved survival outcomes have been observed for breast cancer patients diagnosed during the early stages of disease.[5] But not all the breast cancers can be diagnosed by the existing biomarkers used in clinic. Thus novel biomarkers for further subclassification of the breast cancer are needed.

For biomarker discovery, blood or serum, urine, cancer cell line and tumor biopsy tissue are major sample sources used in researches.[6-11] Although tumor biopsy tissue is invasive and difficult to obtain as well as have large microheterogeneity, it is expected that the tumor and its immediate microenvironment have the highest concentrations of potential biomarkers.[7] Thus, in this work, we chose breast cancer tissues for our biomarker discovery study.

Compared to other methods for biomarker discovery like 2-Dimensional Difference Gel Electrophoresis (2D-DIGE)[11], Surface-Enhanced Laser Desorption /Ionization-Time-of-Flight (SELDI-TOF) [5,7,12], Accurate Mass and Time (AMT) tag Strategy combine with LC-FT-ICR MS[10], isotope labeling followed by MS analysis is the most accurate and reliable quantification method. In this chapter, breast cancer tissues from patients were compared by quantitatively analyzing the proteome profiles of malignant tissue and pooled normal tissues from healthy people using 2-MEGA (N-terminal dimethylation after lysine guanidination) isotopic labeling LC/MS to find differentially expressed proteins. The 2-MEGA method blocks the lysine side chain via guanidiation, followed by reductive methylation of the N-terminus using regular or isotopically enriched

formaldehyde. The protocol eliminates the multiple labeling of peptides and improves the detectability of lysine containing peptides by MS, since lysine residues are converted to the more basic homoarginine. In addition, the reagents used in 2-MEGA method are relatively inexpensive, compared with commercial labeling kits.[13-15] Labeled samples were analyzed by 2D LC MS/MS to increase the likelihood of identifying low abundance proteins.[16]

4.2 Experimental

4.2.1 Chemicals and Reagents

Phenylmethanesulfonyl fluoride (PMSF), sodium chloride (NaCl), potassium chloride (KCl), disodium hydrogen phosphate (Na_2HPO_4), monopotassium phosphate (KH_2PO_4), sodium hydroxide (NaOH), phosphoric acid (H_3PO_4), dithiothreitol (DTT), ammonium bicarbaonate (NH_4HCO_3), iodoacetamide, trifluoroacetic acid (TFA), acetone, methanol, LC-MS grade formic acid (FA), *O*-methylisourea hemisulfate, sodium bicarbaonate ($NaHCO_3$), glacial acetic acid, sodium acetate, formaldehyde, and 2-picoline borane were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). Isotopic formaldehyde ($^{13}CD_2O$) was from Cambridge Isotope Laboratories (Andover, MA, USA). Sequencing grade modified trypsin, LC-MS grade water, acetonitrile (ACN), and ProteaseMAX^{TM} were from Fisher Scientific Canada (Edmonton, Canada). The BCA assay kit was from Pierce (Rockford, IL). The total protein extraction kit (TM buffer) was purchased from Biochain Institute (Hayward, CA).

4.2.2 Breast Samples

Frozen human breast tumor tissues were obtained from the PolyomX tumor bank (Cross Cancer Institute, Edmonton, AB, Canada). Informed consent was obtained from the patient for banking and use of tissue for research. This study was approved by the University of Alberta ethics board. The three breast tumor samples, CT0018, MT1275 and MT699, were human small cell carcinomas of breasts from relapsed female patients from 50 to 60 years old. All breast cancers were luminal A type (estrogen receptor +, progestereone receptor +, and epidermal growth factor receptor 2 (ERBB2 or Her2) -). Ten normal breast tissues were from the breasts of healthy women 20 to 60 years old who experienced reduction mammoplasty. All tissues were embedded in OCT (optimal cutting temperature compound) and cut into 20 μm thick curls, then stored in a -80 °C freezer before protein extraction.

4.2.3 Protein Extraction and Digestion of Breast Tissues

The workflow of the experiment is illustrated in Figure 4.1. Each tissue was defrost at 4 °C for 10 min, cut into small pieces of ~1 mm length using a blade, and rinsed with 200 μL PBS (phosphate buffered saline) buffer for 3 times. 200 μL TM buffer (HEPES of pH 7.9, $MgCl_2$, KCl, EDTA, sucrose, glycerol, sodium deoxycholate, NP-40, and sodium orthovanadate)[17] and 4 μL PI (a cocktail of protease inhibitors) in the total protein extraction kit, and 2 μL PMSF were added to each tissue. All tissues were immersed in the extraction buffer and kept on ice. Ultrasonication of samples was performed for 10 s × 3 followed by cooling on ice for 30 s and centrifugation at 18000 *g* for 1 min each time. Then all the samples

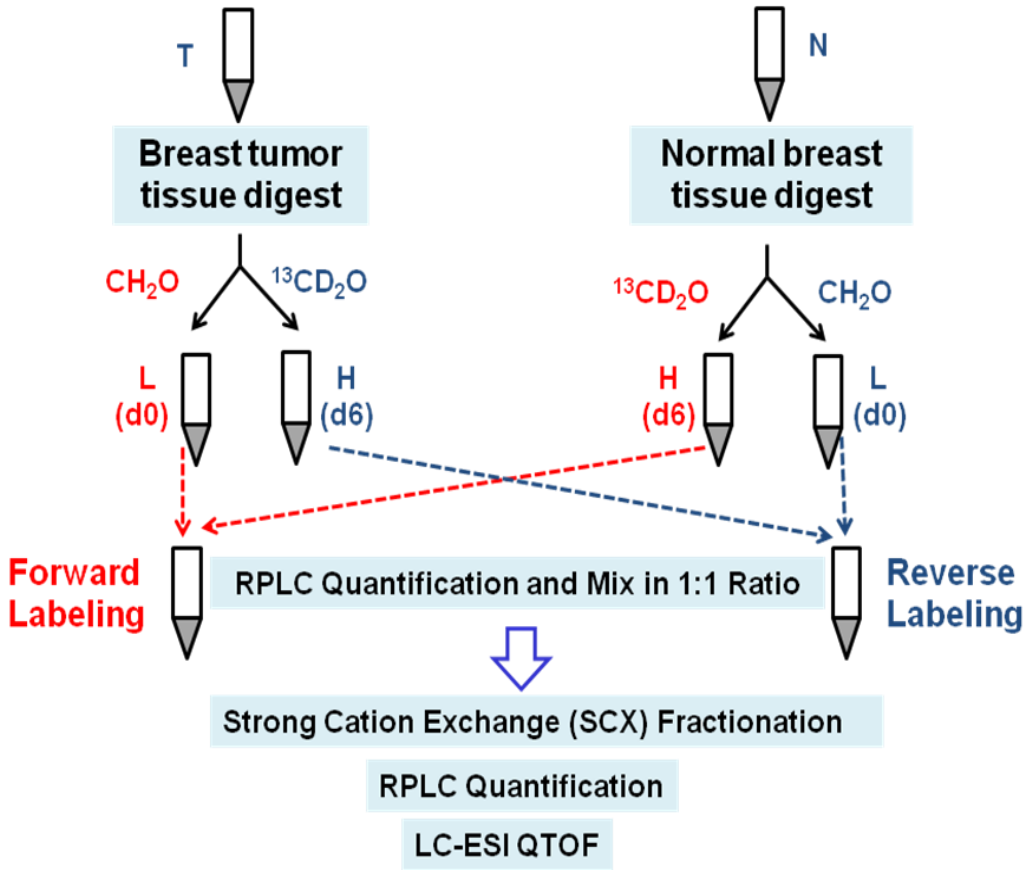


Figure 4.1. Workflow for quantitative proteomic analysis of breast cancer tissues.

were rotated by a Labquake® Shaker Tube Rotator (Lab Industries Inc. Berkeley, Ca) for 20 min in a 4 °C cold room and centrifuged at 18000 *g* for 20 min at 4 ℃. The supernatant was transferred into a 2 mL polyethylene micro-centrifuge tube. Protein concentration was determined using a BCA assay kit.

To process the tissue samples, ~550 µg of each normal tissue were pooled together and made into ten aliquots. Three aliquots of normal tissue extractions were processed in parallel with the three tumor tissue extractions and the rest

were stored at -80 ℃ for future use. The six tissue samples were heated at 65 ℃ for 30 min, reduced with dithiothreitol at 37 ℃ for 30 min and alkylated with iodoacetamide at room temperature in the dark for 15 min. After alkylation, dithiothreitol was added to consume the excess iodoacetamide to prevent over-alkylation.[18] Then an acetone-precipitation step was introduced to remove the salts, lipids and detergents from the extracted proteins. Acetone was precooled to -80 ^{o}C and added to the protein extraction to a final concentration of 80% (v/v), then incubated at -20 ℃ for overnight and centrifuged at 20817 *g* for 15 min. The supernatants were pipetted out, and the remaining pellets were carefully washed twice with cold acetone to effectively remove any residual detergents and lipids. The residual acetone was evaporated at ambient temperature and the pellets were dissolved in 0.2% proteasMAX in 50 mM NH_4HCO_3 in the vial, followed by sonication in an ice bath for ~2 h. After solubilization, 50 mM NH_4HCO_3 was added to dilute the 0.2% proteasMAX to 0.05%. Trypsin solution was then added at a 50:1 ratio (protein:enzyme) and trypsin digestion was performed at 37 ℃ overnight. To ensure complete digestion, fresh trypsin solution was added to each sample at a 500:1 ratio (protein:enzyme) and trypsin digestion was performed at 37 ℃ overnight again.

4.2.4 2-MEAG Labeling

Each digest was separated into two aliquots and labeled using the 2-MEGA protocol ($^{12}CH_2O$ and $^{13}CD_2O$ as the light and heavy labels, respectively). The 2-MEGA labeling method was similar to that described previously.[13] In brief, peptide solutions (0.5 μg/μL) were adjusted to pH 11 using 2 M NaOH. 2 M *O*-

methylisourea hemisulfate solution was prepared in 100 mM $NaHCO_3$ and was adjusted to pH 12 with 2 M NaOH. The guanidination reaction was allowed to proceed for 20 min at 60 ℃ before adjustment to pH 7 using 6 M HCl and further adjustment to pH 7 using 1 M acetate buffer. Formaldehyde ($^{12}CH_2O$) and isotopic formaldehyde $^{13}CD_2O$ (4% in H_2O, (w/w)) were added to aliquoted breast tissue digests separately, followed by subsequent addition of 2-picoline borane (1 M in methanol). The dimethylation reaction was allowed to proceed to completion (20 min) at 37 ℃ to complete before adding 1 M ammonium bicarbonate to quench the reaction. After 30 min incubation at 37 ℃, the solutions were adjusted to pH <2 using 10% TFA.

4.2.5 Desalting and Quantification

Samples were desalted and quantified using an LC-UV method as described previously.[19] The desalting and quantification setup consisted of an Agilent 1100 HPLC system (Palo Alto, CA) with a UV detector. A 4.6 mm i.d. × 5 cm Polaris C18-A column with a particle size of 3 μm and 180 Å pore sizes (Varian, Palo Alto, CA, USA) was used for the desalting of the labeled tryptic peptides at room temperature. After loading 200 μL peptide sample, the column was flushed with 97.5% mobile phase A (0.1% TFA in water). Thus salts and other interferences, such as DTT and IAA, were effectively removed. Subsequently, the concentration of mobile phase B (0.1% TFA in acetonitrile) was step-wise increased to 85% to ensure the complete elution of the peptides from the column, followed by 15 min re-equilibration with 97.5% mobile phase A. The eluted peptides were monitored and quantified using the UV detector operated at 214 nm.

The peptides were fractionated between 7.55 and 8.25 min. During the peptide elution process a chromatographic peak was produced and the amount of peptides was determined based on the peak area. The calibration curve was generated as y=451.8x - 77.6 using various amounts of BSA tryptic digests, where y refers to the peak area of a sample injection minus the system peak area of a blank injection and x refers to the peptide amount analyzed.

4.2.6 Strong Cation Exchange (SCX) Chromatography

The desalted and quantified labeled digests were completely dried using a SpeedVac (Thermo Savant, Milford, MA), and then reconstituted in 50 μL 0.1% formic acid. The light labeled tryptic digests of each tumor tissue was mixed with heavy labeled normal tissue at 1:1 (w/w) ratio. The forward labeling and reverse labeling (or AB/BA labeling) were performed by switching the isotopic labeling assignments, as shown in Figure 4.1. The 100 μL mixtures were separated by SCX chromatography equipped with a UV detector. A highly hydrophilic polysulfoethyl A column (2.1 mm i.d. x 250 mm with particle size of 5 μm diameter and 300 Å pores) from PolyLC was used for the strong cation exchange separation of the mixed labeled tryptic peptides of the breast tissue proteins. Gradient elution was performed with mobile phases A (10 mM KH_2PO_4, pH 2.76) and B (10 mM KH_2PO_4, pH 2.76, 500 mM KCl) at a flow rate of 0.2 mL/min. The gradient profile was as follows: 0 min, 0% B; 10 min, 0% B; 11 min, 4% B; 27 min, 20% B; 49 min, 60% B; 55 min, 100% B; 60 min, 100% B; 62 min, 0% B; 72 min, 0% B. The fractions were collected every 5 min for the first 20 min, every 1 min between 20 and 60 min, and every 1 min thereafter. Therefore, a total of 46

fractions were collected and directly desalted and quantified by the desalting setup described above. The fractions to be pooled were dependent on the peptide amount of each fraction. About 1 μg of peptide sample should be injected into the LC-ESI MS/MS instrument for sequencing.[19] The SCX fraction that had less than 1 μg amount would be combined with the next SCX fraction(s) to make the total amount of peptides in a pooled sample greater than 1 μg.

4.2.7 LC-ESI MS/MS

The desalted SCX fractions were completely dried by SpeedVac and reconstituted to 0.2 μg/μL with 0.1% formic acid, then injected 5 μL onto a 75 μm × 100 mm Atlantis dC18 column (Waters, Milford, MA) and analyzed by a QTOF Premier mass spectrometer (Waters, Manchester, U.K.) equipped with a nanoACQUITY Ultra Performance LC system (Waters, Milford, MA). Solvent A used was 0.1% formic acid in water, and Solvent B was 0.1% formic acid in ACN. Peptides were separated at 35 ℃ by a 120 min gradient (2-6% Solvent B for 2 min, 6-25% Solvent B for 95 min, 30-50% Solvent B for 10 min, 50-90% Solvent B for 10 min, 90-5% Solvent B for 5 min) after column equilibration at 2% Solvent B for 20 min and electrosprayed into the mass spectrometer fitted with a nanoLockSpray source at a flow rate of 300 nL/min.

A survey MS scan was acquired from m/z 350-1600 for 0.8 s, followed by 4 data-dependent MS/MS scans from m/z 50-1900 for 0.8 s each. For both dynamic and precursor ion mass exclusion, a mass tolerance window of 80 mDa was applied. The collision energy used to perform MS/MS was varied according to the mass and charge state of the eluting peptide ion. A mixture of leucine enkephalin and

(Glu1)-fibrinopeptide B, used as mass calibrants (i.e., lock-mass), was infused at a flow rate of 300 nL/min, and a 1 s MS scan was acquired every 1 min throughout the run.[18] The peptide precursor ion exclusion (PIE) strategy[20] was applied to exclude relatively high-abundance peptides identified from the adjacent two SCX fractions to enable additional and less abundance peptides to be analyzed and identified. An exclusion list was generated as previously,[20,21] based on MASCOT (Matrix Science, London, U.K.) searching results. The excluded peptides in the list generally had a MASCOT score 10 points equal to or higher than the identification threshold score at the 95% confidence level. These peptides with a certain m/z value would be excluded at its previous retention time for 150 s during the new LC-ESI run.

4.2.8 Data Analysis

Raw LC-ESI data were lock-mass corrected, de-isotoped, converted to peak lists, and identified via Mascot database searching using the peak lists, as well as quantified using the summed peak areas of all isotopic peaks in the XIC (extracted ion chromatogram) by the MASCOT distiller program (version 2.2). Database searching was restricted to Homo sapiens (human) in the SWISSPROT database (October 4, 2007) and 17317 entries were searched. The following search parameters were selected for all database searching: enzyme, trypsin; missed cleavages, 1; peptide tolerance, 30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, (1+, 2+, and 3+); fixed modification, Carbamidomethyl (C), Guanidinyl (K), Dimethylation (N-term; $+C_2H_4$) or Dimethylation (N-term; $+^{13}C_2D_4$); variable modifications, oxidation (M). The instrument-type setting was

modified from standard ESI-QTOF fragmentations by adding a-ion fragments. For isotopic peak fitting for quantification, the correlation threshold was 0.9. Peptides identified had at least one hit with an ion score above the threshold. Reported quantification ratios as well as protein names, access IDs, molecular mass, peptide sequences, ion score, correlation score, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses were exported to Microsoft Excel for further analysis. The quantification ratios (tumor vs. normal) of the same peptide across multiple fractions and of different charge states were averaged together using the geometric average. The quantification ratio of a protein was obtained by the geometric average of ratios of different peptides from the same protein.

The Gene Ontology (GO) terms of cellular component, molecular function and biological process for the identified proteins were extracted from the ExPASy (Expert Protein Analysis System) Proteomics Server according to their Swiss-Prot IDs (http://ca.expasy.org). This information was referenced from the QuickGO GO browser (http://www.ebi.ac.uk/ego).[19] Thus, proteins not described in the Gene Ontology database and proteins with unspecified cellular locations were categorized as "other".

Ingenuity pathways analysis (IPA)[22,23] and Metacore pathways analysis (GeneGO)[24,25] were used to map the differentially expressed proteins into the protein-protein interaction networks. Differentially expressed proteins were converted into appropriate gene symbols or directly uploaded into IPA and Metacore for analysis. The protein-protein interaction analysis[26] was based on

literature-confirmed interactions from the Human Protein Reference Database, yeast-two-hybrid-defined interactions, and predicted interactions generated by a Bayesian analysis. For network analysis, the bridge interaction algorithm to map the shortest path for interactions was used.[27]

4.3 Results and Discussion

4.3.1 Protein Exaction and Digestion

All tissues were ~20 mg to 80 mg before exaction, including the weight of OCT wrapped outside the tissue. The actual weight of each tissue was unknown. After protein exaction by ultrasonication in TM buffer, BCA assay was carried out to quantify the protein amount. Extraction efficiency was 1.2 – 4.6% as calculated by dividing the amount of extracted proteins by the weight of the tissue. Note the actual extraction efficiencies should be larger than calculated, since the tissue weight included the weight of OCT. Variances of extraction efficiencies were most likely from the different weights of OCT.

Large amounts of tissue debris existed after protein extraction. We transferred the supernatant to another tube and added TM buffer again to the debris of three tumor tissues to do a second protein extraction. BCA assay revealed less than 30 μg proteins were extracted for the second extraction, which was less than 5% of the protein extracted from the first extraction. Therefore it was unnecessary to do a second extraction to ensure the success of protein extraction from the tissue.

Because of lipids and detergents present in the protein extraction, acetone precipitation is irreplaceable. Acetone precipitation was performed after DTT reduction and IAA alkylation to remove salts in case any excess reagent may cause side reactions in the 2-MEAG labeling process. The amounts of lipids were so large that we could see white lipids almost cover all the surface of the protein extract, especially for the extraction of normal breast tissues. We were not surprised to find the lipids of enormous amounts, considering the breast tissue is filled with lipids. We washed the pellets twice using cold acetone to ensure the efficient removals of detergents and lipids after acetone precipitation. Then the pellets were reconstituted in 0.2% proteasMAX in 50 mM NH_4HCO_3. No lipids were seen in the solution. ProteasMAX is an acid-labile surfactant suitable for solubilization of proteins to be analyzed by mass spectrometry. It denatures the protein in the same manner as SDS, helps hydrophobic proteins dissolve into the solution and makes proteases access to the cleavage sites more easily. After sonicating for ~1 h, the solution had already become clear and no obvious precipitation was seen, demonstrating the good solubilizing ability of proteasMAX. However, the concentration of proteasMAX must be diluted before digestion to prevent protease denaturation. ProteasMAX gradually degrades to hydrophobic and hydrophilic products (structures shown in Figure 4.2) during the digestion process. Thus, no additional degradation step is needed after digestion. The degraded hydrophobic product could be removed by centrifugation, as its solubility is low in aqueous solution. The solution became cloudy when the trypsin digestion was finished. The precipitates were likely the degraded hydrophobic products of proteasMAX and highly hydrophobic peptides or

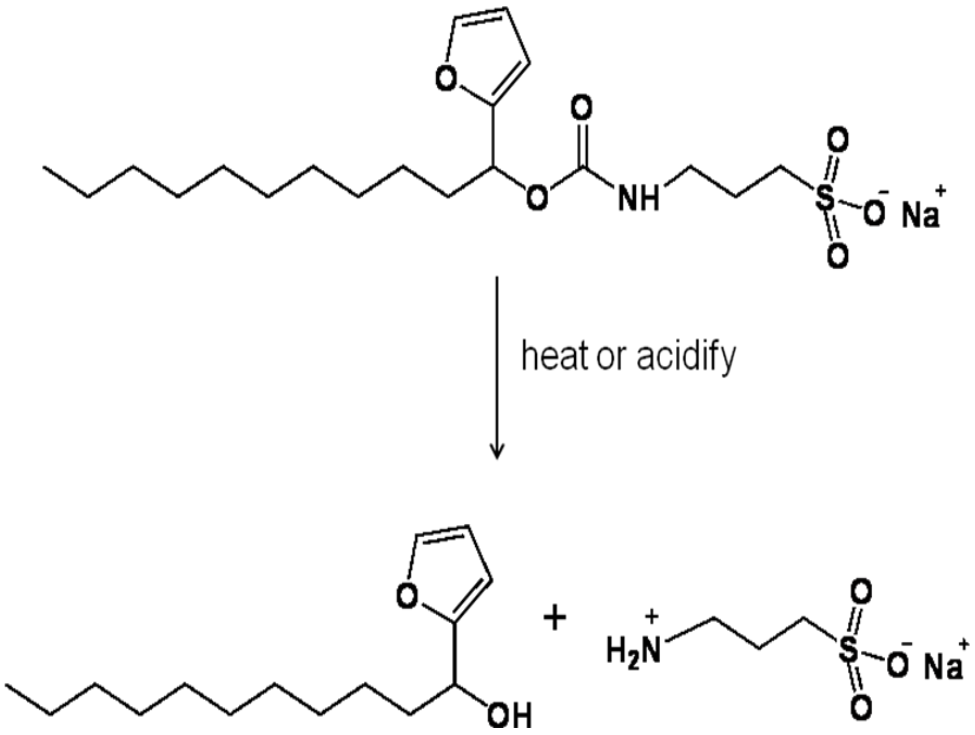


Figure 4.2 The structure of the acid-labile surfactant, proteasMAX, and its degraded products when heated or acidified.

proteins. They precipitated out from the solution without proteasMAX present and were removed by centrifugation afterwards. In the final LC ESI MS/MS analysis, no polymers appeared in the chromatogram and peptide signals were adequate, indicating the successful removal of lipids and detergents by acetone precipitation and there was no interference from proteasMAX.

4.3.2 2-MEGA Labeling

The reaction scheme is illustrated in Figure 4.3a. 2-MEGA stands for "N-terminal dimethylation after lysine guanidination". Briefly speaking, O-methylisourea selectively guanidinates lysine residues of peptides, blocking them from being dimethylated in the subsequent reaction. Then in the presence of formaldehyde and borohydride, the N terminals of peptides were dimethylated. With regular formaldehyde, the masses of peptides increase about 28 Da; these light-labeled peptides were called d0. While using heavy formaldehyde, the masses of peptides increase about 34 Da; these heavy-labeled peptides were called d6, which were 6 Da heavier than the light labeled peptides. Comparing to other methods, 2-MEGA is an indiscriminate peptide labeling method and inexpensive. The reaction schemes of side products (guanidination at N-terminal and dimethylation at lysine) are also demonstrated in Figure 4.3b and 4.3c. Peptides with missed labeling are rather rare when excess reagents are used. Generally speaking, all these by-products take up ~5%. In this work, we checked the efficiencies of 2-MEGA labeling using one of the breast tumor tissues first before we started the quantification experiment to ensure the sample handling procedure used was compatible with the 2-MEGA labeling. About 1 μg tryptic digest of the breast tissues labeled with regular or isotopic formaldehyde was analyzed by 1-D LC MS/MS with Waters nanoACQUITY Ultra Performance LC system and QTOF Premier mass spectrometer. The labeling efficiencies were ~94% for the breast tumor tissue, which was calculated by dividing the number of correctly labeled peptides by the number of all identified peptides. This result is consistent with what we observed for labeling other cellular proteomic samples. Thus, 2-MEGA

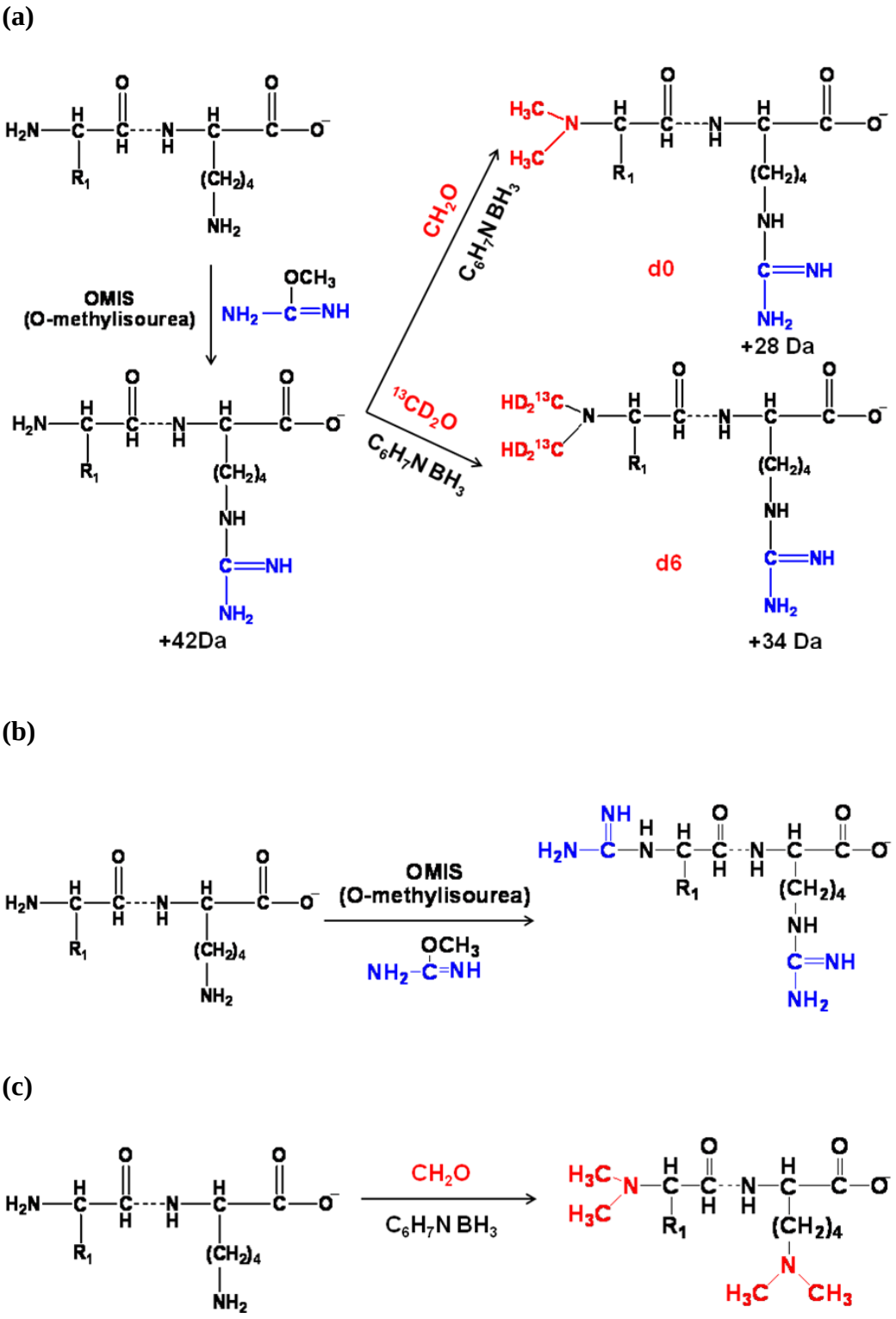
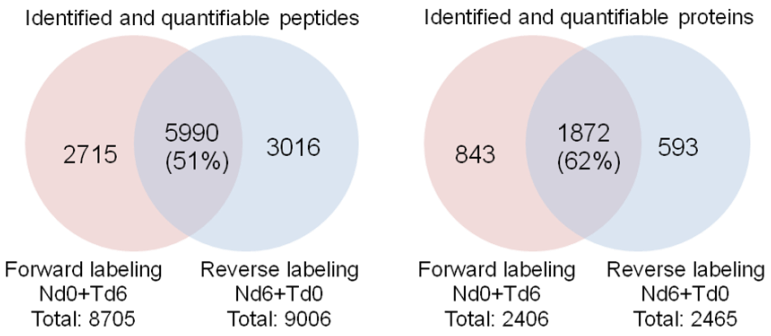
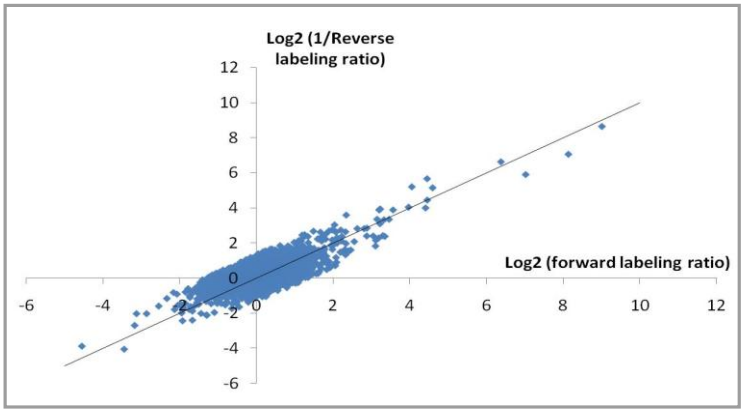


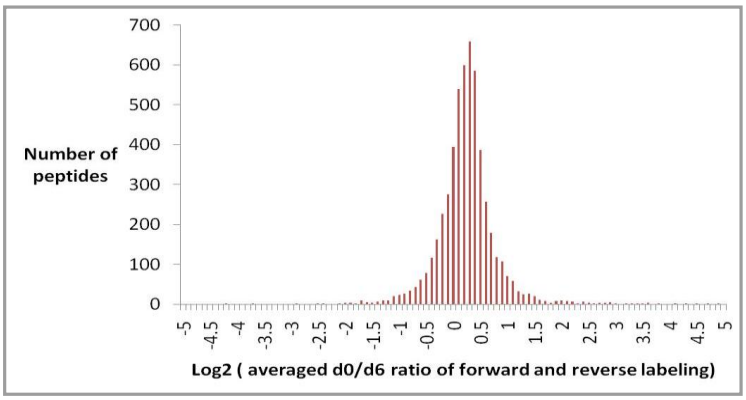
Figure 4.3 (a) Reaction scheme of 2MEAG labellng reaction. The major by-products formed by (b) guanidination at N-term, and (c) dimethylation at lysine.



(a)



(b)



(c)

Figure 4.4 2D LC MS/MS analysis results of the comparison experiment of tumor MT1275 vs. normal tissue. (a) The diagram showing the number of identified peptides and proteins in forward and reverse labeling samples. (b) Log-log plot of quantification ratios of peptides in forward and reverse labeling. (c) Distribution histogram of the log2 value of geometric averaged peptide quantification ratios.

labeling could be used for quantitative tissue proteomics using the sample preparation protocol described in the experimental section.

4.3.3 Comparison Experiment

At first, the breast tumor of case MT1275 was compared with the normal tissue from the same individual subject. The normal tissue was in a region of the same tissue where no or a few tumor cells are found. ~125 μg sample was loaded onto the SCX column and 23 fractions after pooling adjacent SCX fractions were injected for LC MS/MS analysis. Figure 4.4 shows the results. Approximately 9000 peptides and 2400 proteins were identified and quantified for each sample, and ~60% common proteins were found between the two (Figure 4.4a). We discarded about 12% of the peptides because their relative difference between the forward and reserve labeling was larger than 0.77. Log-log plot of quantification ratios was constructed (Figure 4.4b), as well as a distribution histogram of averaged log ratios (Figure 4.4c) for the remaining peptides. In the remaining 1657 proteins, 10% were up regulated, while 3% were down regulated.

Variances of different samples might come from genetic polymorphism, gender, age, ethnicity, life style, dietary influences, collection, storage, processing time relative comparison so that the variations originating from different people could be minimized. However, the real normal tissues are not easy to obtain. Oppenheimer *et al.* confirmed the adjacent normal tissue around the tumor region marked by a pathologist was not so normal as it looked like by MALDI MS analysis.[28,29] In addition to that, the amount of normal tissue around the tumor region is usually too little to be used for the relative quantification experiment.

Table 4.1 The number of identified peptides and proteins of normal and tumor samples with average score and the total number of identified peptides and proteins in each forward and reverse labeling experiment.

	Peptides		Proteins		Score		Total ID	
	N	T	N	T	N	T	Peptides	Proteins
CT0018_F	1539	3462	538	1337	51	50	4542	1591
CT0018_R	1869	4075	638	1539	55	52	5312	1804
MT1275_F	1731	6967	541	2271	56	56	8030	2483
MT1275_R	1879	6802	589	2353	60	58	7875	2539
MT699_F	2386	2895	865	1142	51	50	4317	1538
MT699_R	1994	4198	619	1365	54	51	5163	1568

F: forward labeling
R: reverse labeling
N: normal
T: tumor

Table 4.2 The number of identified and quantified peptides and proteins and percentage of common peptides and proteins in forward and reverse labelling experiment for each case.

	Identified and Quantified		Common	
	Peptides	Proteins	Peptides	Proteins
CT0018_F	4771	1526	43%	54%
CT0018_R	5665	1726		
MT1275_F	8744	2283	44%	59%
MT1275_R	8559	2343		
MT699_F	4738	1440	40%	50%
MT699_R	5516	1451		

F: forward labelling
R: reverse labelling

Thus we used pooled normal tissues in the following experiments to do the comparison.

Three cases of tumors of luminal A type were compared with aliquots of pooled normal tissues separately. The number of identified peptides and proteins of normal and tumor samples for each comparison experiment is listed in Table 4.1. In total, ~4000 to 8000 peptides and 1500 to 2500 proteins were identified for each 2D LC MS/MS analysis. It was not surprising to identify more peptides and proteins for MT1275, for its starting amount was almost twice that of the other two. The average Mascot search score of MT1275 was also higher, telling us the search results were more confident. Thus, for 2D LC MS/MS analysis, it is better to start from more samples; however, in real situations, tissue samples might be available in limited quantity. The numbers of identified peptides from the tumor samples were much larger than that of the normal tissues. The normal sample was a mixture of ten normal breast tissues that was apparently much more complex than the tumor samples. Therefore, fewer peptides of sufficient abundances can be identified for the normal sample.

The numbers of identified and quantified peptides and proteins are listed in Table 4.2. Not all the identified peptides could be quantified, so the number of quantified peptides and proteins is smaller than its counterpart in Table 4.1. A quantification example of a 2+ peptide (EEAENTLQSFR) from human vimentin protein is shown in Figure 4.5. This peptide is doubly charged, so in the spectra, the mass difference between the monoisotopic peaks of d0 and d6 is 3 m/z. The continuous red curve is the mass spectrum of the real sample, while the

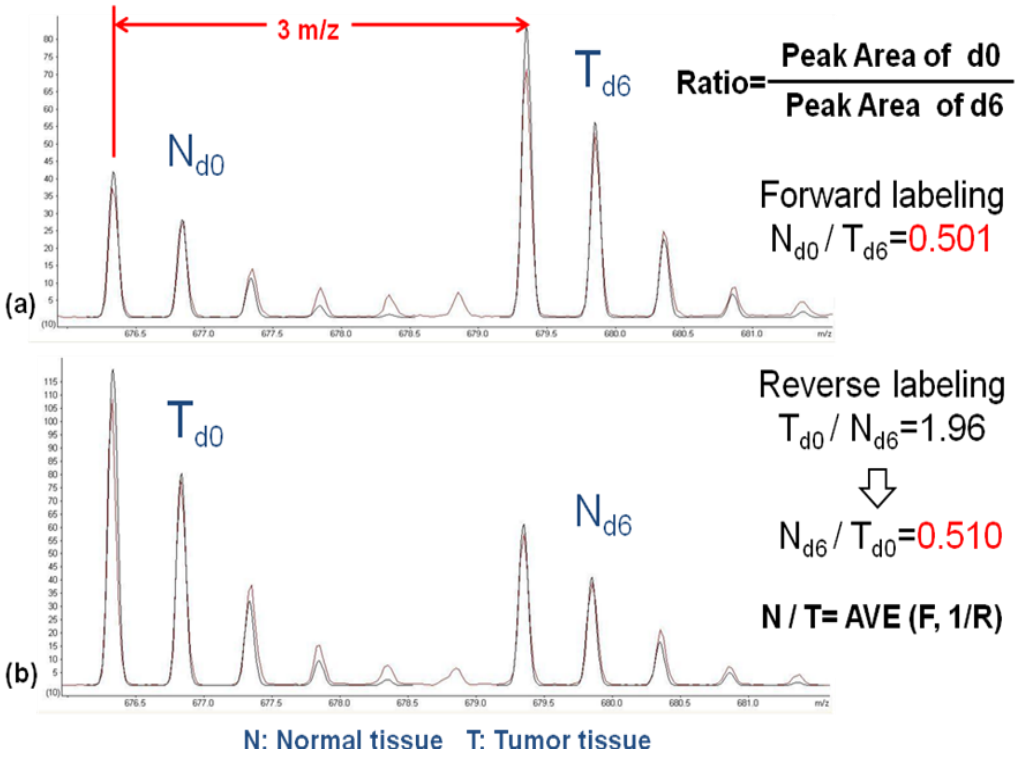


Figure 4.5 Quantification example of 2+ peptide EEAENTLQSFR from Human Vimentin Protein. The black line is the fitting curve and the red line is the mass spectrum.

discontinuous black curve is the fitting curve by the software. Quantification ratio as calculated by dividing the total peak area of d0 by the total peak area of d6. In the forward labeling experiment, where d0 of normal tissue was mixed with d6 of tumor tissue, the quantification ratio of d0 to d6 (i.e. normal to tumor) is around 0.5. In the reverse labeling experiment, where d0 of tumor tissue was mixed with

d6 of normal tissue, the quantification ratio of d0 to d6 (i.e. tumor to normal) is around 2. The reciprocal of it is 0.5, which is in coincidence with the quantification ratio in the forward labeling experiment. The averaged forward labeling quantification ratio and the reciprocal of the reverse labeling quantification ratio is the final quantification result for this peptide.

Figure 4.6 shows the log-log plots of quantification ratios of peptides in forward and reverse labeling for CT0018, MT1275 and MT699 after we removed peptides with relative differences of larger than 0.77. Linear regression was carried out for each case. The slope of each line was around 0.94, indicating a good correlation for these peptides between the forward and reverse labeling experiments. The averaged relative quantification ratio of all the peptides also shifted to 1.73, 2.67, and 1.34 from 1 for CT0018, MT1275, and MT699 respectively. The basis for the biomarker quantification experiment was that most peptides were unchanged. Thus the quantification ratios were normalized to make the average ratio equal to 1. Then the quantification ratio of a protein was obtained by the geometric average of ratios of different peptides from the same protein. 606, 880 and 456 differentially expressed proteins with a quantification ratio of less than 0.67 or larger than 1.5 were found in CT0018, MT1275 and MT699 respectively (Figure 4.7a). 119 proteins were differentially expressed in all three cases. These common proteins along with MASCOT peptide identification information are listed in Table 4.3 and Table 4.4. More proteins (76) were down-regulated than up-regulated (43).

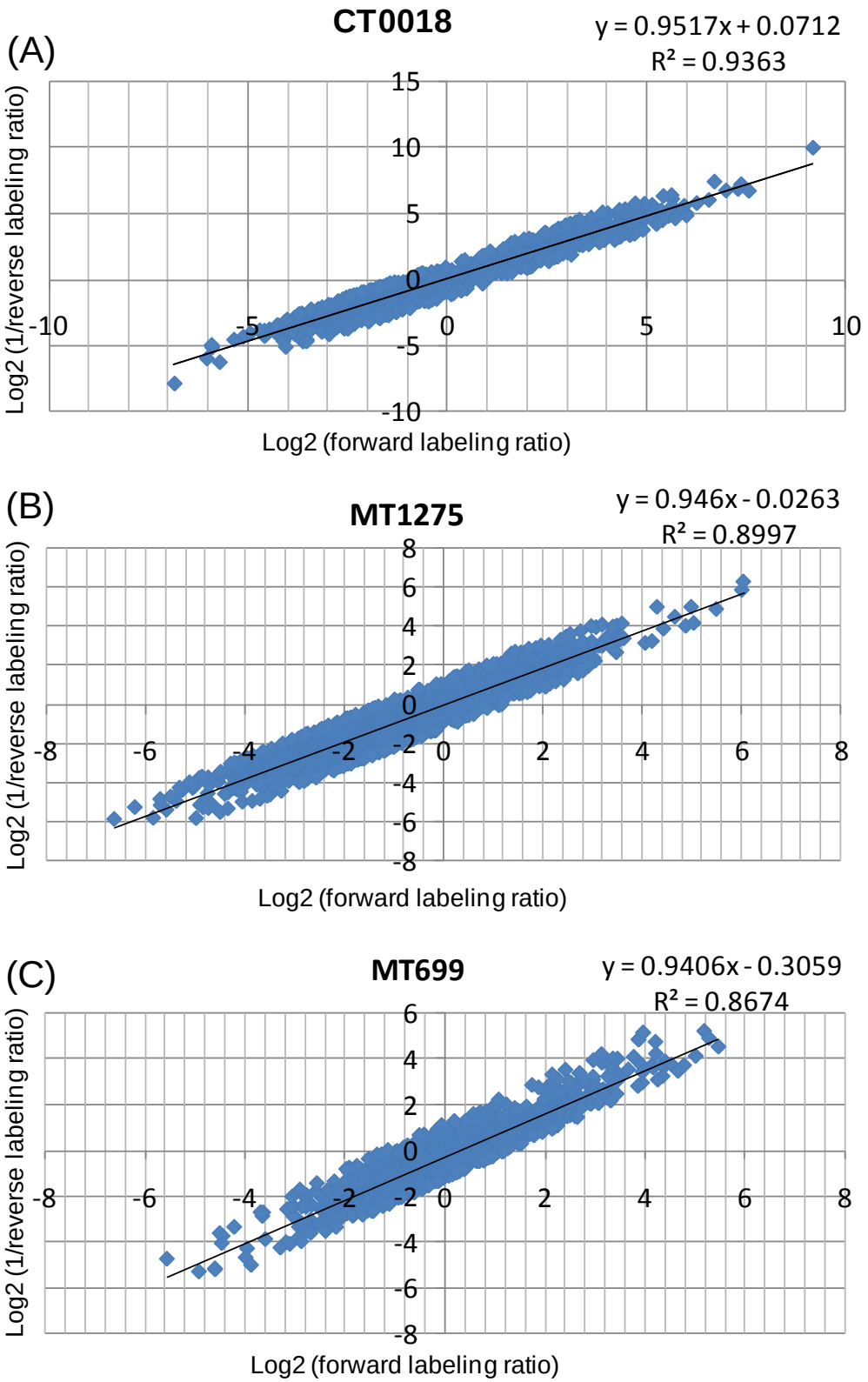


Figure 4.6 Log-log plot of quantification ratios of peptides in forward and reverse labeling for (a) CT0018, (b) MT1275 and (c) MT699.

Table 4.3 The 43 proteins up-regulated in all three cases.

Cancer	Protein Name	Quantification ratio			Gene Name	AccessID	MW(kDa)	Gravy
		CT0018	MT1275	MT699				
b	1433Z_HUMAN	2.67	1.81	1.59	YWHAZ	P63104	28.773	0.48
b	ANM1_HUMAN	2.48	2.47	2.82	PRMT1	Q99873	43.221	1.10
O	ARC1B_HUMAN	3.69	1.74	3.11	ARPC1B	O15143	42.842	-0.22
O	ARP2_HUMAN	2.47	3.28	3.71	ACTR2	P61160	46.096	0.16
O	ARPC2_HUMAN	3.18	1.80	2.72	ARPC2	O15144	35.51	-0.30
O	CALR_HUMAN	1.53	1.97	2.13	CALR	P27797	50.076	-1.26
N	CAN1_HUMAN	3.28	1.89	2.09	CAPN1	P07384	84.048	0.08
N	CAND1_HUMAN	1.70	1.79	2.09	CAND1	Q86VP6	141.311	1.10
b	CAPG_HUMAN	6.05	1.74	1.98	CAPG	P40121	39.815	1.10
O	CDS2_HUMAN	1.55	1.68	1.57	CDS2	O95674	52.655	-0.41
O	CLH1_HUMAN	1.63	2.01	2.20	CLTC	Q00610	197.322	-0.24
N	COPB_HUMAN	3.18	1.90	4.58	COPB1	P53618	111.148	-1.22
O	COPG_HUMAN	2.17	5.62	2.01	COPG	Q9Y678	101.264	0.24

b	CSPG2_HUMAN	2.15	1.77	7.26	VCAN	P13611	380.923	-1.40
N	DC1I2_HUMAN	2.78	2.12	1.88	DYNC1I2	Q13409	73.526	-1.20
O	EF1D_HUMAN	2.06	2.57	2.55	EEF1D	P29692	32.085	-0.05
O	EF1G_HUMAN	3.21	2.85	5.70	EEF1G	P26641	51.802	-0.72
O	GRP75_HUMAN	2.02	1.83	1.53	HSPA9	P38646	76.139	0.46
b	HNRH1_HUMAN	1.92	2.04	2.43	HNRNPH1	P31943	50.184	-0.48
b	HNRPK_HUMAN	2.29	2.04	2.25	HNRNPK	P61978	52.188	0.62
b	HNRPM_HUMAN	2.80	1.93	2.54	HNRNPM	P52272	79.464	-1.21
N	HS90B_HUMAN	2.81	2.62	2.41	HSP90AB1	P08238	86.74	-1.20
N	K1C18_HUMAN	2.71	2.76	2.57	KRT18	P05783	48.939	-0.19
O	K1C19_HUMAN	2.42	1.59	1.75	KRT19	P08727	44.687	-0.39

b	K2C8_HUMAN	3.25	2.95	2.03	KRT8	P05787	55.044	-0.65
b	LG3BP_HUMAN	1.68	2.65	1.92	LGALS3BP	Q08380	67.028	0.07
N	LMAN2_HUMAN	1.55	2.50	2.01	LMAN2	Q12907	41.294	-0.31
O	MAP1B_HUMAN	3.47	3.18	4.14	MAP1B	P46821	281.429	1.22
b	PSD11_HUMAN	2.35	1.91	2.37	PSMD11	O00231	49.224	-0.24
N	PSMD6_HUMAN	2.39	2.68	2.10	PSMD6	Q15008	46.914	0.07
b	PTBP1_HUMAN	1.84	2.64	2.30	PTBP1	P26599	58.771	-1.67
O	RRBP1_HUMAN	1.87	2.12	3.07	RRBP1	Q9P2E9	160.42	0.16
O	RUVB1_HUMAN	1.55	2.00	1.78	RUVBL1	Q9Y265	52.295	-0.47
N	SC22B_HUMAN	2.22	1.73	4.70	SEC22B	O75396	25.518	0.53
N	SC24C_HUMAN	3.56	1.54	3.75	SEC24C	P53992	121.368	0.59

b	SEPT2_HUMAN	2.06	2.03	1.53	SEPT2	Q15019	42.81	-0.33
N	SFRS1_HUMAN	1.53	1.57	1.78	SRSF1	Q13243	32.108	-0.96
N	SFRS5_HUMAN	1.95	1.67	1.72	SRSF5	Q13243	32.108	-0.96
O	TAGL2_HUMAN	3.75	2.40	1.86	TAGLN2	P37802	23.081	-1.01
b	THY1_HUMAN	1.89	1.70	1.82	THY1	P04216	18.516	-0.02
N	TNPO1_HUMAN	2.43	1.80	2.24	TNPO1	Q92973	104.604	-0.17
N	U2AF2_HUMAN	2.08	4.20	2.62	U2AF2	P26368	55.02	0.87
b	YBOX1_HUMAN	4.02	2.97	2.53	YBX1	P67809	36.603	-1.05

b: found differentially expressed in breast cancer

o: found differentially expressed in cancers other than breast cancer

n: not found differentially expressed in any cancer

Table 4.4 The 76 proteins down-regulated in all three cases.

Cancer	Protein Name	Quantification ratio			Gene Name	AccessID	MW(kDa)	Gravy
		CT0018	MT1275	MT699				
O	A1BG_HUMAN	0.18	0.31	0.52	A1BG	P04217	55.305	0.52
O	A2MG_HUMAN	0.07	0.13	0.62	A2M	P01023	168.368	-0.50
O	ALBU_HUMAN	0.08	0.09	0.37	ALB	P02768	73.867	1.17
O	AMPN_HUMAN	0.04	0.44	0.32	ANPEP	P15144	111.795	0.06
O	ANT3_HUMAN	0.07	0.10	0.51	SERPINC1	P01008	54.608	0.58
b	ANXA1_HUMAN	0.45	0.34	0.42	ANXA1	P04083	40.291	-0.72
O	AOC3_HUMAN	0.33	0.09	0.16	AOC3	Q16853	85.755	-0.15

b	APOA2_HUMAN	0.06	0.36	0.58	APOA2	P02652	11.73	-0.25
b	APOB_HUMAN	0.07	0.24	0.42	APOB	P04114	531.702	-0.58
b	APOC3_HUMAN	0.04	0.13	0.30	APOC3	P02656	11.132	-0.64
b	APOD_HUMAN	0.07	0.49	0.53	APOD	P05090	22.037	-0.56
b	APOE_HUMAN	0.08	0.30	0.55	APOE	P02649	36.826	-0.11
N	B3AT_HUMAN	0.07	0.36	0.26	SLC4A1	P02730	103.259	0.49
b	BGH3_HUMAN	0.23	0.19	0.35	TGFBI	Q15582	76.682	-1.22
b	CAV1_HUMAN	0.60	0.65	0.51	CAV1	Q03135	21.168	-0.68
O	CBPA3_HUMAN	0.33	0.11	0.10	CPA3	P15088	50.537	0.75
b	CD36_HUMAN	0.03	0.16	0.33	CD36	P16671	54.884	-0.50
b	CERU_HUMAN	0.10	0.08	0.41	CP	P00450	125.79	-0.57

b	CO1A1_HUMAN	0.03	0.06	0.15	COL1A1	P02452	142.276	-1.06
b	CO1A2_HUMAN	0.02	0.02	0.08	COL1A2	P08123	131.858	-0.03
O	CO6A1_HUMAN	0.02	0.14	0.46	COL6A1	P12109	111.9	-1.03
O	CO6A2_HUMAN	0.02	0.15	0.46	COL6A2	P12110	111.965	0.00
O	CO6A3_HUMAN	0.05	0.16	0.45	COL6A3	P12111	351.837	0.79
N	CO6A6_HUMAN	0.01	0.13	0.33	COL6A6	A6NMZ7	255.088	1.17
O	COEA1_HUMAN	0.12	0.19	0.20	COL14A1	Q05707	198.12	0.07
b	COIA1_HUMAN	0.43	0.39	0.20	COL18A1	P39060	181.188	1.26
b	CRYAB_HUMAN	0.21	0.32	0.18	CRYAB	P02511	20.595	-0.83
O	EHD2_HUMAN	0.40	0.41	0.42	EHD2	Q9NZN4	62.925	-0.39
b	EMIL1_HUMAN	0.19	0.64	0.40	EMILIN1	Q9Y6C2	108.403	-0.39

b	ENOA_HUMAN	0.63	0.46	0.31	ENO1	P06733	49.106	0.33
b	FBN1_HUMAN	0.18	0.62	0.55	FBN1	P35555	337.375	-0.65
b	FETUA_HUMAN	0.05	0.15	0.48	AHSG	P02765	40.798	0.14
b	HBA_HUMAN	0.18	0.13	0.27	HBA1 & HBA2	P69905	15.801	-0.22
b	HBB_HUMAN	0.19	0.18	0.44	HBB	P68871	16.593	-0.75
b	HBD_HUMAN	0.14	0.11	0.20	HBD	P02042	16.656	-0.22
O	HEMO_HUMAN	0.10	0.11	0.36	HPX	P02790	53.301	-0.63
b	HPT_HUMAN	0.08	0.09	0.55	HP	P00738	47.36	0.50
O	IGHA1_HUMAN	0.10	0.24	0.46	IGHA1	P01876	39.06	-0.43
b	IGHG1_HUMAN	0.10	0.08	0.39	IGHG1	P01857	37.807	0.16
N	IGHG2_HUMAN	0.22	0.21	0.44	IGHG2	P01859	37.542	0.59

b	IGHG3_HUMAN	0.08	0.06	0.37	IGHG3	P01860	43.449	0.15
O	IGJ_HUMAN	0.17	0.39	0.23	IGJ	P01591	16.321	0.46
N	IGKC_HUMAN	0.09	0.17	0.33	IGKC	P01834	12.143	-2.08
b	K2C5_HUMAN	0.31	0.48	0.22	KRT5	P13647	63.821	-1.58
b	KV104_HUMAN	0.11	0.13	0.36	/	P01596	12.097	-0.26
b	KV119_HUMAN	0.03	0.13	0.28	/	P01611	11.917	-0.25
b	KV402_HUMAN	0.11	0.09	0.18	/	P01625	13.032	0.02
b	LAC_HUMAN	0.13	0.19	0.50	/	P0CG04	11.771	-0.31
b	LAMC1_HUMAN	0.22	0.36	0.31	LAMC1	P11047	186.629	0.41
b	LDHB_HUMAN	0.59	0.30	0.25	LDHB	P07195	38.027	1.44
b	LEG1_HUMAN	0.48	0.45	0.65	LGALS1	P09382	15.413	-1.60

b	LMNA_HUMAN	0.52	0.54	0.53	LMNA	P02545	76.089	-0.52
b	LUM_HUMAN	0.10	0.14	0.46	LUM	P51884	39.832	0.08
b	LV403_HUMAN	0.09	0.24	0.53	/	P01717	11.784	-0.64
b	MET7A_HUMAN	0.60	0.65	0.56	METTL7A	Q9H8H3	29.394	-0.46
b	MIME_HUMAN	0.02	0.05	0.26	OGN	P20774	35.412	-0.63
b	MUC18_HUMAN	0.17	0.26	0.13	MCAM	P43121	73.911	-0.12
b	MUCL1_HUMAN	0.07	0.08	0.04	MUCL1	Q96DR8	9.251	0.27
N	MYO1C_HUMAN	0.52	0.30	0.26	MYO1C	O00159	125.311	-0.16
O	NID1_HUMAN	0.18	0.27	0.25	NID1	P14543	140.464	-1.06
b	PGS2_HUMAN	0.05	0.23	0.62	DCN	P07585	41.227	-1.50
O	PIGR_HUMAN	0.11	0.63	0.23	PIGR	P01833	86.222	-0.57

b	PIP_HUMAN	0.04	0.49	0.60	PIP	Q9Y2Q5	13.851	0.28
N	PLIN_HUMAN	0.14	0.05	0.16	PLIN1	O60240	57.151	0.46
b	PTRF_HUMAN	0.26	0.24	0.19	PTRF	Q6NZI2	45.117	-0.84
b	S10A4_HUMAN	0.59	0.64	0.62	S100A4	P26447	12.481	0.24
b	SODE_HUMAN	0.10	0.19	0.22	SOD3	P08294	26.535	0.71
O	SPTA2_HUMAN	0.60	0.63	0.49	SPTAN1	Q13813	293.517	0.64
O	SPTB2_HUMAN	0.67	0.60	0.55	SPTBN1	Q01082	282.793	-0.22
b	SYUG_HUMAN	0.11	0.13	0.09	SNCG	O76070	13.987	-1.59
O	TARSH_HUMAN	0.20	0.39	0.66	ABI3BP	Q7Z7G0	122.901	-1.18
b	TENX_HUMAN	0.11	0.07	0.10	TNXB	P22105	476.249	-0.81
b	TRFE_HUMAN	0.09	0.07	0.28	TF	P02787	81.746	-1.09

b	TRFL_HUMAN	0.14	0.36	0.34	LTF	P02788	81.975	-1.18
b	VTDB_HUMAN	0.12	0.10	0.55	GC	P02774	56.403	-0.18
N	VTNC_HUMAN	0.14	0.31	0.65	VTN	P04004	55.938	-0.64

b: found differentially expressed in breast cancer

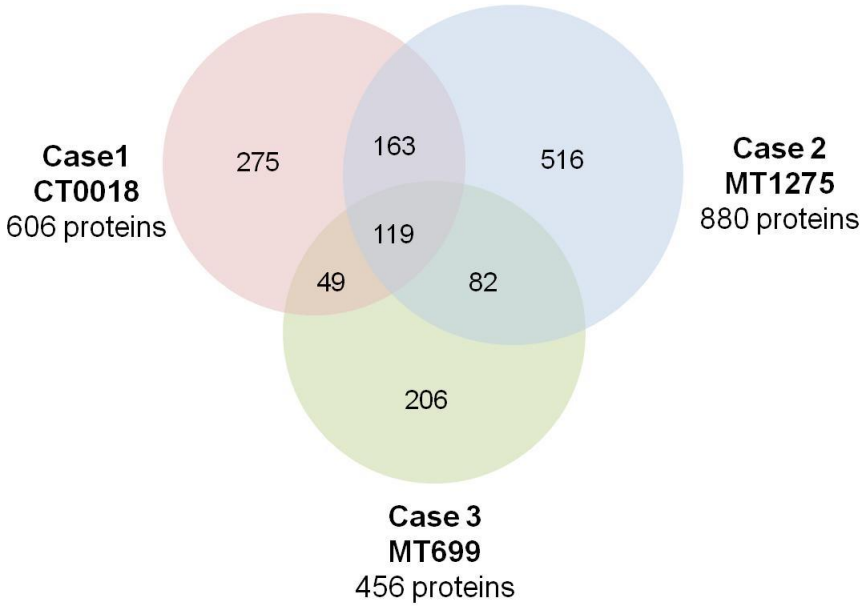
o: found differentially expressed in cancers other than breast cancer

n: not found differentially expressed in any cancer

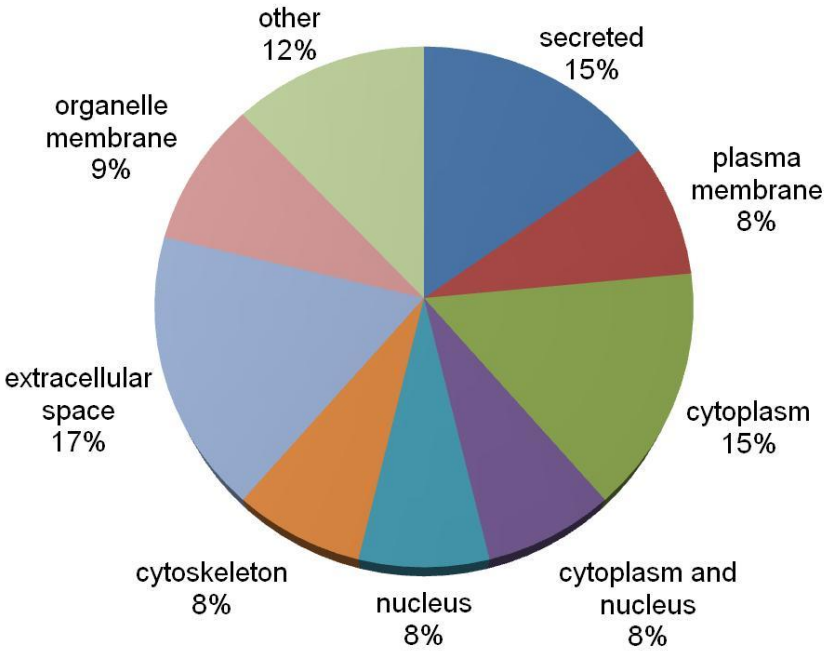
4.3.4 Bioinformatic Analysis

Figure 4.7b shows subcelluar locations of the common differentially expressed proteins. These proteins are mostly (17%) distributed in extracellular space, followed by the cytoplasm and secretion (15%). ~8% proteins are located in plasma membrane, cytoplasm and nucleus, nucleus, cytoskeleton and organelle membrane. The molecular functions of the common differentially expressed proteins are shown in Figure 4.7c. Major proteins (81%) have binding functions. 22% proteins have catalytic function, 7% have structural molecule activity, 6% have transporter activity, 2% have signal transducer activity and only 1% have motor activity. Many proteins have more than one function, e.g., protein AMPN_HUMAN has both binding and catalytic activity, protein ANXA1_HUMAN has binding and structural molecule activity, and APOD_HUAMN has binding and transporter activity. The biological processes related to cancer for the common differentially expressed proteins are shown in Figure 4.7d. No common proteins are associated with the metastasis process. Almost equal numbers of up-regulated proteins and down-regulated proteins are associated with the biological processes. In total, 8% in apoptosis, 5% in proliferation, 4% in differentiation and 2% in cell cycle.

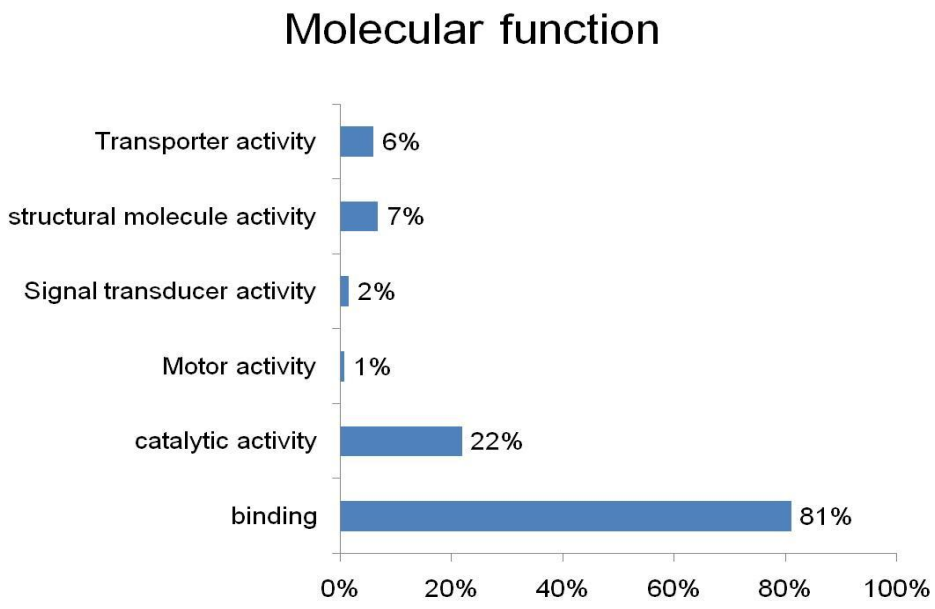
Pathways analysis results of the common differentially expressed proteins using Ingenuity pathways analysis (IPA) and Metacore pathways analysis (GeneGO) are illustrated in Figure 4.8 and figure 4.9 respectively. The protein-protein interaction networks were constructed using the identified common proteins from our experiment and other known or predicted proteins that are related to the


Case1
CT0018
606 proteins
275
163
516
Case 2
MT1275
880 proteins
119
49
82
206
Case 3
MT699
456 proteins

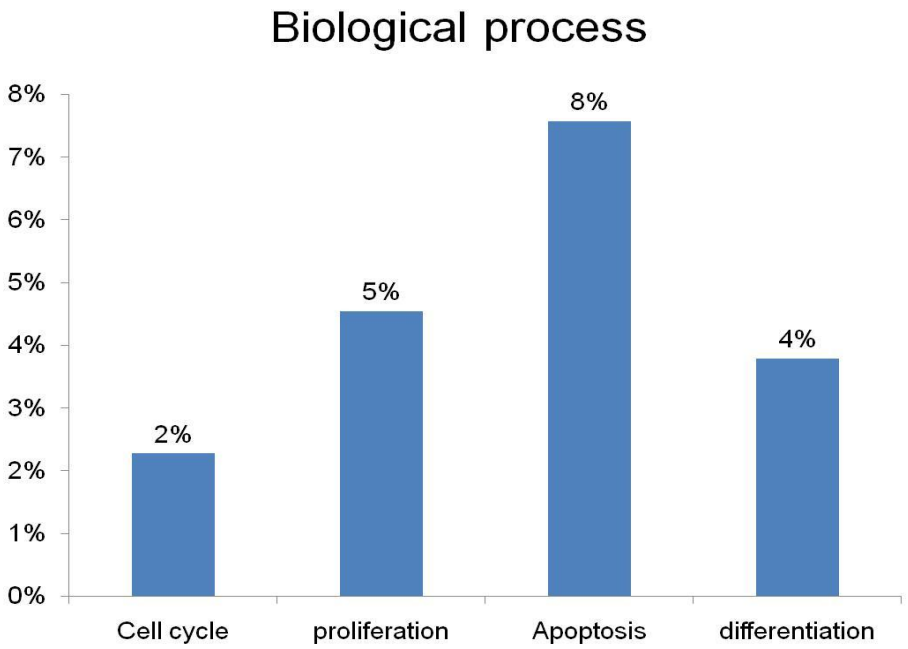
(a)


other
12%
secreted
15%
organelle
membrane
9%
plasma
membrane
8%
extracellular
space
17%
cytoplasm
15%
cytoskeleton
8%
nucleus
8%
cytoplasm and
nucleus
8%

(b)



(c)



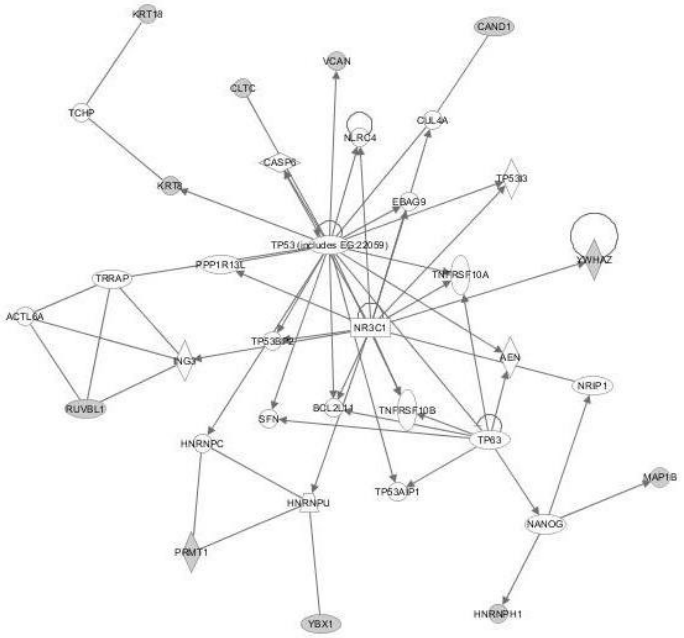
(d)

Figure 4.7 Figure showing (a) overlaps of differentially expressed proteins in three cases, and (b) subcelluar locations, (c) molecular functions and (d) biological processes of common differentially expressed proteins.

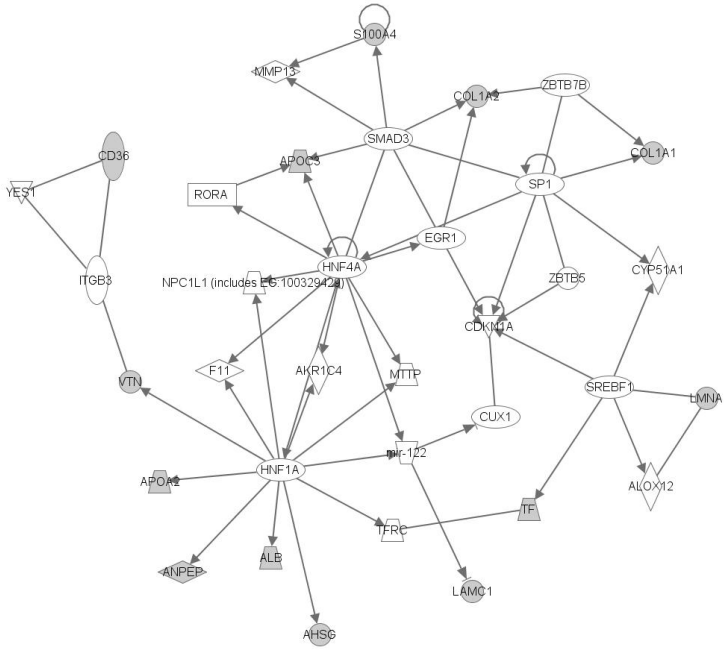
identified proteins. Some proteins are connected to well-known biomarkers inbreast cancer. ANXA1, CAV1, COL18A1, CP, DCN, FBN1, HBB, HP, MCAM, NID1, SNCG, SOD3 (gene names) are connected to EGFR (Epidermal growth factor receptor; ErbB-1; HER1)[30-32] as well as one of the most frequently mutated genes in breast cancer BRCA1[33-35] (Figure 4.8c). A2M, APOB, APOD, APOE, CRYAB, GC, LDHB, KRT5, and PIGR (gene names) are connected to EGFR and CD44[36-38] (Figure 4.8d). Protein hnRNP K (gene name: HNRPK), Keratin 8 (gene name: KRT8), GRP75 (gene name: HSPA9), and Calreticulin (gene name: CALR) are connected to EGFR and p53[39-41] (Figure 4.9a). hnRNP K, Versican (gene name: VCAN), PRMT1 (gene name: PRMT1), and COPG1 (gene name: COPG or COPG1) are connected to EGFR, p53 and CD44 (Figure 4.9b). Caveolin-1 (gene name: CAV1), Decorin (gene name: DCN), Alpha crystallin B (gene name: CRYAB), and Annexin I (gene name: ANXA1) are connected to EGFR, p53, and ErbB-2 (Human Epidermal Growth Factor Receptor 2; HER2/neu)[42-44] (Figure 4.9c). Caveolin-1, Fetuin-A (gene name: AHSG), Decorin, and PIGR (gene name: PIGR) are connected to EGFR and ErbB-2 (Figure 4.9d).

4.3.5 Biomarker Candidates for Verification

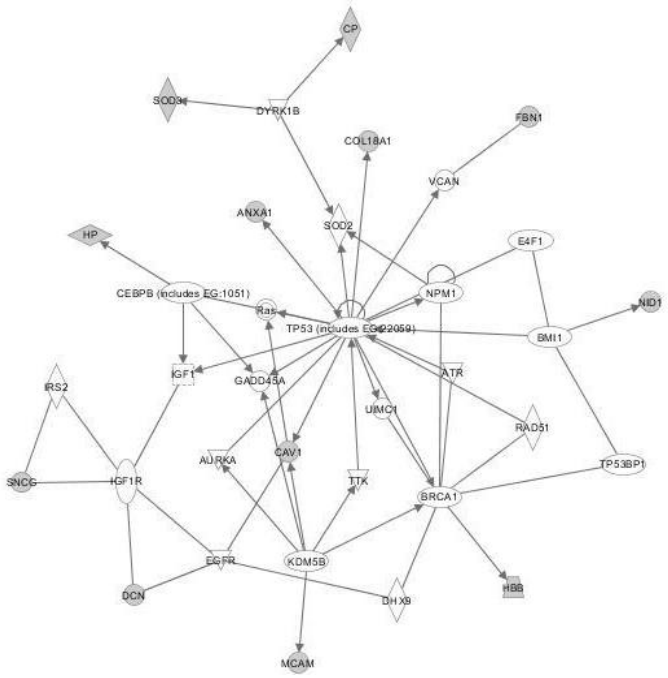
The common differentially expressed proteins can be considered as the putative biomarkers. Further verification needs to be conducted with a larger sample set of tissues from different breast cancer patients using immunohistochemistry. By searching through literatures, we picked three kinds of proteins for verification as potential biomarkers. The first kind of proteins, GRP75_HUMAN (gene name: HSPA9; up-regulated) and PIGR_HUAMN (gene name: PIGR; down-regulated),


KRT18
CAND1
VCAN
CLTC
TCHP
CUL4A
NLRC4
CASP8
KRT8
TP53I3
EBAG9
TP53 (includes EG:22059)
YWHAZ
PPP1R13L
TNFRSF10A
TRRAP
ACTL6A
NR3C1
TP53BP2
ING3
AEN
NRIP1
RUVBL1
SFN
BCL2L1
TNFRSF10B
TP63
HNRNPC
MAP1B
TP63AIP1
HNRNPU
NANOG
PRMT1
HNRNPH1
YBX1

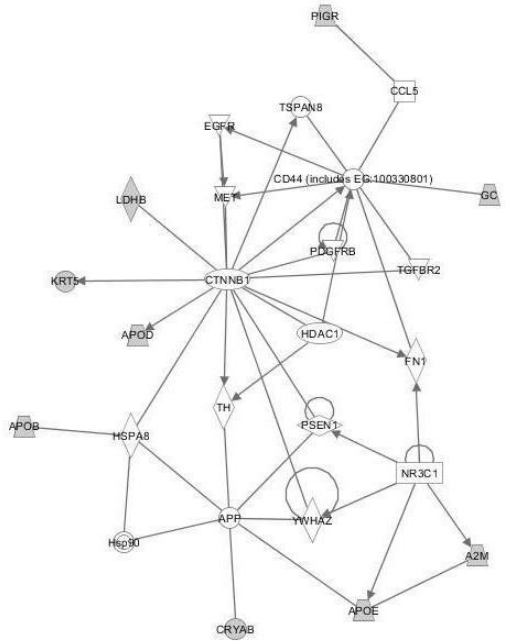
(a)


S100A4
MMP13
COL1A2
ZBTB7B
SMAD3
CD36
APOC3
COL1A1
YES1
RORA
SP1
EGR1
HNF4A
ITGB3
NPC1L1 (includes EG:100329428)
ZBTB5
CYP51A1
CDKN1A
F11
AKR1C4
MTTP
SREBF1
LMNA
VTN
CUX1
miR-122
HNF1A
APOA2
ALOX12
TF
TFRC
ANPEP
ALB
LAMC1
AHSG

(b)



(c)



(d)

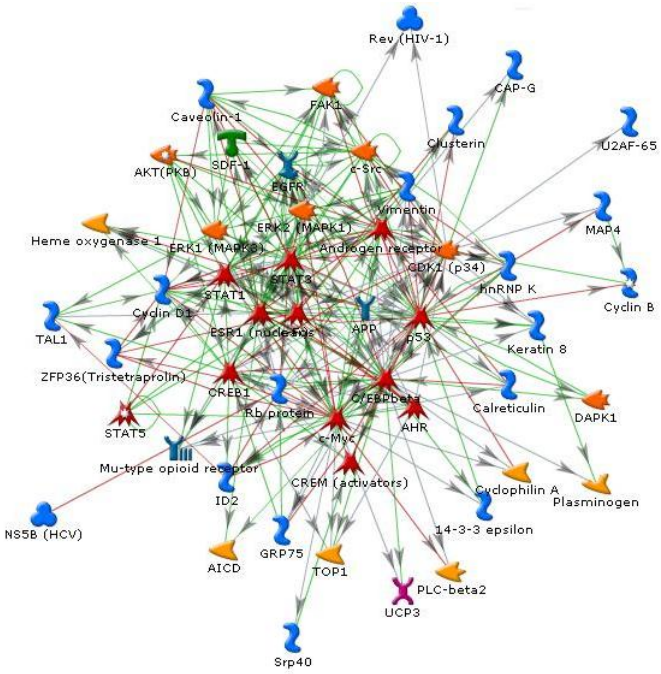
Figure 4.8 Protein-protein interaction networks of some (a) up-regulated proteins and (b) (c) (d) down-regulated proteins. The networks were generated through the use of IPA (Ingenuity® Systems, www.ingenuity.com). Proteins labeled by grey color are the identified common differentially expressed proteins from the comparison experiment.

are connected to well-known breast cancer biomarkers and found differentially expressed in many cancers other than breast cancer. The second kind of proteins, TAGL2_HUMAN (gene name: TAGLN2; up-regulated) and AOC3_HUAMN (gene name: AOC3; down-regulated), are not connected to well-known breast cancer biomarkers but are differentially expressed in many cancers other than breast cancer. The third kind of proteins, CALR_HUMAN (gene name: CALR; up-regulated), COPG_HUMAN (gene name: COPG; up-regulated), A2MG_HUAMN (gene name: A2M; down-regulated) and NID1_HUAMN (gene name: NID1; down-regulated), are connected to well-known breast cancer biomarkers and differentially expressed in only a few cancers other than breast cancer. These protein biomarker candidates should be of great interest for the next phase of research: using a large number of tissue samples for verification. This work will need to be conducted in the future in collaboration with cancer centers having well-characterized tumor and control samples.

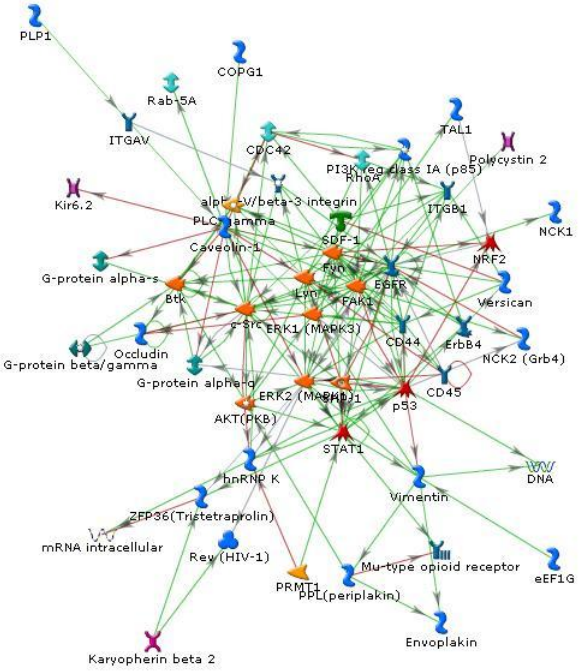
4.4 Conclusions

We have successfully applied the 2-MEGA labeling method combined with 2D LC-MS/MS for relative quantification of the proteomes of breast cancer tissues vs. pooled normal tissues.

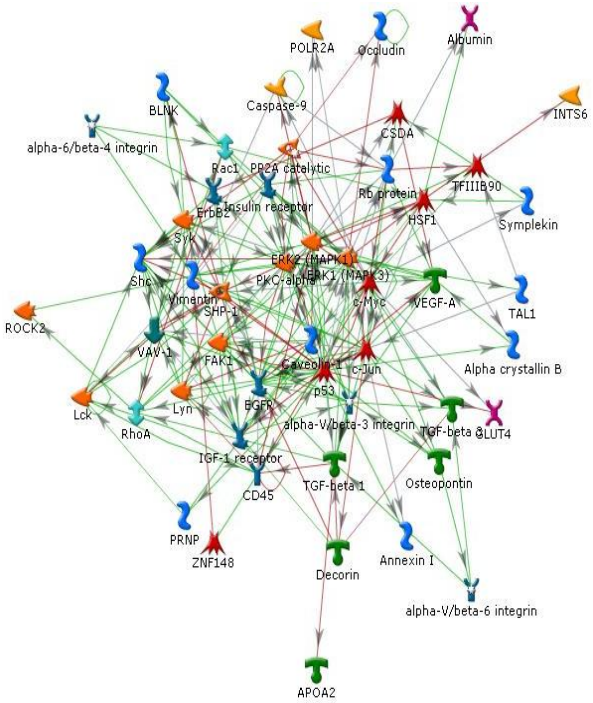
Acetone precipitation successfully removed lipids and salts in the samples. The acid labile detergent proteasMAX used to redissolve the protein pellets after acetone precipitation, had no interference with the LC ESI MS/MS analysis. The 2-MEGA labeling efficiencies were ~94% for the breast tumor tissues, similar to those obtained from other types of samples, such as breast cancer cell lines. Thus,


Rev (HIV-1)
CAP-G
Caveolin-1
FAK1
Clusterin
U2AF-65
AKT(PKB)
SDF-1
c-Src
EGFR
Vimentin
MAP4
Heme oxygenase 1
ERK2 (MAPK1)
ERK1 (MAPK3)
Androgen receptor
CDK1 (p34)
hnRNP K
Cyclin B
STAT3
STAT1
Cyclin D1
TAL1
ESR1 (nucleus)
APP
p53
Keratin 8
ZFP36(Tristetraprolin)
CREB1
C/EBPbeta
Calreticulin
DAPK1
Rb protein
AHR
STAT5
c-Myc
Mu-type opioid receptor
CREM (activators)
Cyclophilin A
Plasminogen
ID2
NS5B (HCV)
14-3-3 epsilon
GRP75
AICD
TOP1
PLC-beta2
UCP3
Srp40

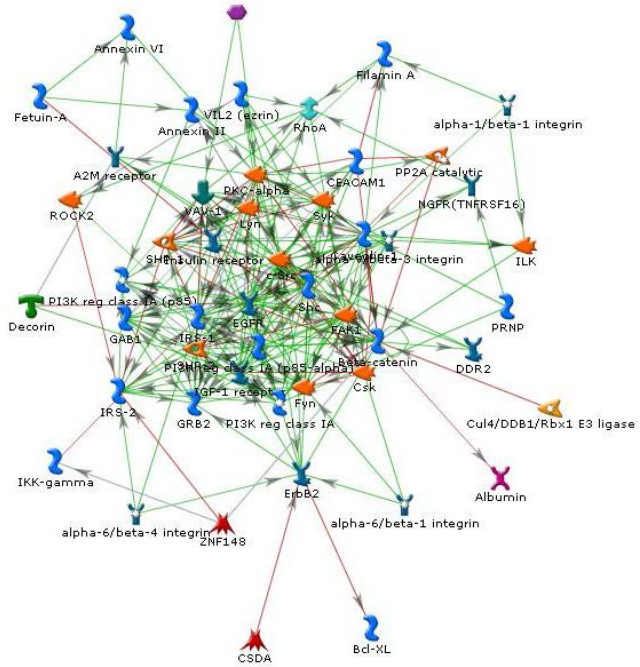
(a)


PLP1
COPG1
Rab-5A
TAL1
ITGAV
CDC42
Polycystin 2
PI3K reg class IA (p85)
RhoA
Kir6.2
alpha-V/beta-3 integrin
ITGB1
NCK1
PLC-gamma
SDF-1
Caveolin-1
NRF2
G-protein alpha-s
FYN
Btk
Lyn
EGFR
Versican
FAK1
c-Src
ERK1 (MAPK3)
CD44
ErbB4
Occludin
NCK2 (Grb4)
G-protein beta/gamma
G-protein alpha-q
CD45
ERK2 (MAPK1)
SHP-1
p53
AKT(PKB)
STAT1
hnRNP K
DNA
Vimentin
ZFP36(Tristetraprolin)
mRNA intracellular
Rev (HIV-1)
Mu-type opioid receptor
eEF1G
PRMT1
PPL(periplakin)
Envoplakin
Karyopherin beta 2

(b)



(c)



(d)

Figure 4.9 Protein-protein interaction networks of some (a) (b) up-regulated proteins and (c) (d) down-regulated proteins. The networks were generated through the use of Metacore program.

2-MEGA labeling could be used to quantify proteins of breast tissues.

In the comparison experiments, ~4000 to 8000 peptides and 1500 to 2500 proteins were identified for each 2-D LC MS/MS analysis in total. ~23% peptides in the comparison experiments were removed after we applied the threshold of relative difference of 0.77 from the forward and reverse labeling experiments. 606, 880 and 456 differentially expressed proteins with a quantification ratio of less than 0.67 or larger than 1.5 were found in tumor cases CT0018, MT1275 and MT699 respectively. 119 proteins were found differentially expressed in all three cases. These proteins are mostly (17%) distributed in extracellular space. 81% common proteins have binding functions. No common proteins are associated with metastasis process. Almost equal numbers of up-regulated proteins and down-regulated proteins are associated with the biological processes.

Ingenuity pathways analysis (IPA) and Metacore pathways analysis (GeneGO) were used to construct the protein-protein interaction networks of the common differentially expressed proteins. The results show some proteins are related to well-known biomarkers in breast cancer. Finally, several protein candidates were suggested for future verification work as potential biomarkers for breast cancer tumors.

4.5 Literature Cited

(1) Cole, K. D.; He, H. J.; Wang, L. *Proteomics. Clinical Applications* **2013**, *7*, 17.

(2) Fry, S. A.; Sinclair, J.; Timms, J. F.; Leathem, A. J.; Dwek, M. V. *Cancer Letters* **2013**, *328*, 335.
(3) Maxwell, C. *Clinical Journal of Oncology Nursing* **2010**, *14*, 771.
(4) Diamandis, E. P. *Mol Cell Proteomics* **2004**, *3*, 367.
(5) Zhang, J.; Wang, Y.; Yin, Q.; Zhang, W.; Zhang, T.; Niu, Y. *International Journal of Clinical and Experimental Pathology* **2013**, *6*, 1380.
(6) Bouchal, P.; Roumeliotis, T.; Hrstka, R.; Nenutil, R.; Vojtesek, B.; Garbis, S. D. *J Proteome Res* **2009**, *8*, 362.
(7) Simpson, R. J.; Bernhard, O. K.; Greening, D. W.; Moritz, R. L. *Current Opinion in Chemical Biology* **2008**, *12*, 72.
(8) Ilyin, S. E.; Belkowski, S. M.; Plata-Salaman, C. R. *Trends in Biotechnology* **2004**, *22*, 411.
(9) Kulasingam, V.; Diamandis, E. P. *Int J Cancer* **2008**, *123*, 2007.
(10) Umar, A.; Kang, H.; Timmermans, A. M.; Look, M. P.; Meijer-van Gelder, M. E.; den Bakker, M. A.; Jaitly, N.; Martens, J. W.; Luider, T. M.; Foekens, J. A.; Pasa-Tolic, L. *Mol Cell Proteomics* **2009**, *8*, 1278.
(11) Kondo, T. *BMB reports* **2008**, *41*, 626.
(12) He, Q. Y.; Chiu, J. F. *Journal of Cellular Biochemistry* **2003**, *89*, 868.
(13) Ji, C.; Guo, N.; Li, L. *Journal of Proteome Research* **2005**, *4*, 2099.
(14) Ji, C.; Lo, A.; Marcus, S.; Li, L. *J Proteome Res* **2006**, *5*, 2567.
(15) Ji, C.; Zhang, N.; Damaraju, S.; Damaraju, V. L.; Carpenter, P.; Cass, C. E.; Li, L. *Anal Chim Acta* **2007**, *585*, 219.
(16) Thierolf, M.; Hagmann, M. L.; Pfeffer, M.; Berntenis, N.; Wild, N.; Roessler, M.; Palme, S.; Karl, J.; Bodenmuller, H.; Ruschoff, J.; Rossol, S.; Rohr, G.; Rosch, W.; Friess, H.; Eickhoff, A.; Jauch, K. W.; Langen, H.; Zolg, W.; Tacke, M. *Proteomics. Clinical Applications* **2008**, *2*, 11.
(17) Chang, Z.; Yamada, R.; Suzuki, A.; Sawada, T.; Yoshino, S.; Tokuhiro, S.; Yamamoto, K. *Rheumatology* **2005**, *44*, 40.
(18) Boja, E. S.; Fales, H. M. *Anal Chem* **2001**, *73*, 3576.
(19) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Anal Chem* **2009**, *81*, 1049.
(20) Wang, N.; Li, L. *Anal Chem* **2008**, *80*, 4696.
(21) Gong, Y.; Wang, N.; Wu, F.; Cass, C. E.; Damaraju, S.; Mackey, J. R.; Li, L. *J Proteome Res* **2008**, *7*, 3583.
(22) Racine, J.; Gerber, V.; Feutz, M. M.; Riley, C. P.; Adamec, J.; Swinburne, J. E.; Couetil, L. L. *BMC Veterinary Research* **2011**, *7*, 48.
(23) Jimenez-Marin, A.; Collado-Romero, M.; Ramirez-Boo, M.; Arce, C.; Garrido, J. J. *BMC Proceedings* **2009**, *3 Suppl 4*, S6.
(24) Ekins, S.; Nikolsky, Y.; Bugrim, A.; Kirillov, E.; Nikolskaya, T. *Methods Mol Biol* **2007**, *356*, 319.
(25) Ekins, S.; Bugrim, A.; Brovold, L.; Kirillov, E.; Nikolsky, Y.; Rakhmatulin, E.; Sorokina, S.; Ryabov, A.; Serebryiskaya, T.; Melnikov, A.; Metz, J.; Nikolskaya, T. *Xenobiotica* **2006**, *36*, 877.
(26) Rhodes, D. R.; Tomlins, S. A.; Varambally, S.; Mahavisno, V.; Barrette, T.; Kalyana-Sundaram, S.; Ghosh, D.; Pandey, A.; Chinnaiyan, A. M. *Nat Biotechnol* **2005**, *23*, 951.
(27) Wang, P.; Lo, A.; Young, J. B.; Song, J. H.; Lai, R.; Kneteman, N. M.; Hao, C.; Li, L. *J Proteome Res* **2009**, *8*, 3403.

(28) Oppenheimer, S. R.; Mi, D.; Sanders, M. E.; Caprioli, R. M. *J Proteome Res* **2010**, *9*, 2182.
(29) Cottingham, K. *J Proteome Res* **2010**, *9*, 2049.
(30) Bhargava, R.; Gerald, W. L.; Li, A. R.; Pan, Q.; Lal, P.; Ladanyi, M.; Chen, B. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **2005**, *18*, 1027.
(31) Foley, J.; Nickerson, N. K.; Nam, S.; Allen, K. T.; Gilmore, J. L.; Nephew, K. P.; Riese, D. J., 2nd *Seminars in Cell & Developmental Biology* **2010**, *21*, 951.
(32) Liu, Z. B.; Hou, Y. F.; Zhu, J.; Hu, D. L.; Jin, W.; Ou, Z. L.; Di, G. H.; Wu, J.; Shen, Z. Z.; Shao, Z. M. *Oncogene* **2010**, *29*, 2996.
(33) Ford, D.; Easton, D. F.; Stratton, M.; Narod, S.; Goldgar, D.; Devilee, P.; Bishop, D. T.; Weber, B.; Lenoir, G.; Chang-Claude, J.; Sobol, H.; Teare, M. D.; Struewing, J.; Arason, A.; Scherneck, S.; Peto, J.; Rebbeck, T. R.; Tonin, P.; Neuhausen, S.; Barkardottir, R.; Eyfjord, J.; Lynch, H.; Ponder, B. A.; Gayther, S. A.; Zelada-Hedman, M.; et al. *American Journal of Human Genetics* **1998**, *62*, 676.
(34) Couch, F. J.; Weber, B. L. *Hum Mutat* **1996**, *8*, 8.
(35) King, M. C.; Marks, J. H.; Mandell, J. B. *Science* **2003**, *302*, 643.
(36) Al-Hajj, M.; Wicha, M. S.; Benito-Hernandez, A.; Morrison, S. J.; Clarke, M. F. *Proc Natl Acad Sci U S A* **2003**, *100*, 3983.
(37) Phillips, T. M.; McBride, W. H.; Pajonk, F. *Journal of the National Cancer Institute* **2006**, *98*, 1777.
(38) Leth-Larsen, R.; Terp, M. G.; Christensen, A. G.; Elias, D.; Kuhlwein, T.; Jensen, O. N.; Petersen, O. W.; Ditzel, H. J. *Mol Med* **2012**, *18*, 1109.
(39) Gasco, M.; Shami, S.; Crook, T. *Breast Cancer Research : BCR* **2002**, *4*, 70.
(40) Miller, L. D.; Smeds, J.; George, J.; Vega, V. B.; Vergara, L.; Ploner, A.; Pawitan, Y.; Hall, P.; Klaar, S.; Liu, E. T.; Bergh, J. *Proc Natl Acad Sci U S A* **2005**, *102*, 13550.
(41) Allred, D. C.; Clark, G. M.; Elledge, R.; Fuqua, S. A.; Brown, R. W.; Chamness, G. C.; Osborne, C. K.; McGuire, W. L. *Journal of the National Cancer Institute* **1993**, *85*, 200.
(42) Harari, D.; Yarden, Y. *Oncogene* **2000**, *19*, 6102.
(43) Moulder, S. L.; Yakes, F. M.; Muthuswamy, S. K.; Bianco, R.; Simpson, J. F.; Arteaga, C. L. *Cancer Research* **2001**, *61*, 8887.
(44) Revillion, F.; Bonneterre, J.; Peyrat, J. P. *Eur J Cancer* **1998**, *34*, 791.

Chapter 5

Microwave-Assisted Acid Hydrolysis (MAAH) of Proteins Electroeluted from Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

5.1 Introduction

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is a powerful tool for protein separation and identification in proteomics.[1] In particular, two-dimensional (2-D) SDS-PAGE has been widely used for the comprehensive analysis of protein composition in biological systems.[2] SDS-PAGE combined with sensitive mass spectrometry analysis is an important tool for protein characterization and proteome analysis.

Several methods have been used for analyzing proteins in a gel band by MS analysis. In-gel digestion of SDS-PAGE separated proteins is one of these methods.[3] Proteins can be digested in-gel either by enzymes like trypsin[4-18] or by chemicals like CNBr[13,19-22] and acids[23,24]. After proteins are digested to peptides, fragments of smaller size are more easily extracted out of the gel by solvents. However, it usually takes a long time to finish the in-gel digestion (e.g., overnight). There are reports of applying microwave irradiation for in-gel digestions to shorten the digestion time.[25-27]

To recover the intact gel-separated proteins for protein analysis, passive

extraction, electroblotting and electroelution can be used.[28] Passive extraction involves direct protein extraction with an acidic organic solvent.[29-34] The extract is salt and SDS free, but the extraction efficiency is rather low. Electroblotting transfers proteins from gels to membranes, e.g., poly (vinylidene fluoride) (PVDF)-based membranes.[35-38] Proteins can be analyzed after protein extraction from the membrane. Alternatively, on-membrane MALDI MS analysis can be performed after matrix solution is added to the membrane.[1] Proteins blotted onto membranes can be also subjected to *in situ* enzymatic or chemical digestion, followed by peptide extraction and MS analysis.[28,36,38,39] However, this approach is not efficient at recovering large fragments or intact proteins.[1] Electroelution, on the other hand, is a method where the proteins are transferred in an electric field from the gel into solution.[40,41] Although the protein recovery rate of this method can be as high as 90%, its disadvantages include: low sample throughput; contamination of the eluted protein with SDS, salts and other impurities; and protein cleavage or chemical modification during elution.[42] For MS analysis, the eluate must be purified to remove the salts and SDS accompaning with the proteins. Seelert and Krause wrote a comprehensive review on the elution of proteins from gels, including discussion of different electroelution devices.[43]

Our research objective was to develop improved analytical tools for detailed characterization of intact proteins including sequence confirmation and analysis of protein modifications. In this work, we chose one commercially available electroelution apparatus to electroelute gel-separated proteins for terminal peptide analysis with the microwave-assisted acid hydrolysis (MAAH) MS method. We studied the effects of salts and SDS in purification thoroughly and found

equilibrium dialysis followed by acetone precipitation could effectively remove the salts and SDS. Terminal peptide analysis was carried out successfully by MAAH combined with LC-ESI MS/MS analysis of electroeluted proteins.

5.2 Experimental

5.2.1 Chemicals and Reagents

Dithiothreitol (DTT), ammonium bicarbaonate (NH_4HCO_3), β-mercaptoethnanol, sodium dodecyl sulfate (SDS), glycerol, bromophenol blue, tris(hydroxymethyl)aminomethane (tris base), glacial acetic acid, trifluoroacetic acid (TFA), methanol, LC-MS grade formic acid (FA), acetone, bovine serum albumin (BSA) and myoglobin protein standard were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). LC-MS grade water and acetonitrile (ACN) were from Fisher Scientific Canada (Edmonton, AB, Canada). ACS grade 37% HCl was from Merck (KGaA, Darmstadt, Germany).

5.2.2 SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE was carried out in a Bio-Rad mini-PROTEAN 3 system using 1 mm thick 12% Mini-PROTEAN® TGX™ precast gel of 10 wells of 50 µL from Bio-rad (Mississauga, ON, Canada). The sample was mixed with the same volume of glycine loading buffer (62.5 mM tris base, pH 6.8, 5% β-mercaptoethnanol (v/v), 2% SDS, trace amount bromphenol blue, 12.5% glycerol), heated at 95 ℃ for 5 min, cooled to room temperature, and loaded onto the gel. For the molecular weight markers, the Precision Plus Protein™ All Blue Standard of 10 – 25 kDa

from Bio-Rad (Mississauga, ON, Canada) was used. The electrophoresis was performed at 50 mA. The gel was fixed for 30 min in 10% acetic acid /40% methanol and then stained with Coomassie Brilliant Blue G-250 (Bio-Rad) for one hour. The gel was destained in water for two hours. For copper stain, the gel was washed in fresh distilled, deionized water (DDI water) for 5 min, stained in diluted Copper Stain from Bio-Rad and washed in fresh DDI water for 3 min. The protein bands were excised, transferred to a 1.5 mL polyethylene tube with 1 mL water and stored in -80 ℃ freezer prior to electroelution.

5.2.3 Electroelution of Proteins from Gel Bands

An electroeluter, Bio-Rad Model 422 Electro-Eluter (Bio-Rad), shown in Figure 5.1, was used for electroelution. The procedures operated were similar as what Schuhmacher *et al*. described.[39] In general, the stained protein bands were placed into one elution tube. Membrane caps with a molecular weight cut-off of 3.5 kDa were used and a current of 9 mA/elution tube was applied. The elution buffer was 50 mM ammonium bicarbonate (pH 8.0) containing 0.1% SDS. Elution was performed for about 2 h until the blue gel bands were completely colorless. After electroelution, the elution buffer was replaced with fresh DDI water without SDS, and equilibrium dialysis was performed overnight. Eluted proteins were collected in a volume of approximately 400 μL of buffer in the space between the frit and the bottom membrane. The collected protein solution was transferred into a 1.5 mL polyethylene tube. 200 μL fresh DDI water was used to rinse the membrane cap twice. The rinsed Proteins migrate out of the gel slices with the electrical current, through the frit, solution was also transferred to the same tube with the

(a)

(b)

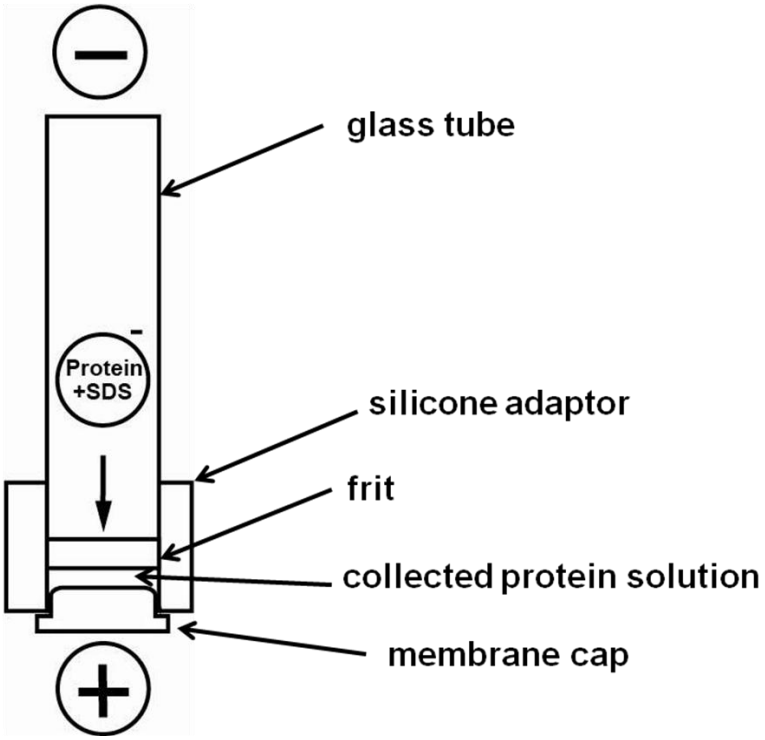


Figure 5.1 (a) The picture of Bio-Rad Model 422 Electro-Eluter. (b) Schematic view of electroelution apparatus. The glass tube is filled with elution buffer. go into the membrane cap and are retained by the dialysis membrane of the cap.

eluted protein solution.

5.2.4 LC Desalting and Quantification of Protein Elution

The collected protein solution was concentrated to about 200 μL in a SpeedVac vacuum centrifuge, followed by acetone precipitation at -20 °C overnight. The protein pellets were rinsed with 400 μL of -20 °C precooled acetone twice. After evaporation of the residue acetone, the protein pellets were reconstituted with 100 μL 0.1% TFA, and LC desalting and quantification were carried out on a 3 mm i.d. × 15 cm Zorbax 300 - SB C_3 column with a particle size of 3.5 μm and 300 Å pore (Agilent, Mississauga, ON, Canada) in an Agilent 1100 HPLC system (Palo Alto, CA, USA) at room temperature. After loading of the acidified protein sample, the column was flushed with 97.5% mobile phase A (0.1% TFA in water) and 2.5% mobile phase B (0.1% TFA in ACN) at a flow rate of 0.4 mL/min for 5 min and the residue salts were effectively removed. Subsequently, the concentration of phase B in the mobile phase was step-wise increased to 85% for 5 min to elute the proteins from the column, followed by 15 min re-equilibration with 97.5% mobile phase A. Quantification was performed by referring the peak area detected by UV to the previously established external calibration curve.

5.2.5 MALDI-TOF MS analysis

The collected fractions were concentrated using a SpeedVac, then mixed with the matrix solution of α-cyano-4-hydrocynnamic acid (CHCA) and spotted on a MALDI plate for MALDI analysis by an AB SCIEX 4800 Plus MALDI TOF/TOF Analyzer (AB SCIEX, Foster City, CA, USA) equipped with a diode-

pumped Nd:YAG laser at 355 nm. A linear or reflector mode was operated for MALDI-TOF mass spectra collection. The peak picking software in the 4800 Plus system was used for automatic peptide ion peak picking and mass assignment.

5.2.6 Microwave-Assisted Acid Hydrolysis of Proteins

The dried LC elution of BSA was reconstituted with 80 μL 3 mM DTT in 3M HCl in a 1.5 mL polypropylene vial. The vial was capped and then placed inside a household microwave oven (Panasonic, London Drugs, Edmonton) with 1200 W output at 2,450 MHz. A container with 100 mL of water was placed beside the sample vial to absorb the excess microwave energy. The sample was subjected to 60 s microwave irritation, then cooled and dried in a SpeedVac vacuum centrifuge to remove all the acid. 30 μL of 250 mM NH_4HCO_3 was used to reconstitute the sample. 1 μL 500 mM DTT was added and reduction of the disulfide bond was performed by incubation for 60 min at 37 ℃. Then the peptide solution was acidified to pH 2 with 50% TFA and diluted to 105 μL with 0.1% TFA.

5.2.7 LC Desalting or Fractionation of Hydrolysates

Desalting or fractionation was carried out on the 3 mm i.d. × 15 cm Zorbax 300 - SB C_3 column with a particle size of 3.5 μm and 300 Å pore in an Agilent 1100 HPLC system at room temperature. After loading of 100 μL acidified polypeptide sample, the column was flushed with 97.5% mobile phase A (0.1% TFA in water) and 2.5% mobile phase B (0.1% TFA in ACN) at a flow rate of 0.4 mL/min for 5 min and the salts were effectively removed. Subsequently, the concentration of phase B in the mobile phase was step-wise increased to 30% for 9 min to elute the low molecular weight polypeptides from the column for LC ESI MS/MS analysis,

or increased to 85% for 5 min to elute all peptides from the column for MALDI-TOF MS analysis, followed by 15 min re-equilibration with 97.5% mobile phase A.

5.2.8 LC ESI MS/MS analysis

The collected fractions were dried in the SpeedVac vacuum centrifuge to remove all the solvents and acid, reconstituted with 0.1 % formic acid and analyzed by a quadrupole time-of-flight (QTOF) Premier mass spectrometer (Waters, Manchester, U.K.) equipped with a nanoACQUITY Ultra Performance LC system (Waters, Milford, MA, USA). In brief, 5 μL of peptide solution was injected onto a 75 μm i.d. × 150 mm Atlantis dC_{18} column with 3 μm particle size (Waters). Solvent A consisted of 0.1% formic acid in water, and Solvent B consisted of 0.1% formic acid in ACN. Peptides were separated at 35 ℃ using 15 min gradients (2%–7% Solvent B for 1 min, 7%–25% Solvent B for 3 min, 25%–50% Solvent B for 2 min, 50%–90% Solvent B for 1 min, 90%–95% Solvent B for 8 min; the column was pre-equilibrated at 2% Solvent B for 20 min) and electrosprayed into the mass spectrometer (fitted with a nanoLockSpray source) at a flow rate of 350 nL/min. Mass spectra were acquired from m/z 400 to 1600 for 0.8 s, followed by six data-dependent MS/MS scans from m/z 50–1900 for 0.8 s each. The collision energy used to perform MS/MS was varied according to the mass and charge state of the eluting peptide. Leucine enkephalin and (Glu1)-fibrinopeptide B, a mixed mass calibrant (i.e., lock-mass), was infused at a rate of 300nL/min, and an MS scan was acquired for 1 s every 1 min throughout the run.[44]

5.2.9 Protein Database Search

Raw MS and MS/MS data were lock-mass-corrected, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.3 (Waters). Peptide sequences were identified via automated database searching of peak list files using the MASCOT search program (http://www.matrixscience.com). Database searches were restricted to the protein sequence of the corresponding protein downloaded from the SwissProt database. The following search parameters were selected for all database searching: enzyme, nonspecified; missed cleavages, 0; peptide tolerance, ±30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, (1+, 2+, and 3+); variable modifications, deamidation of asparagine and glutamine. The search results, including unique peptide sequences, ion score, MASCOT threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses were extracted to Excel files. All the identified peptides with scores lower than the MASCOT identity threshold scores for identity were then deleted from the list.

5.3 Results and discussion

5.3.1 Method Development

5.3.1.1 SDS-PAGE analysis

Figure 5.2 shows the SDS-PAGE image of 40 μg BSA protein. When the protein loading amount was as large as 40 μg, some impurities accompanied with the

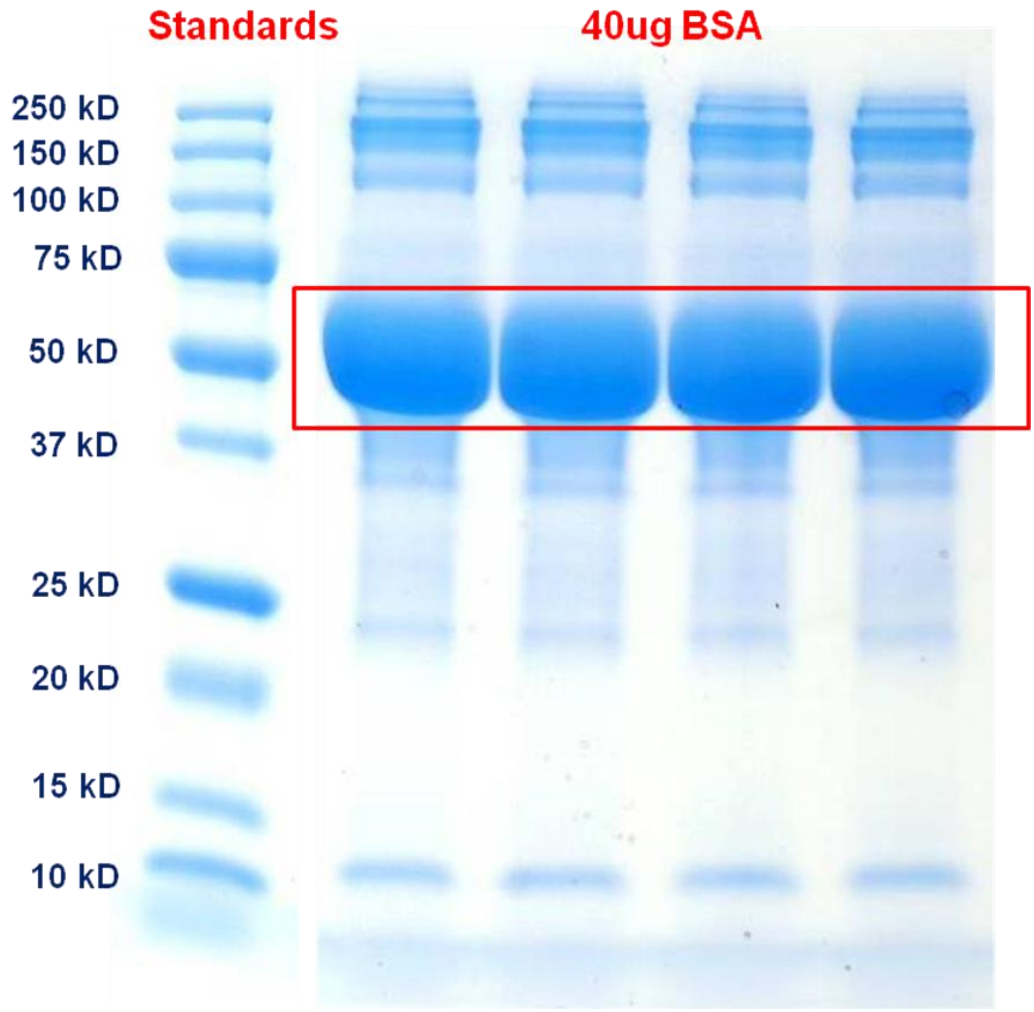


Figure 5.2 The SDS-PAGE image of 40 μg BSA protein. The red rectangle shows the gel bands where the BSA protein is located.

protein could be seen clearly in SDS-PAGE. Only the dark protein bands were cut for electroelution.

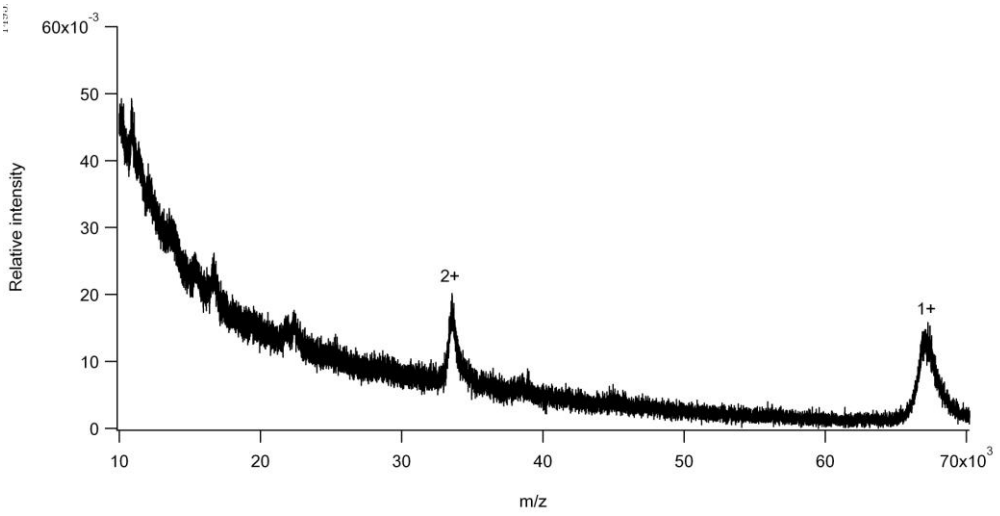
5.3.1.2 Removal of Ammonium Bicarbonate (NH_4HCO_3)

At first, electroelution was followed by electrodialysis. Then the collected protein solution was concentrated in the SpeedVac at room temperature. However, there were always white solids present at the tube bottom even when we dried the eluate overnight. These were likely the salts and SDS powders. We then performed acetone precipitation to remove the Coomassie blue dye and SDS and in the meantime, the salts would precipitate out with SDS.

To remove ammonium bicarbonate, we heated the dried eluate at 75°C. The white salts were all gone after drying for half an hour. The proteins were redissolved in fresh DDI water, followed by acetone precipitation and MALDI-TOF MS analysis. But the MALDI spectrum, shown in Figure 5.3, had poor signals.

In order to figure out the problems, we prepared a standard BSA protein sample at the same concentration as that in the elution buffer, i.e., 0.1% SDS in 50 mM ammonium bicarbonate, to mimic the protein elution. We tried to heat the sample at 75 °C after drying down the sample solutions to remove residue ammonium bicarbonate in the tube, or dry the elution at different temperatures in SpeedVac, i.e., 65 °C, 43 °C and room temperature. Figure 5.4 shows the results. Surprisingly, the spectra of BSA with drying down at different temperatures were as good as that of the standard BSA protein. Although the spectrum of BSA with

(a)



(b)

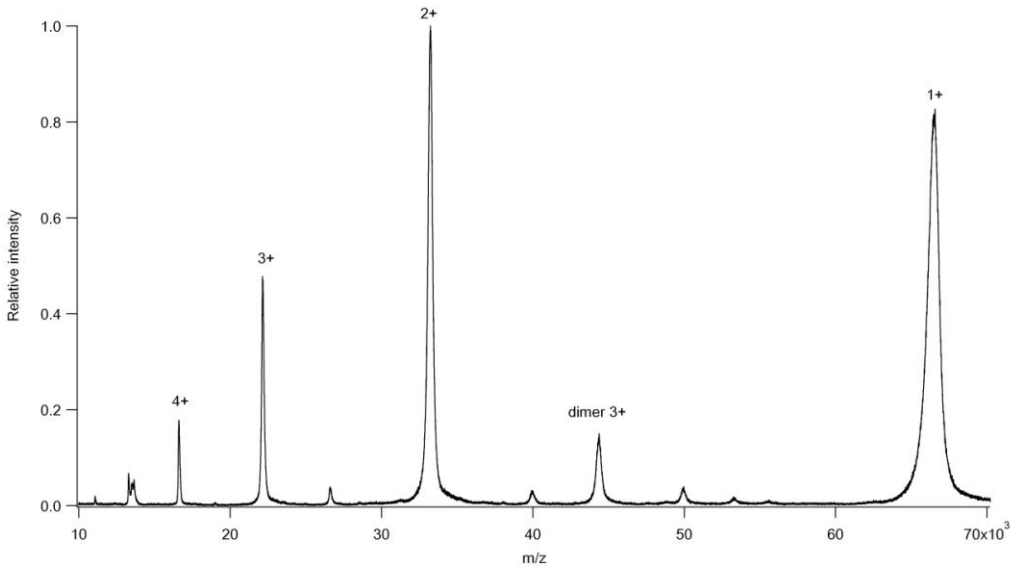
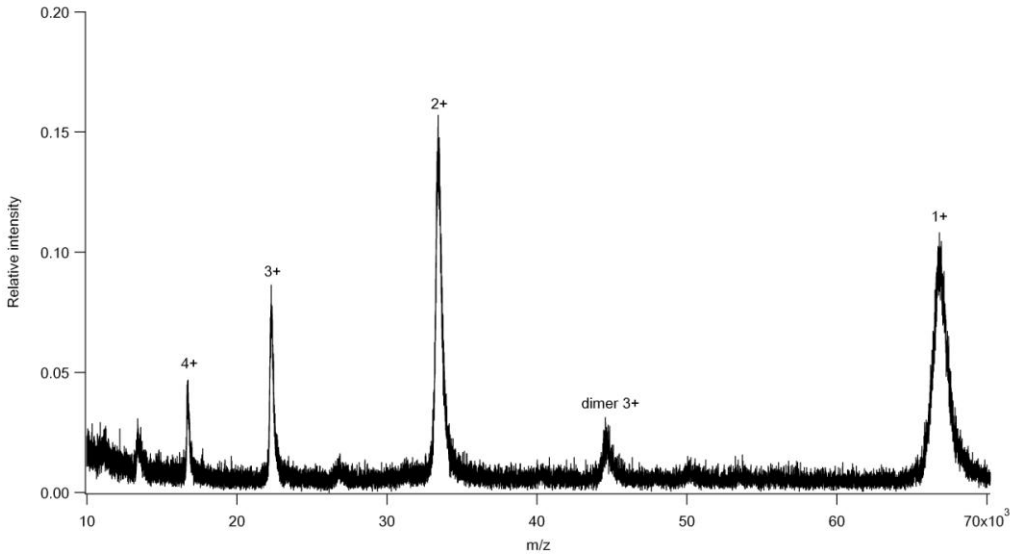
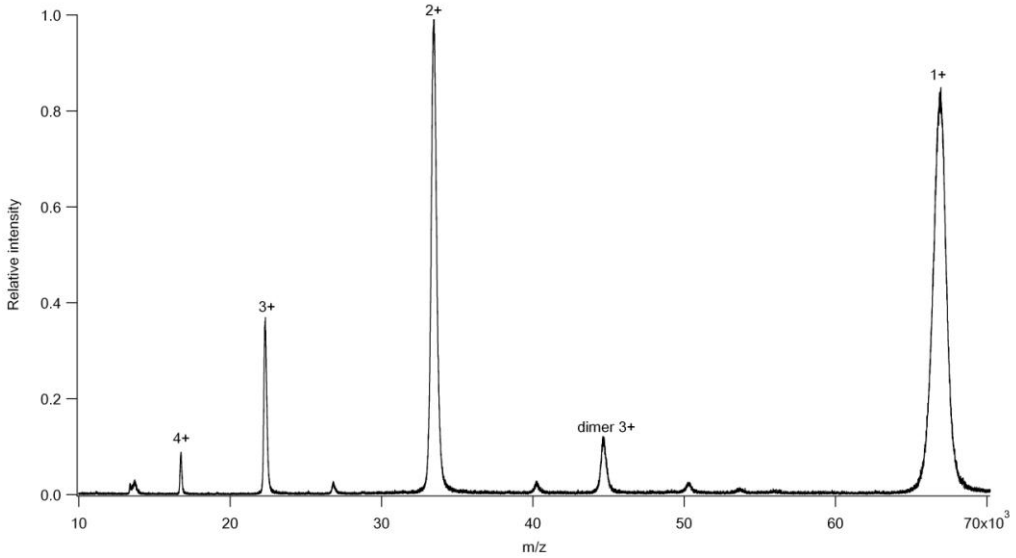


Figure 5.3 The spectra of (a) electroeluted BSA from gel, with 75°C heating to remove ammonium bicarbonate and (b) standard BSA protein.

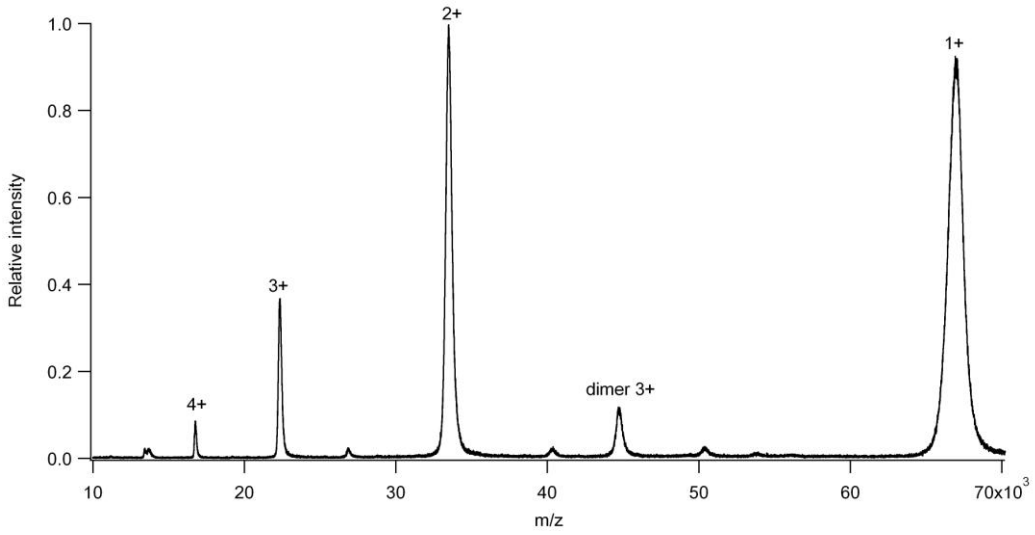
(a)


0.20
0.15
0.10
0.05
0.00
Relative intensity
4+
3+
2+
dimer 3+
1+
10
20
30
40
50
60
70x10^3
m/z

(b)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
4+
3+
2+
dimer 3+
1+
10
20
30
40
50
60
70x10^3
m/z

(c)



(d)

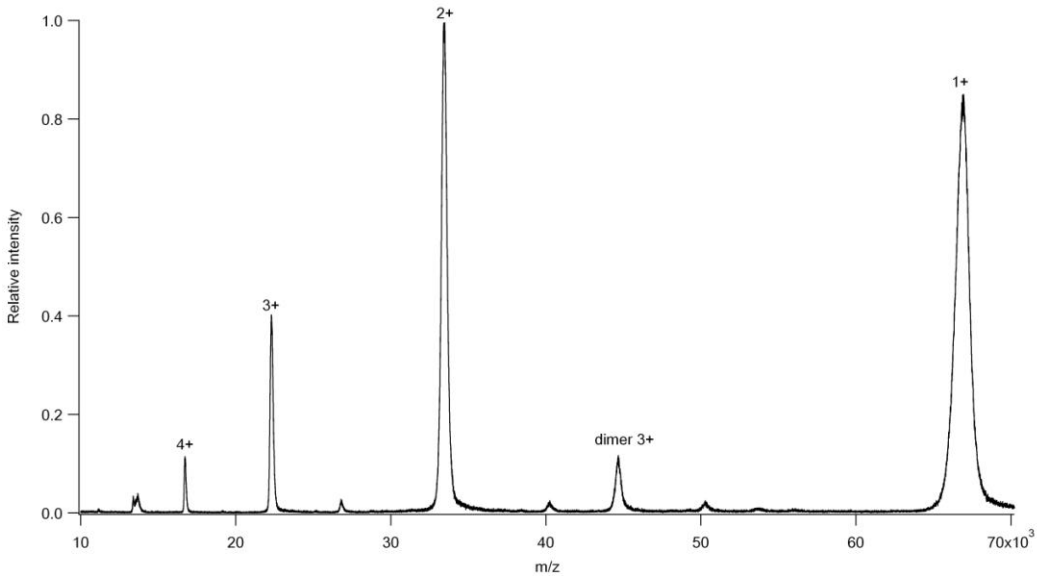


Figure 5.4 The spectra of standard BSA protein in 0.1% SDS in 50 mM ammonium bicarbonate after purification. (a) with 75 °C heating to remove ammonium bicarbonate after drying down, (b) with drying down at 65 °C in SpeedVac, (c) with drying down at 43 °C in SpeedVac, and (d) with drying down at room temperature in SpeedVac.

heating at 75 °C had poorer quality than others, it was still better than the elution that had undergone the same purification procedures. Heating as high as 75°C might make proteins bind with SDS more tightly so that SDS could not be removed in acetone precipitation.

Notably drying the standard BSA solution containing SDS and NH_4HCO_3 at room temperature overnight was able to remove NH_4HCO_3 in the solution, while it was not the case for the protein eluate. We thought the only difference was the presence of the dye, Coomassie blue. So we tried copper stain/destain after running SDS-PAGE, then performed electroelution and collected the protein elution. However, it turned out overnight drying at room temperature in SpeedVac still could not remove all of the NH_4HCO_3 in the eluate.

Heating could help with the evaporation of NH_4HCO_3 effectively. So we chose to dry the protein elution at 65 °C in SpeedVac to remove NH_4HCO_3 in the eluate. Figure 5.5 shows that the spectrum was better than before, but the peaks were broaden and the baseline was wavy, clearly indicating the presence of SDS.

5.3.1.3 Removal of Sodium Dodecyl Sulfate (SDS)

After electrodialysis, many bubbles appeared in the space between the frit and the membrane cap when collecting the protein eluate, demonstrating that SDS was not removed by electrodialysis completely.

We suspected that the SDS concentration in the eluate was not the same as that in the elution buffer. Thus, there were still SDS residues left after acetone precipitation.

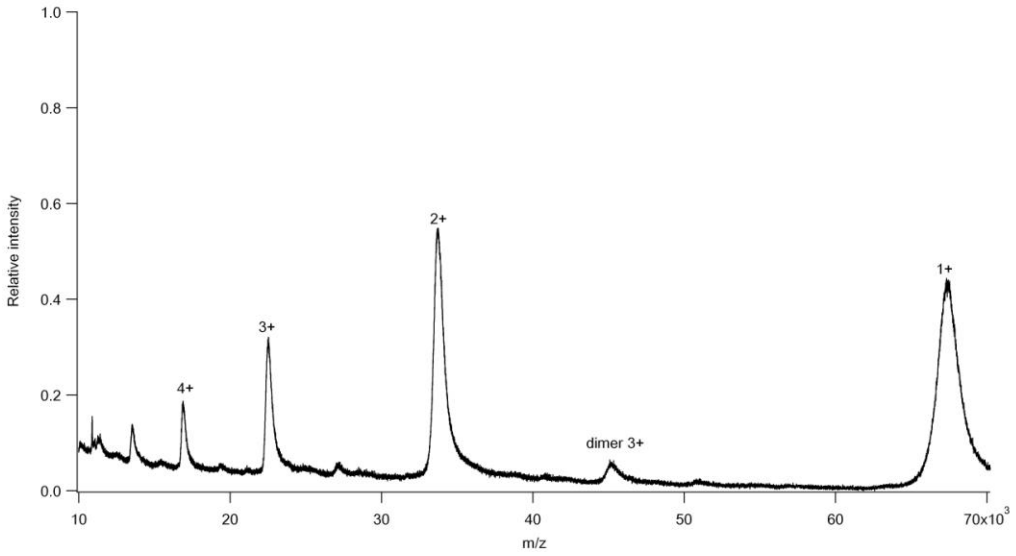


Figure 5.5 The spectrum of electroeluted BSA from gel, with drying at 65 °C in SpeedVac to remove ammonium bicarbonate.

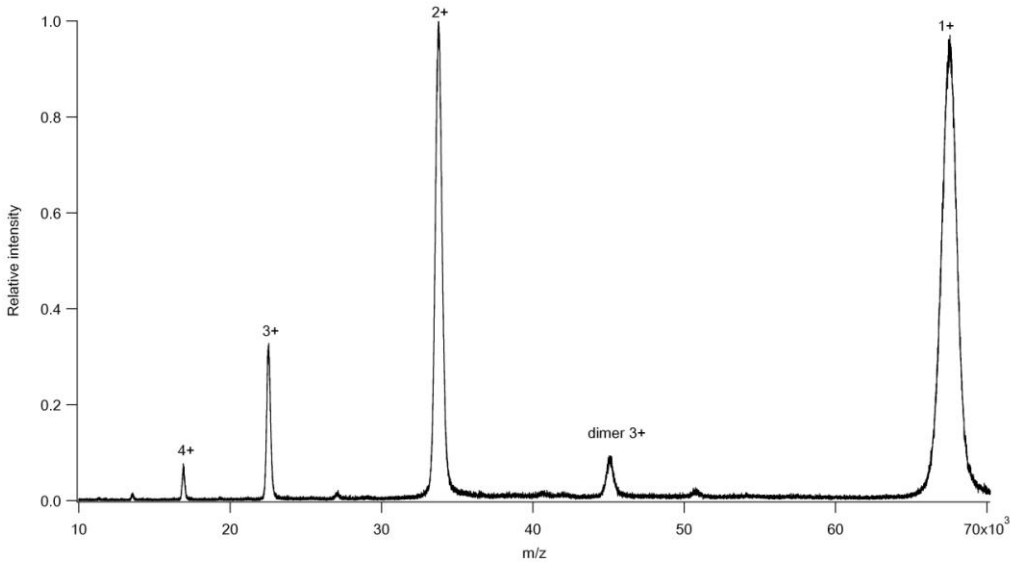


Figure 5.6 The spectrum of electroeluted BSA from gel, with equilibrium dialysis to remove ammonium bicarbonate.

To determine the concentration of SDS in the eluate, we concentrated the protein elution from 800 μL to 200 μL, and performed GC-MS analysis using the method described by Campbell *et al*.[45] The GC-MS results indicated that the SDS concentration with or without electrodialysis was about 2% to 3% in 200 μL concentrated eluate, compared to 0.5% to 0.75% in the original protein elution. The critical micelle concentration (CMC) of SDS in pure water at 25 °C is 8 mM (0.24%).[46] Furthermore, the CMC decreases in the presence of salts.[47,48] So actually SDS continuously moved to the space between the frit and the membrane cap with current and formed micelles there. After the formation of micelles, more and more SDS and salts were retained in the space. The real concentration of NH_4HCO_3 in the protein eluate was unknown, but it should be much higher than its concentration in the elution buffer which is possibly retained with the concentrated SDS. This was why the standard protein prepared in the elution buffer could evaporate all NH_4HCO_3 in the solution at room temperature in SpeedVac, while the protein eluate could not.

Now the question became how to remove SDS as well as the concentrated salts in the protein eluate. Heating might make the concentrated SDS difficult to remove, as demonstrated in Figure 5.5. Thus we stopped using heating to remove NH_4HCO_3.

Initially we concentrated the protein elution and then performed acetone precipitation. However, too much salt would precipitate out during acetone precipitation. To reduce the salt and SDS concentration, we tried to avoid concentrating the protein elution. Instead, we aliquoted the protein eluate to

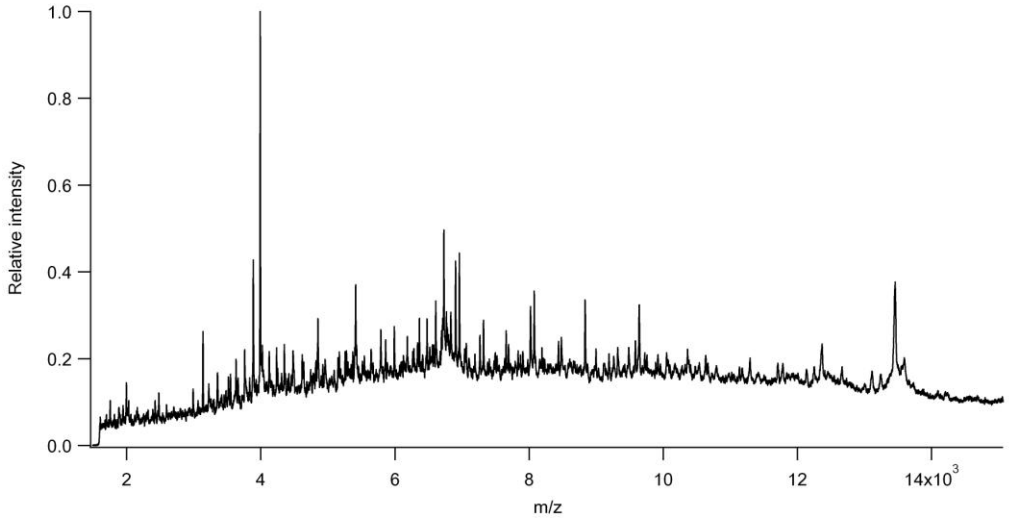
several tubes and performed acetone precipitation in each tube. However, the recovery of protein was rather poor. The protein amount in each tube was so small after split that the protein loss in the tube was severe. Therefore, it was necessary to concentrate the protein elution while removing SDS in the presence of concentrated salts.

We examined the ability of removing SDS using another classic SDS removal method, mehthanol/chloroform/water protein precipitation (CMW precipitation method), based on the protocol described by Wessel and Fugge.[49] The quantification of the protein amount after CMW precipitation was done by LC-UV analysis. For 20 μg BSA in DDI water, the protein recovery of CMW precipitation was around 65%. However, if we added 50 mM NH_4HCO_3 or 4% SDS to the protein solution, the recovery dropped to about 25%. Thus CMW precipitation was not effective for handling our protein eluate containing concentrated salts and SDS.

Next, we tested the protein recovery of acetone precipitation. For the same 20 μg BSA in DDI water, no proteins were recovered. However, for BSA in 50 mM NH_4HCO_3 or 4% SDS, the recovery was up to around 80%. Thus, acetone precipitation was a good way to remove the concentrated SDS.

At last, we chose to perform equilibrium dialysis overnight with the same eletroelution apparatus. Salts were effectively removed, but SDS still existed in the protein elution, as indicated by bubbles in the solution. No salts precipitated out in the acetone precipitation step and thus SDS removal was successful. The recovery was around 25% for 40 μg BSA loaded into the gel. Considering the

(a)



(b)

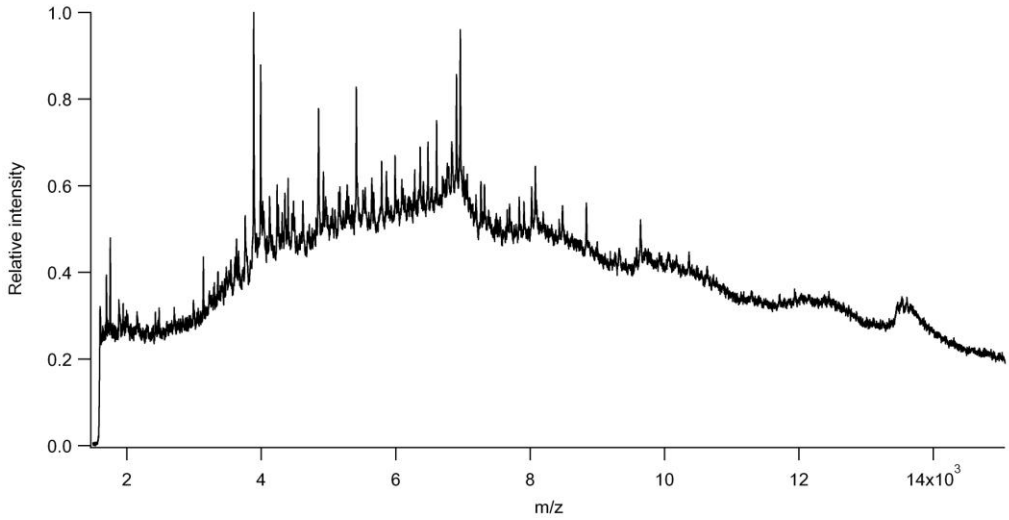


Figure 5.7 The spectra of BSA hydrolysates generated by MAAH for (a) standard BSA protein and (b) purified BSA protein electroeluted from SDS-PAGE.

high recovery rate in the acetone precipitation, we believe that most proteins were lost in the electroelution and dialysis process. Figure 5.6 shows the MALDI mass spectrum of electroeluted BSA with equilibrium dialysis and acetone precipitation purification. This spectrum is similar to the spectrum of the standard BSA protein shown in Figure 5.3b. The baseline was still a little wavy, indicating the presence of some residue SDS.

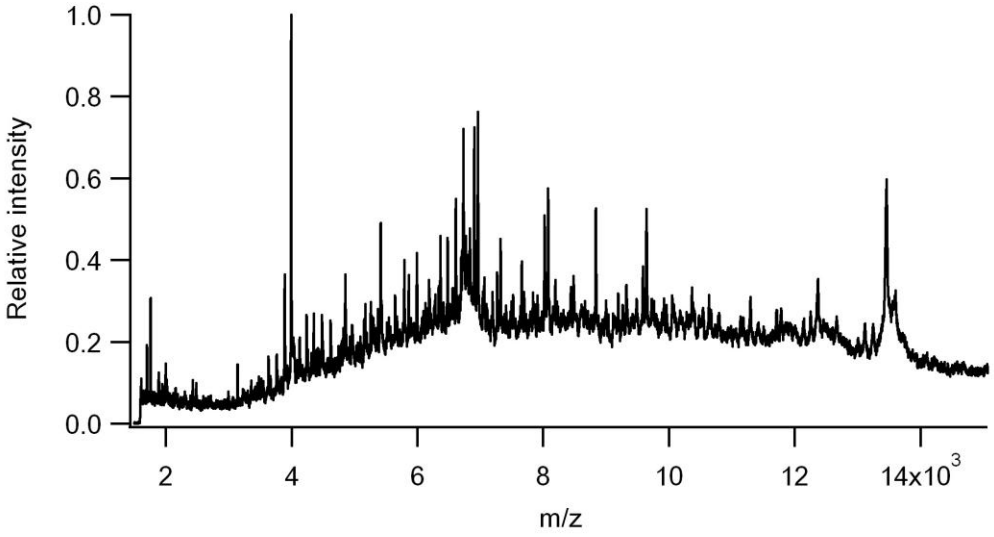
5.3.1.4 MALDI-TOF MS analysis of MAAH hydrolysates of electroeluted BSA

The purified BSA protein electroeluted from SDS-PAGE was hydrolyzed by MAAH, and then analyzed by MALDI-TOF. Figure 5.7 shows that the spectrum quality was much poorer than that of the standard BSA protein hydrolysates. The degradation of spectral quality might be caused by the residual SDS present in the eluate sample.

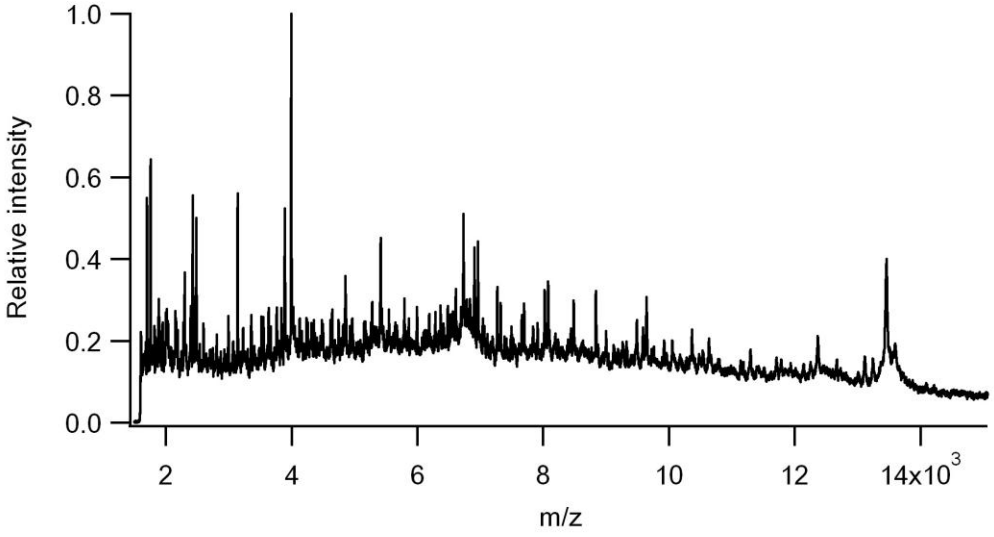
To confirm this, we spiked 0.001%, 0.005%, 0.01% and 0.05% SDS into the desalted hydrolysates of the standard BSA protein. Figure 5.8 shows the results. The baseline became wavy with SDS added into the hydrolysates. The signal suppression became more severe as more SDS was added. When the SDS concentration was increased to 0.05%, the signal was almost completely suppressed. However, no spectrum was similar to the one shown in Figure 5.7b. We also spiked 0.001%, 0.005%, 0.01% and 0.05% SDS into the standard BSA protein before MAAH. The results are shown in Figure 5.9, which were quite similar to those obtained after spiking SDS to the BSA hydrolysates.

Because we used LC to desalt both the BSA protein and hydrolysates before

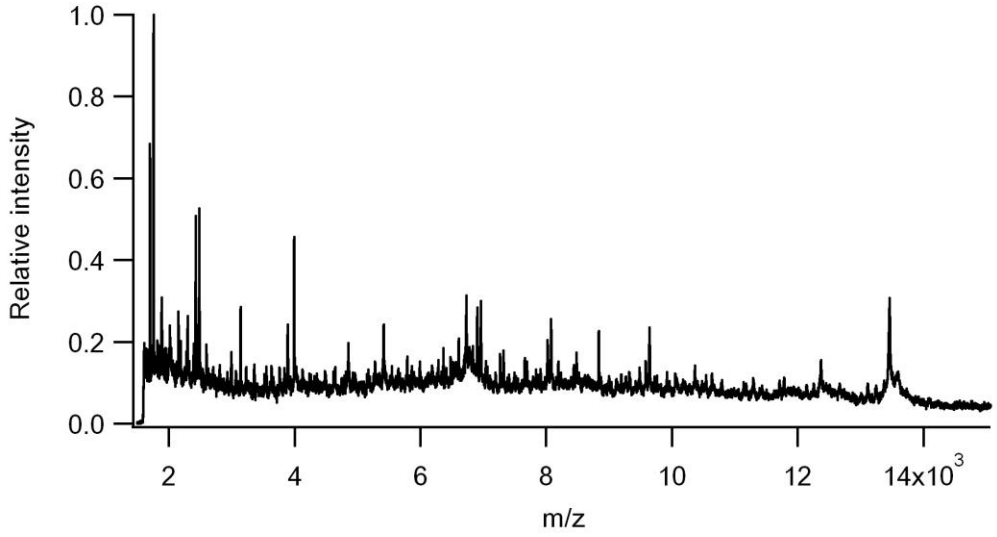
(a)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
2
4
6
8
10
12
14x10^3
m/z

(b)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
2
4
6
8
10
12
14x10^3
m/z

(c)



(d)

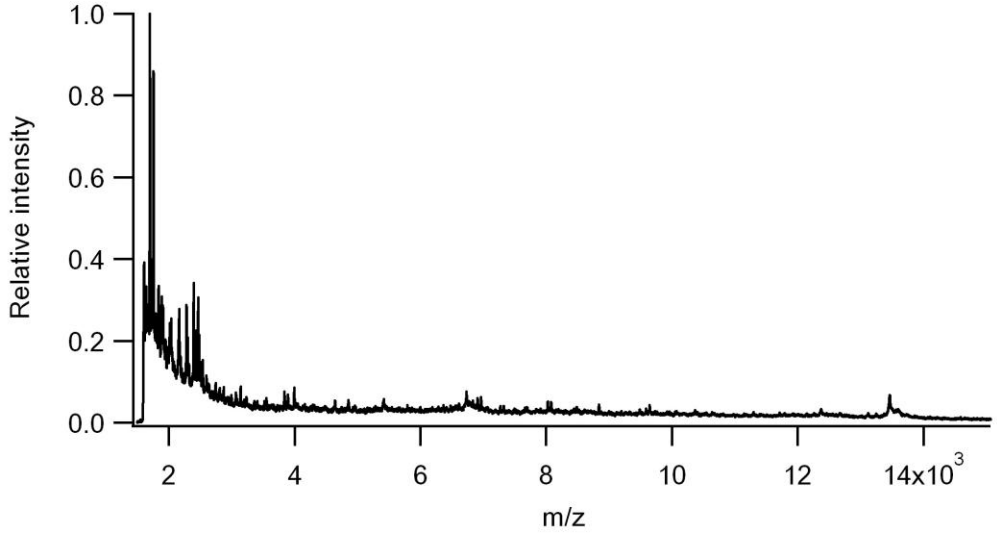
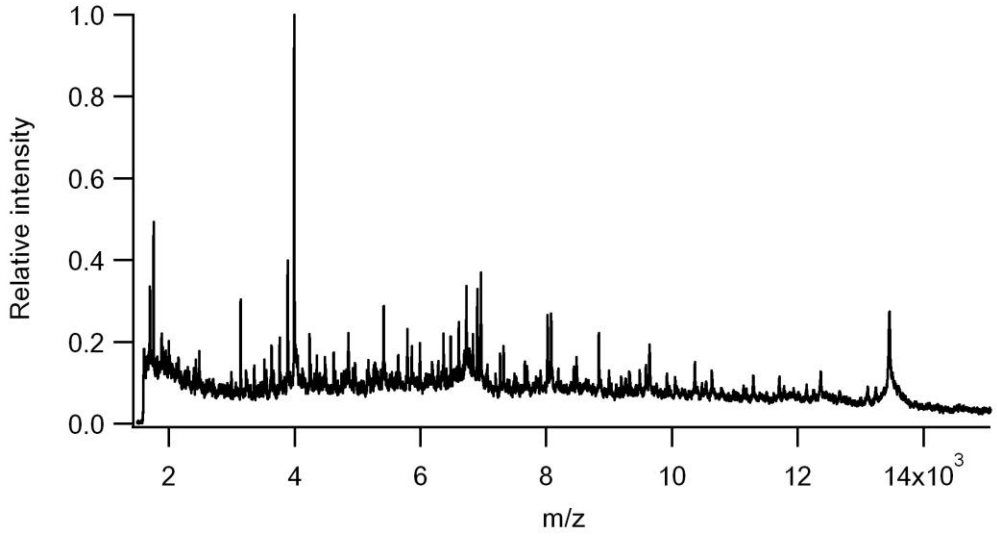
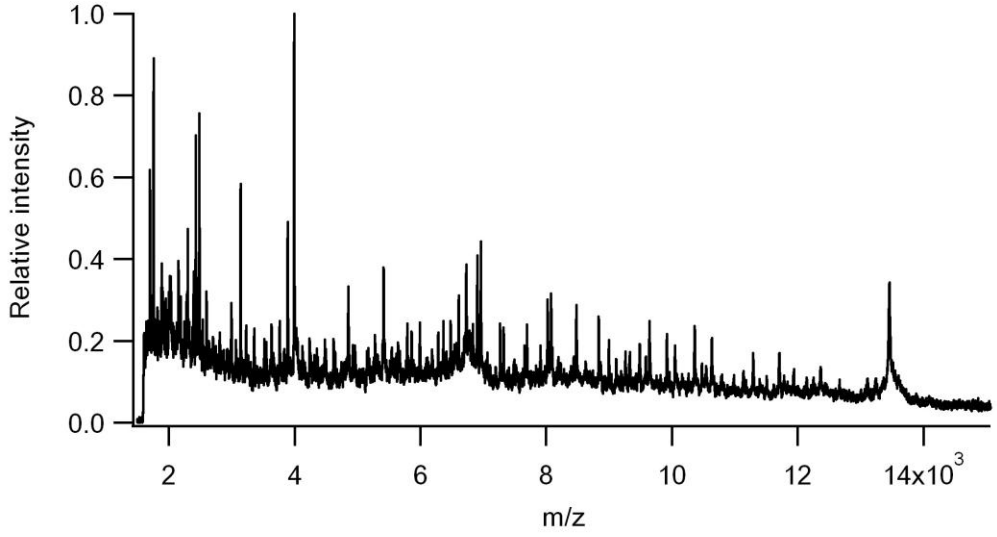


Figure 5.8 The spectra of BSA hydrolysates generated by MAAH, with (a) 0.001%, (b) 0.005%, (c) 0.01% and (d) 0.05% SDS spiked to the hydrolysates separately.

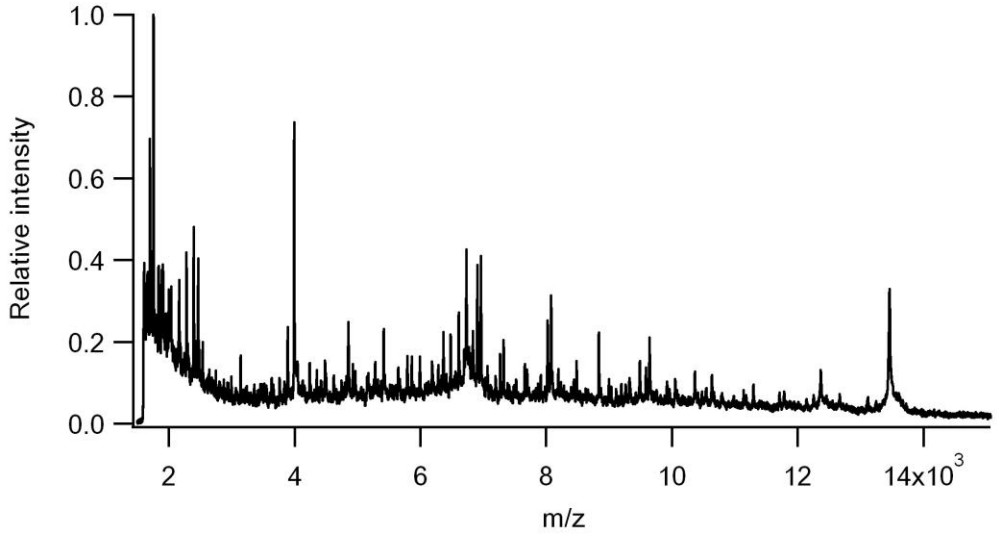
(a)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
2
4
6
8
10
12
14x10^3
m/z

(b)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
2
4
6
8
10
12
14x10^3
m/z

(c)



(d)

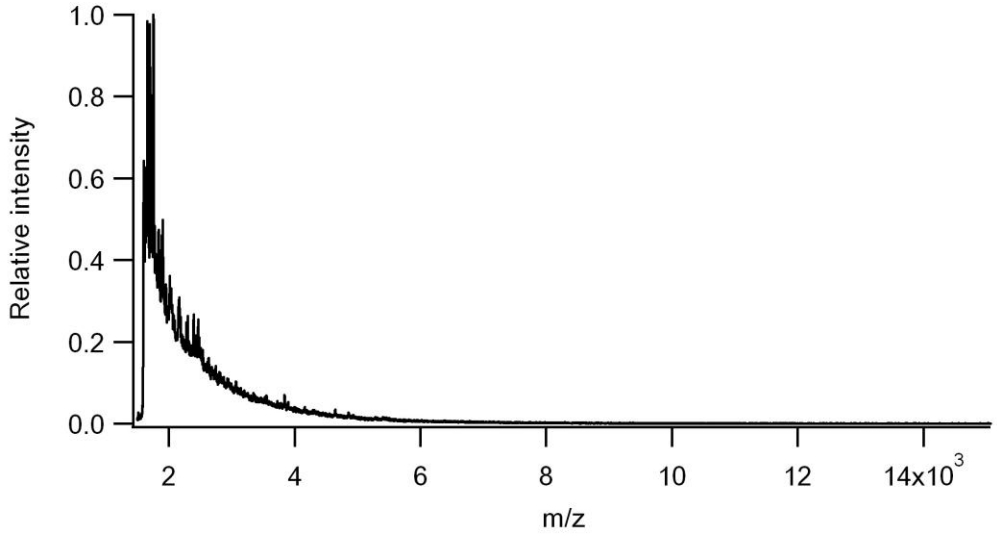
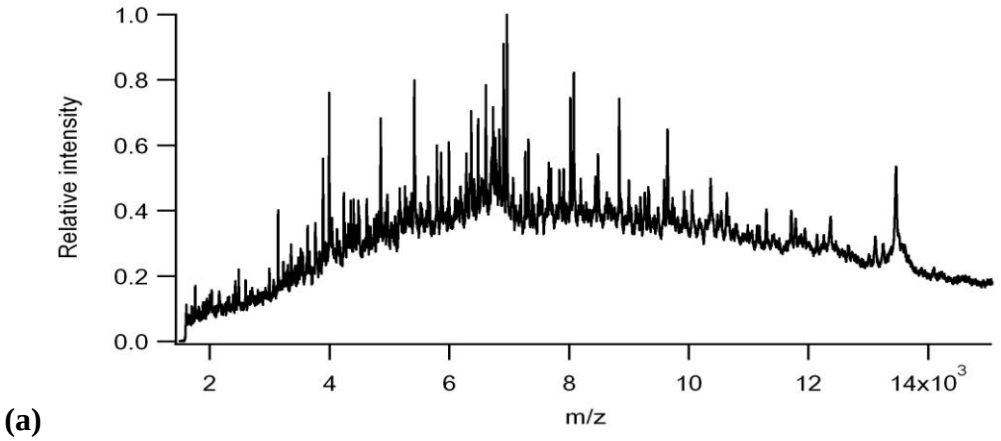
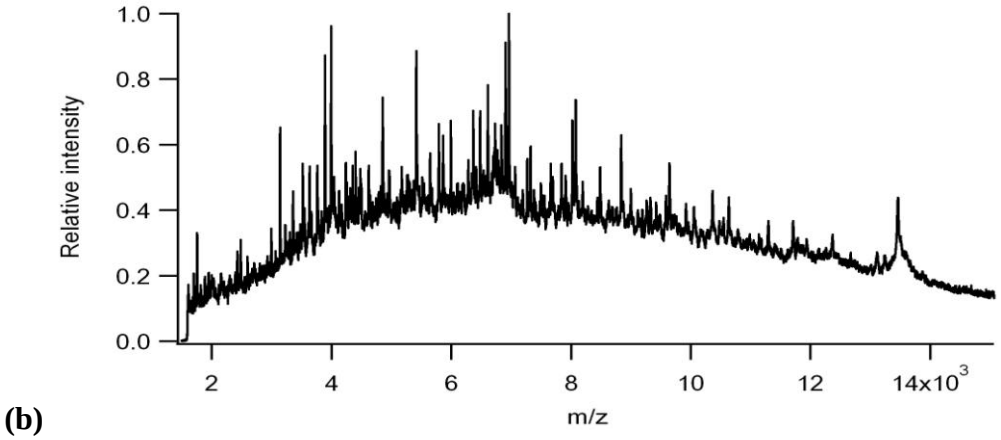


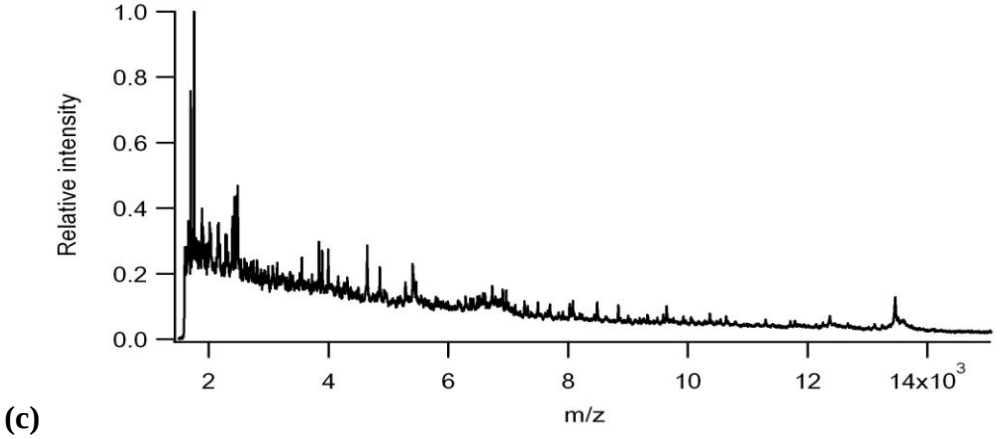
Figure 5.9 The spectra of BSA hydrolysates generated by MAAH with (a) 0.001%, (b) 0.005%, (c) 0.01% and (d) 0.05% SDS spiked to BSA standard protein before MAAH separately.



(a)



(b)



(c)

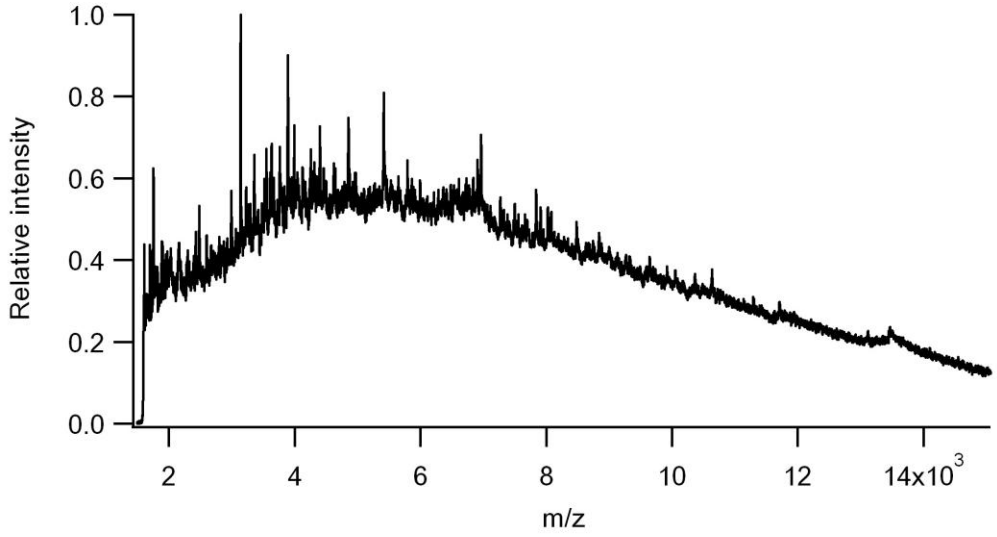
Figure 5.10 The spectra of desalted BSA hydrolysates generated by MAAH (a) without concentration of LC elution, (b) with dried-down LC elution, or (c) with 0.01% SDS spiked before desalting and dried-down LC elution.

MALDI-TOF MS analysis, the difference might rise from the LC process. We tried to desalt the standard BSA protein first with LC. Next, we completely dried down the LC eluate of protein or dried it to a volume of around 20 μL, then performed MAAH, followed by LC desalting again and MALDI-TOF MS analysis. Figure 5.10a and 10b shows that the results of the two were similar to the hydrolysates of the standard BSA protein without desalting. Residue SDS might interfere with the desalting process, too. Thus we also spiked 0.01% SDS to the BSA standard protein, then performed desalting and MAAH. Figure 5.10c demonstrates that the spectrum was similar to the one shown in Figure 5.8c, with 0.01% SDS spiked into the hydrolysates of the BSA standard protein. Thus residue SDS would not affect the desalting process. The problem was not from the LC desalting.

Acetone precipitation was one of the differences between the standard protein and the electroeluted protein. We tried to perform acetone precipitation for the standard BSA protein in 4% SDS in 250 mM NH_4HCO_3, and then reconstituted the protein pellets with DDI water, followed by sonication for 1 min or 30 min and MAAH. The results are shown in Figure 5.11. We could see the baseline was wavy and elevated, and the peaks with molecular masses larger than 10 kDa disappeared, similar to the spectrum shown in Figure 5.7b.

Because acetone precipitation generates the protein pellets, we do not know how the formation of a pellet affects the MAAH process. We tried to apply microwave irradiation for 60, 75, and 90 s separately for acetone precipitated BSA protein pellets. Figure 5.12 shows that increasing microwave irradiation time did not help

(a)



(b)

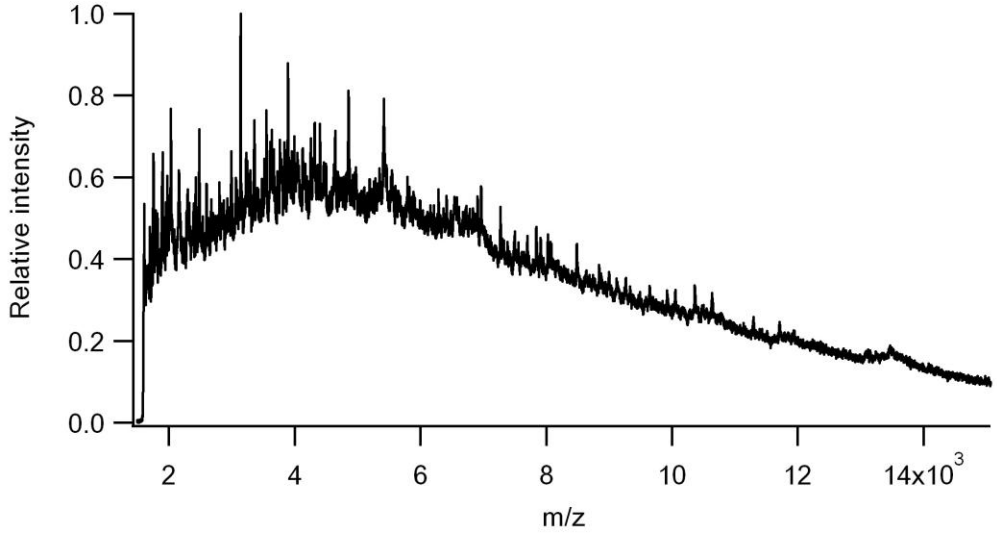
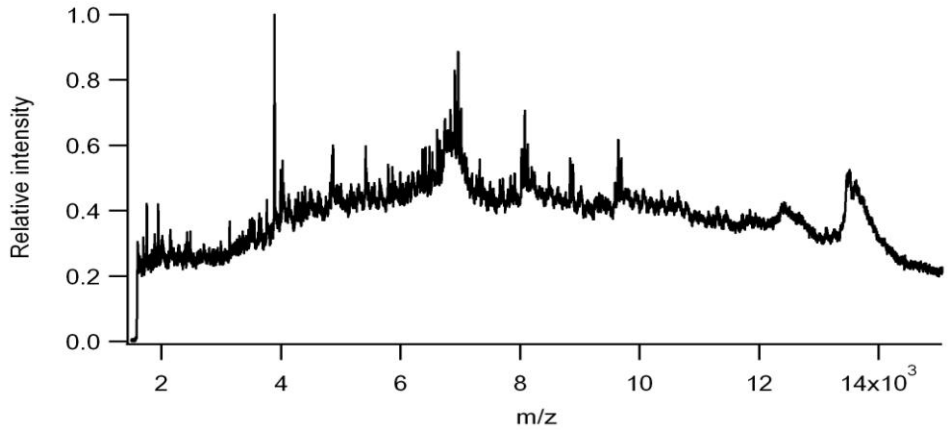
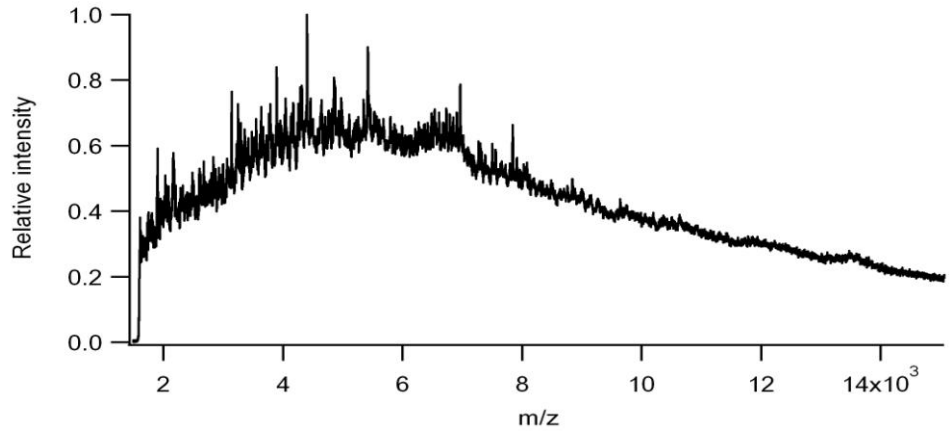


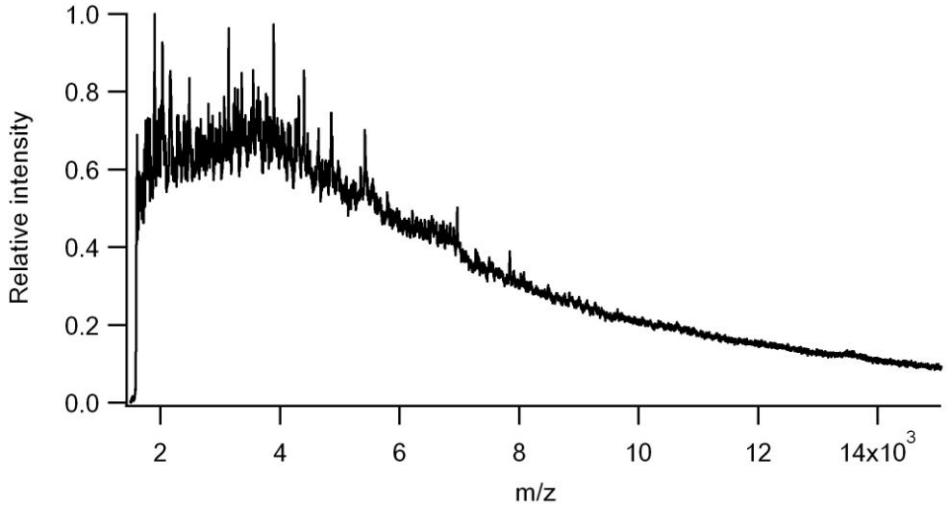
Figure 5.11 The spectra of BSA hydrolysates generated by MAAH with (a) 1 min and (b) 30 min sonication for acetone precipitated protein pellets in DDI water.



(a)



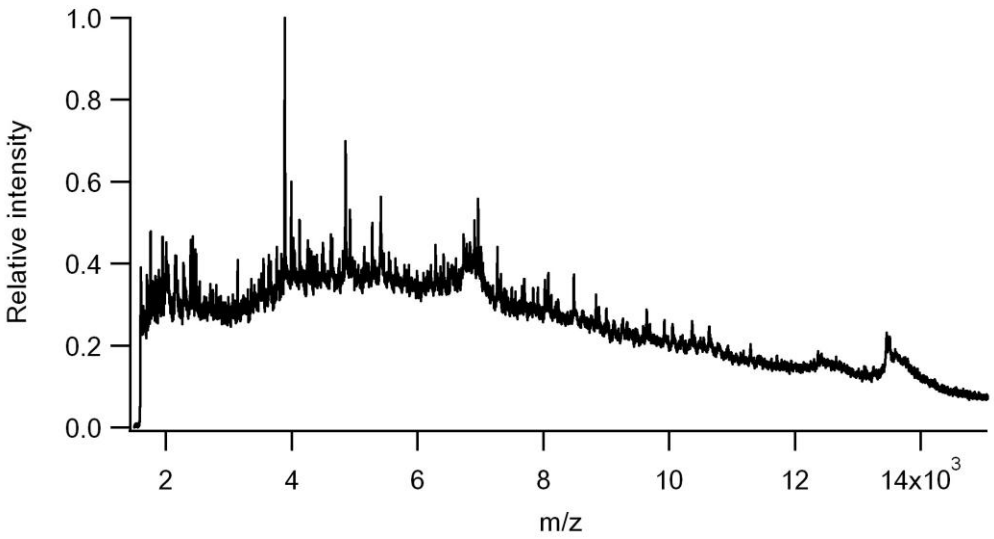
(b)



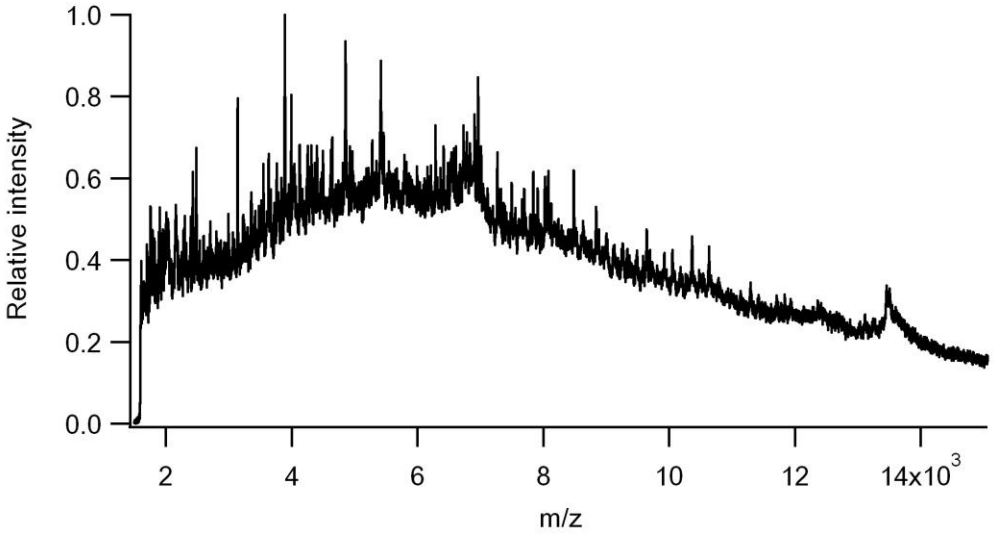
(c)

Figure 5.12 The spectra of BSA hydrolysates generated by MAAH with microwave irradiation applied for (a) 60 s, (b) 75 s, and (c) 90 s, respectively.

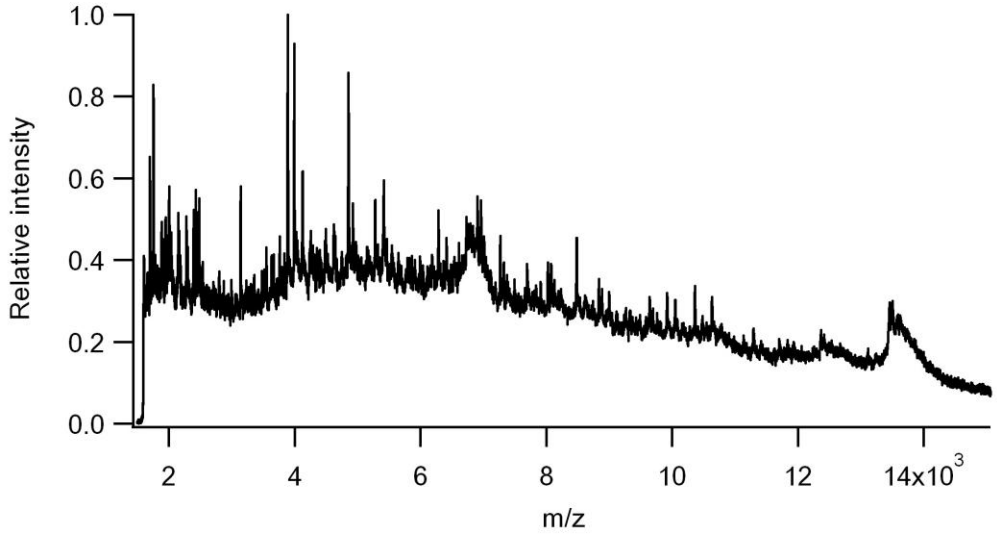
(a)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
2
4
6
8
10
12
14x10^3
m/z

(b)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
2
4
6
8
10
12
14x10^3
m/z

(c)



(d)

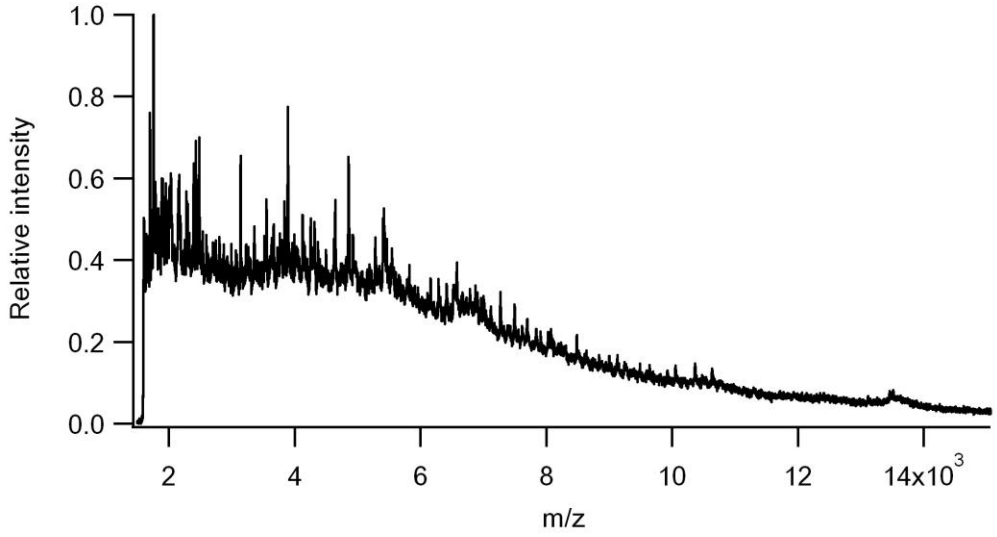


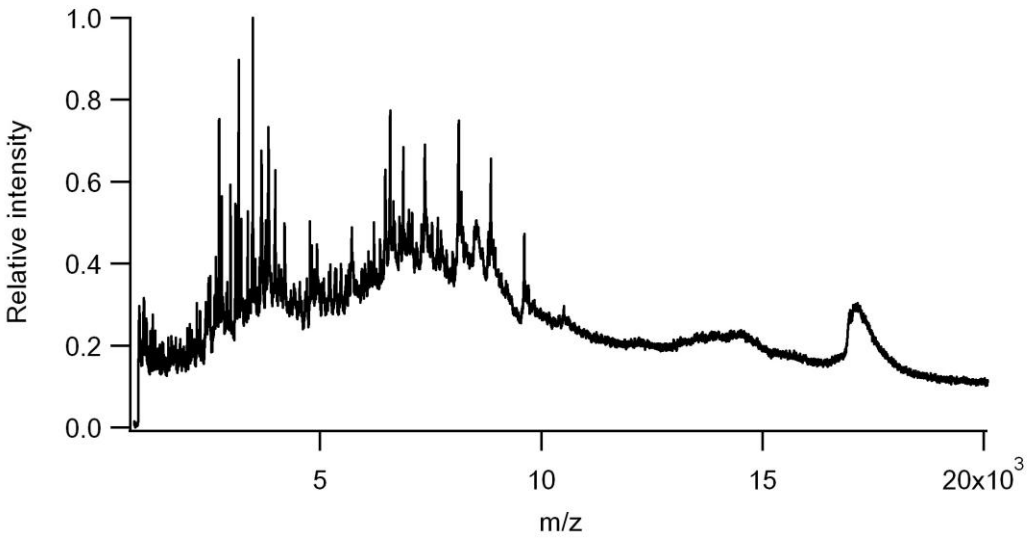
Figure 5.13 The spectra of BSA hydrolysates generated by MAAH with solubilization of acetone precipitated protein pellets by (a) DDI water, (b) 50 mM NH_4HCO_3, (c) 0.1% RapigestSF in DDI water and (d) 0.1% RapigestSF in 50 mM NH_4HCO_3.

us get a similar spectrum as the one of the standard protein hydrolysates. It only hydrolyzed terminal peptides to smaller internal peptides.

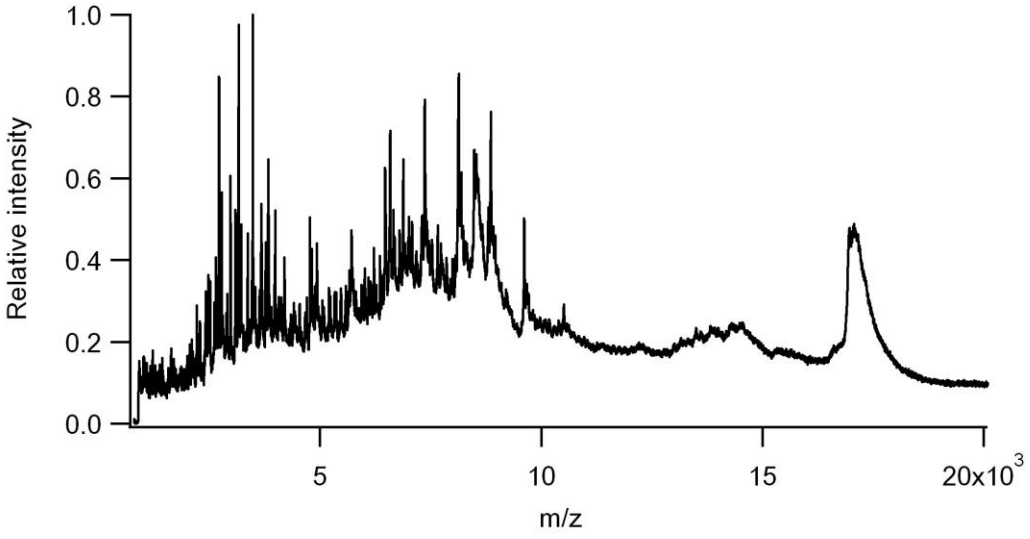
We thought the protein pellets might not dissolve well in the 3 M HCl used in MAAH. We used DDI water, 50 mM NH_4HCO_3, 0.1% RapiGestSF (Waters, Mississauga, ON, Canada) in DDI water and 0.1% RapigestSF in 50 mM NH_4HCO_3 to dissolve the protein pellets, followed by sonication for 1 hour. Then we added the same volume of 6 M HCl and performed MAAH. Since RapiGestSF is an acid labile surfactant, it will hydrolyze to two products when acid is added. One product is water immiscible and can be removed by centrifugation, while the other part is an aqueous product.[50] So RapiGestSF would not interfere with the MS analysis. Figure 5.13 shows that the spectrum qualities of intensive solubilization were still rather poor. More salts present in the MAAH solution only made proteins hydrolyze to smaller peptides by absorbing more energy.

Next, we chose to use myoglobin standard protein as a model to investigate the MAAH problem of acetone precipitated protein further. Because no reduction was needed for myoglobin, we could simplify the experiment procedures. Another reason was that myoglobin was much smaller than BSA so we could detect the peptide peaks as well as the protein peak in the spectrum. Solubilization experiment was also conducted for myoglobin to check if it was the same for another protein. Figure 5.14 shows the results. Compared to the spectrum of standard myoglobin hydrolysates shown in Figure 5.15a, the protein peaks in all spectra were broadened severely, accompanied with wavy baselines. So obviously, resolubilization of protein pellets was not a problem in MAAH.

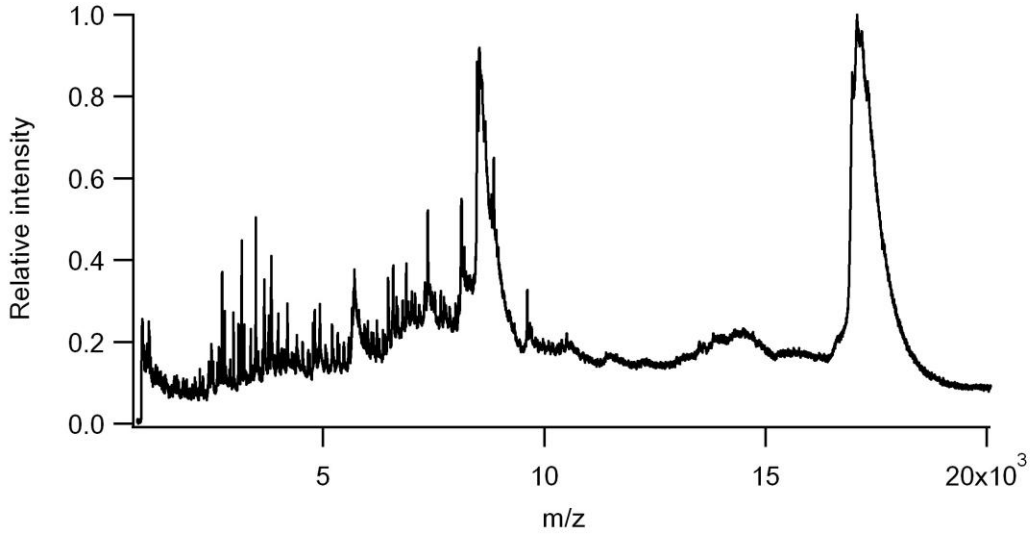
(a)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
5
10
15
20x10^3
m/z

(b)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
5
10
15
20x10^3
m/z

(c)



(d)

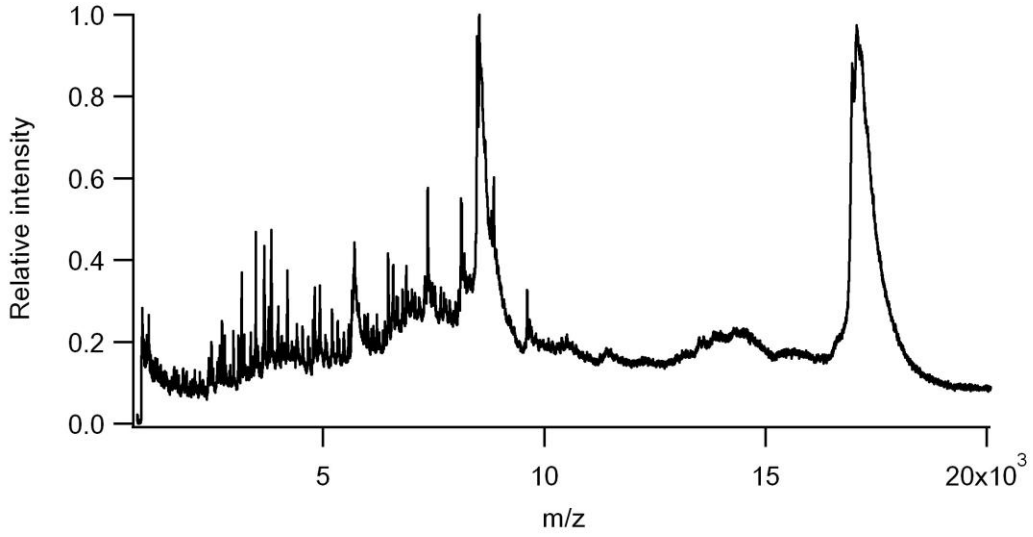
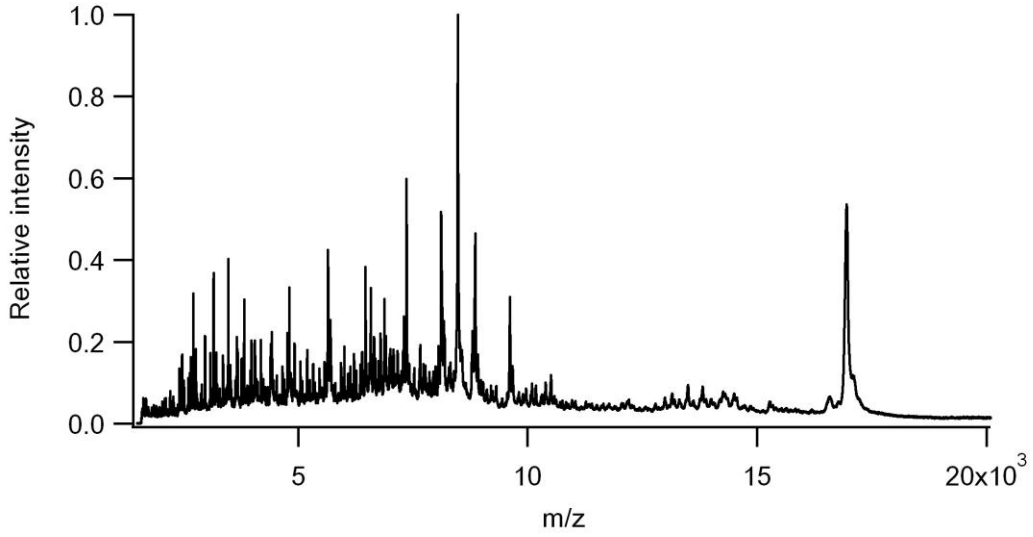
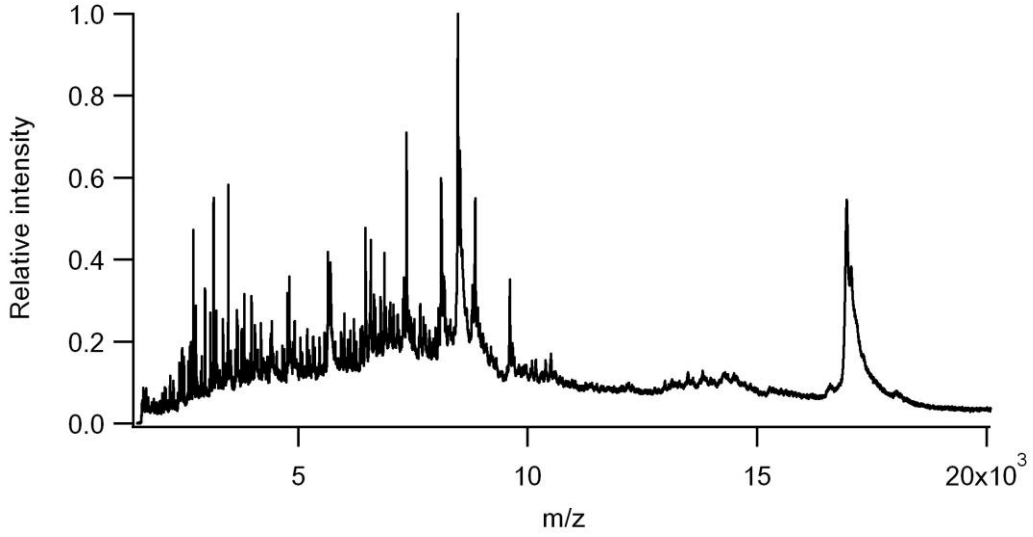


Figure 5.14 Spectra of myoglobin hydrolysates generated by MAAH with solubilization of acetone precipitated protein pellets by (a) DDI water, (b) 50 mM NH_4HCO_3, (c) 0.1% RapigestSF in DDI water and (d) 0.1% RapigestSF in 50 mM NH_4HCO_3.

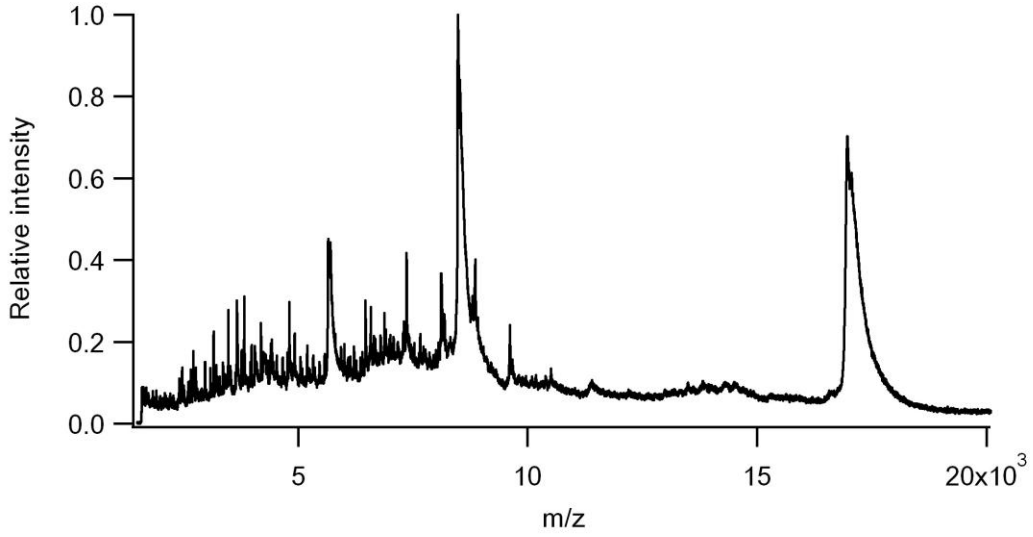
(a)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
5
10
15
20x10^3
m/z

(b)


1.0
0.8
0.6
0.4
0.2
0.0
Relative intensity
5
10
15
20x10^3
m/z

(c)



(d)

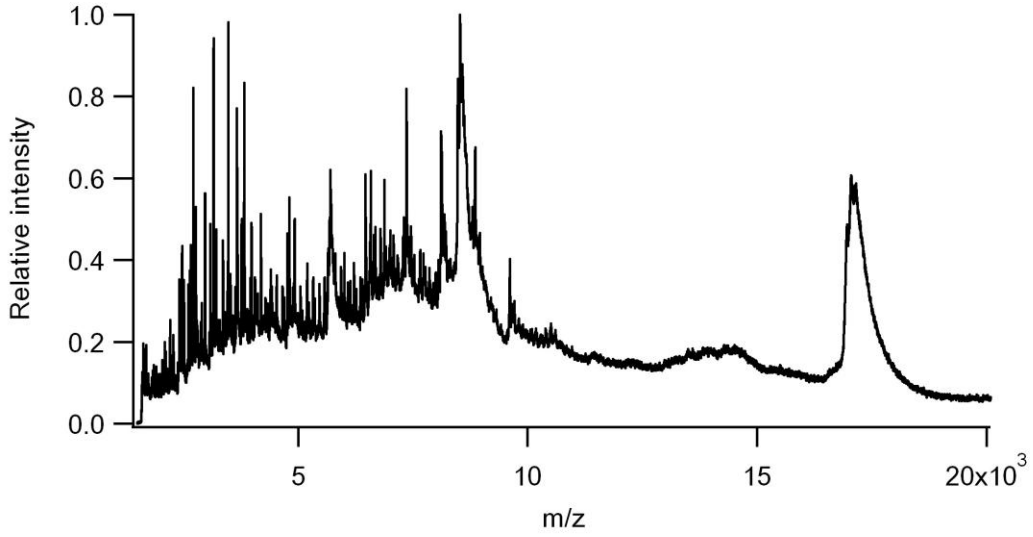
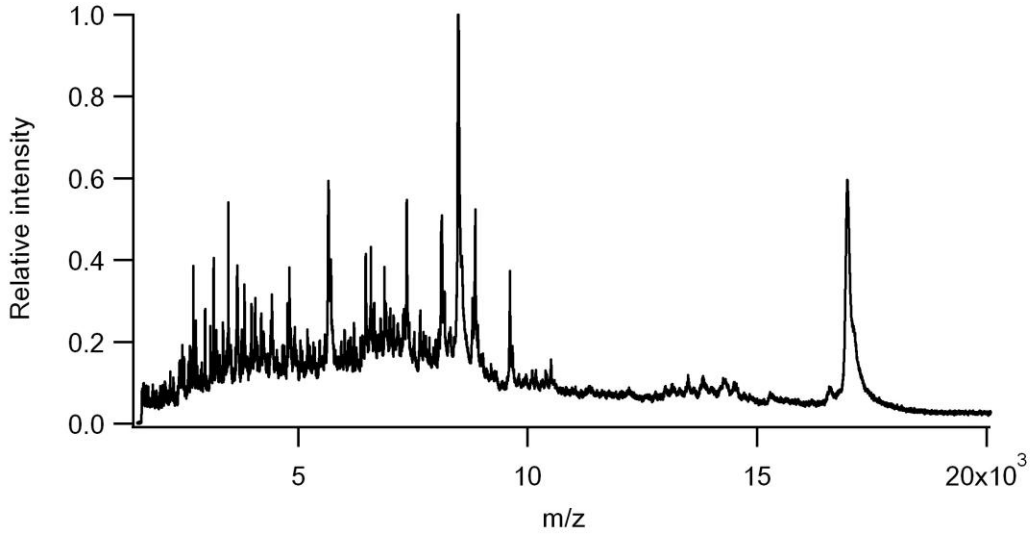


Figure 5.15 Spectra of myoglobin hydrolysates generated by MAAH for (a) myoglobin standard protein in DDI water, (b) acetone precipitated myoglobin in 50 mM NH_4HCO_3, (c) acetone precipitated myoglobin in 4% SDS, and (d) acetone precipitated myoglobin in 4% SDS in 50 mM NH_4HCO_3.

(a)



(b)

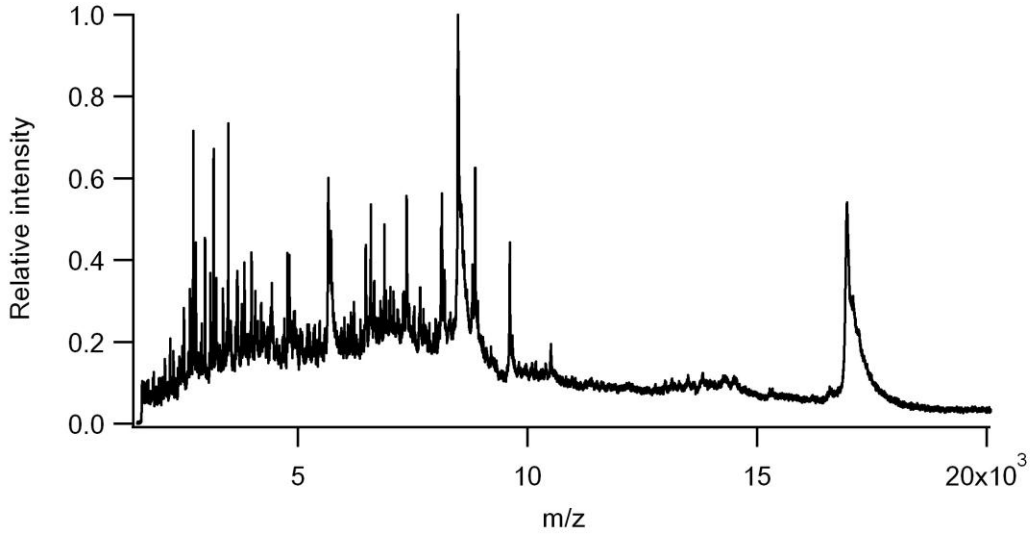
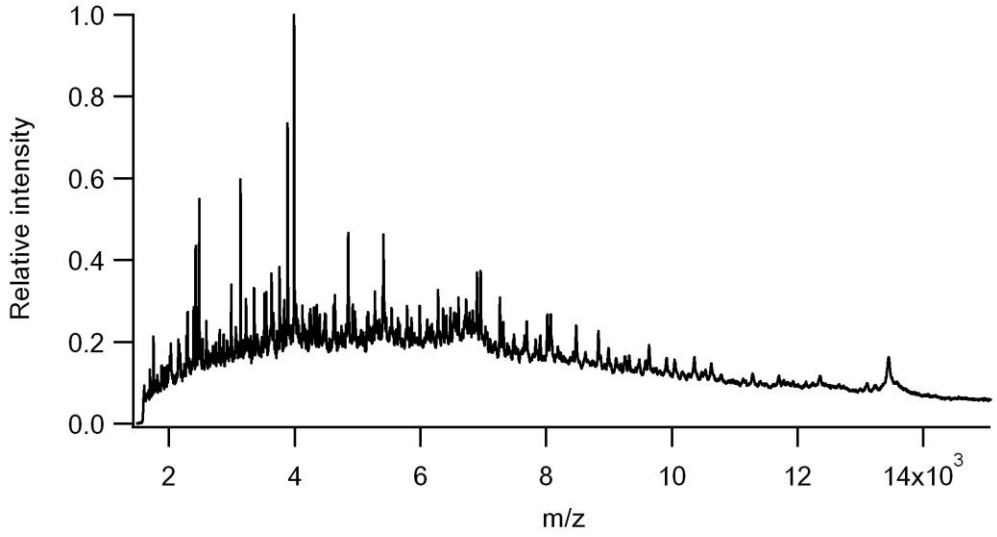


Figure 5.16 Spectra of myoglobin hydrolysates generated by MAAH for (a) acetone precipitated myoglobin in 50 mM NaCl, (b) acetone precipitated myoglobin in 4% SDS in 50 mM NaCl.

(a)



(b)

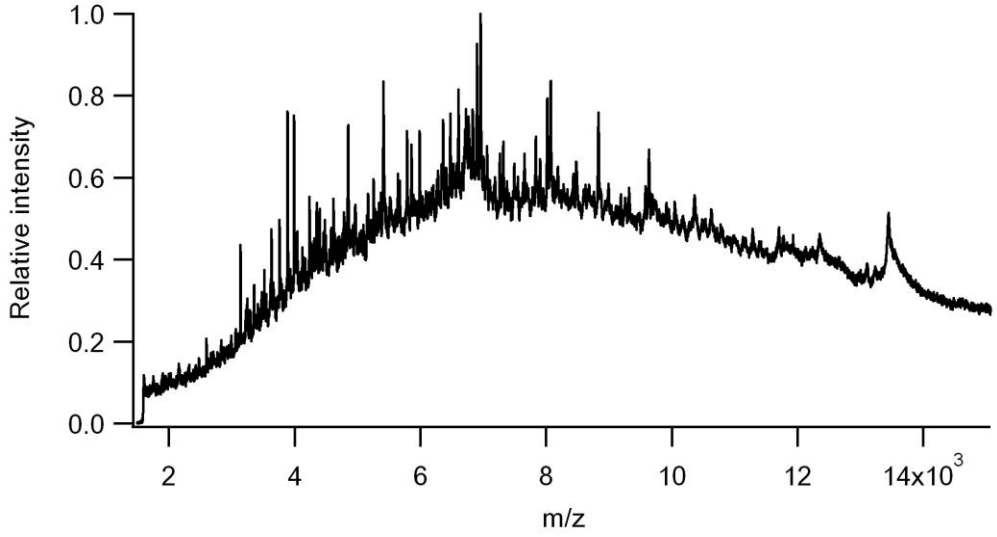


Figure 5.17 Spectra of BSA hydrolysates generated by MAAH for (a) acetone precipitated BSA in 50 mM NaCl, (b) acetone precipitated BSA in 4% SDS in 50 mM NaCl.

To study how salt and SDS in acetone precipitation affect MAAH, we prepared 10 μg myoglobin in 100 μL 50 mM NH_4HCO_3, 4% SDS, and 4% SDS in 50 mM NH_4HCO_3 separately. Then we conducted acetone precipitation and MAAH. Figure 5.15 shows the results. The presence of SDS deteriorated the spectrum. However, acetone precipitation of myoglobin in 50 mM NH_4HCO_3 made the protein peak broaden in spectrum, too. So if SDS and NH_4HCO_3 were both in protein solution before acetone precipitation, the protein peak would be more broadened.

We changed NH_4HCO_3 to NaCl to check if a different salt has the same effect. Surprisingly, as Figure 5.16a shows, the spectrum of acetone precipitated myoglobin in 50 mM NaCl was as good as standard myoglobin hydrolysates. Also the spectrum of acetone precipitated myoglobin in 4% SDS in 50 mM NaCl (Figure 5.16b) was similar as the one shown in Figure 5.15c, i.e., acetone precipitated myoglobin in 4% SDS only. Thus acetone precipitation is not a problem for MAAH and it is actually the kind of salts used matters. How different salts play a role in acetone precipitation and MAAH is still unknown.

We tried to perform the acetone precipitation experiment for BSA in 50 mM NaCl and in 4% SDS in 50 mM NaCl as well. Similar results as myoglobin resulted in Figure 5.17. Although the baseline was elevated because of the presence of residual SDS, the peak broadening problem had been eased.

Then we used 0.1% SDS in 50 mM NaCl in electroelution. However, Cl^- ion would change to Cl_2 in electrochemical reaction and further reaction of Cl_2 and

H_2O would form HCl and HClO. The dialysis membrane of membrane cap was destroyed by the acid formed, so we did not recover any protein in this case.

5.3.1.5 LC ESI MS/MS analysis of MAAH hydrolysates of electroeluted BSA

To better analyze the MAAH hydrolysates of eletroeluted BSA, we used LC ESI MS/MS, instead of MALDI-TOF MS.

In this case, 15 min LC ESI MS/MS analysis identified 12 N-term peptides and 8 C-term peptides. For in-solution HCl MAAH of BSA standard protein, 17±5 N-term peptides and 19±2 C-term peptides could be identified using the same 15 min LC ESI MS/MS analysis. The number of identified terminal peptides of electroeluted protein was fewer than the number of identified terminal peptides of standard protein. Considering so many purifying steps used for electroeluted protein, this number was reasonable and we still could use this technique to identify terminal peptides of proteins separated by SDS-PAGE.

5.3.2 Discussion

The problem of electroelution using the apparatus from Bio-Rad is that many purification steps are needed for the downstream MS analysis. So the protein recovery is rather low, only around 25% recovery rate for 40 μg BSA loaded into the gel. We did not test lower amount samples, but the recovery rates should be even smaller. Larger quantities of sample also overload the gel and lower the separation resolution. The relatively large sample amount needed limits the application areas of this method.

Schuhmacher *et al.*[39] also used the same electroelution apparatus to extract proteins from gels for ESI-MS analysis. They isolated proteins by electroelution with SDS-free ammonium acetate (pH 2.5) buffer and reversed polarity. Since the isoelectric point (pI) values of most proteins range from 3-10, proteins are positively charged at pH < 3 with resultant migration to the cathode upon *in situ* dissociation of the protein-SDS-complex. They determined protein recoveries densitometrically. For 2 to 20 hour elution time, the recoveries of 20 μg BSA were only about 10% to 20%. After 24 hour electroelution, recovery of BSA increased to a maximum of 58%. Although electroelution with SDS-free elution buffer takes a rather long time to get over 50% protein recoveries, the protein elution is SDS-free, and ready for downstream MS analysis. The sensitivity of this method should not be very high, considering the difficulties of electroelution with SDS-free elution buffer.

Their modified electroelution method has better recoveries as well as simpler steps than ours, but our in-depth studies of electroelution with SDS-containing elution buffer could give others references when they try to perform electroelutions combined with MS analysis.

5.4 Conclusions

We have developed a method combining electroelution of a gel-separated protein, MAAH and LC ESI MS/MS for protein sequence analysis.

The electroelution apparatus of Bio-Rad concentrates the SDS and salts in the protein eluate, which presents a major challenge to the purification work. We applied equilibrium dialysis to remove NH_4HCO_3 first, and then used acetone

precipitation to remove SDS and Coomassie blue at the same time. Recovery was around 25% by LC-UV quantification. MALDI-TOF MS analysis shows that the spectrum of purified protein electroeluted from gel was similar as the standard protein.

The spectrum quality of MAAH hydrolysates of electroeluted protein was much poorer than that of the standard protein hydrolysates. We found salts in protein solution and acetone precipitation played a role in the peak broaden problem. Changing NH_4HCO_3 to NaCl would solve the problem, but NaCl was not appropriate for electroelution due to the acid produced in electrochemical reactions. Other salts such as Na_2SO_4 and Na_3HPO_4 are worth trying.

Instead of using MALDI-TOF, LC ESI MS/MS analysis was applied to analyze the low molecular weight peptides in the MAAH hydrolysates of electroeluted protein. 12 N-term peptides and 8 C-term peptides were detected in 15 minutes LC ESI MS/MS analysis.

Combining electroelution and LC ESI MS/MS analysis, we could unambiguously identify terminal peptides of proteins separated by SDS-PAGE.

The problem of this method is that many purifying steps are needed for the downstream MS analysis, and thus the protein recovery is rather low (only around 25% recovery rate for 40 μg BSA loaded into the gel). The relatively large sample amount required limits the application areas of this method.

5.5 Literature Cited

(1) Liang, X.; Bai, J.; Liu, Y. H.; Lubman, D. M. *Anal Chem* **1996**, *68*, 1012.

(2) Cooper, J. W.; Gao, J.; Lee, C. S. *Electrophoresis* **2004**, *25*, 1379.
(3) van Montfort, B. A.; Canas, B.; Duurkens, R.; Godovac-Zimmermann, J.; Robillard, G. T. *J Mass Spectrom* **2002**, *37*, 322.
(4) Mortz, E.; Sareneva, T.; Haebel, S.; Julkunen, I.; Roepstorff, P. *Electrophoresis* **1996**, *17*, 925.
(5) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal Chem* **1996**, *68*, 850.
(6) Hellman, J. *Proteomics* **2007**, *7*, 3441.
(7) Zhou, J.; Hu, Y.; Lin, Y.; Liu, H.; Xie, P. *Journal of Chromatography. B,* **2011**, *879*, 2957.
(8) Medzihradszky, K. F.; Leffler, H.; Baldwin, M. A.; Burlingame, A. L. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 215.
(9) Palviainen, M.; Raekallio, M.; Vainionpaa, M.; Kosonen, S.; Vainio, O. *Toxicon : official journal of the International Society on Toxinology* **2012**, *60*, 1228.
(10) Li, G.; Waltham, M.; Anderson, N. L.; Unsworth, E.; Treston, A.; Weinstein, J. N. *Electrophoresis* **1997**, *18*, 391.
(11) Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P. R. *Rapid Communications in Mass Spectrometry : RCM* **2000**, *14*, 496.
(12) Klarskov, K.; Roecklin, D.; Bouchon, B.; Sabatie, J.; Van Dorsselaer, A.; Bischoff, R. *Anal Biochem* **1994**, *216*, 127.
(13) van Montfort, B. A.; Doeven, M. K.; Canas, B.; Veenhoff, L. M.; Poolman, B.; Robillard, G. T. *Biochimica et Biophysica Acta* **2002**, *1555*, 111.
(14) Borchers, C.; Peter, J. F.; Hall, M. C.; Kunkel, T. A.; Tomer, K. B. *Anal Chem* **2000**, *72*, 1163.
(15) Bantscheff, M.; Weiss, V.; Glocker, M. O. *Biochemistry* **1999**, *38*, 11012.
(16) Sutton, C. W.; Pemberton, K. S.; Cottrell, J. S.; Corbett, J. M.; Wheeler, C. H.; Dunn, M. J.; Pappin, D. J. *Electrophoresis* **1995**, *16*, 308.
(17) Sickmann, A.; Dormeyer, W.; Wortelkamp, S.; Woitalla, D.; Kuhn, W.; Meyer, H. E. *Electrophoresis* **2000**, *21*, 2721.
(18) Larsson, T.; Norbeck, J.; Karlsson, H.; Karlsson, K. A.; Blomberg, A. *Electrophoresis* **1997**, *18*, 418.
(19) Ogorzalek Loo, R. R.; Stevenson, T. I.; Mitchell, C.; Loo, J. A.; Andrews, P. C. *Anal Chem* **1996**, *68*, 1910.
(20) Patterson, S. D. *Electrophoresis* **1995**, *16*, 1104.
(21) Ogorzalek Loo, R. R.; Mitchell, C.; Stevenson, T. I.; Martin, S. A.; Hines, W. M.; Juhasz, P.; Patterson, D. H.; Peltier, J. M.; Loo, J. A.; Andrews, P. C. *Electrophoresis* **1997**, *18*, 382.
(22) Quach, T. T.; Li, N.; Richards, D. P.; Zheng, J.; Keller, B. O.; Li, L. *J Proteome Res* **2003**, *2*, 543.
(23) Li, A.; Sowder, R. C.; Henderson, L. E.; Moore, S. P.; Garfinkel, D. J.; Fisher, R. J. *Anal Chem* **2001**, *73*, 5395.
(24) Xiang, B.; MacIsaac, S.; Lardizabal, K.; Li, B. *Rapid Communications in Mass Spectrometry : RCM* **2010**, *24*, 3447.
(25) Sun, W.; Gao, S.; Wang, L.; Chen, Y.; Wu, S.; Wang, X.; Zheng, D.; Gao, Y. *Mol Cell Proteomics* **2006**, *5*, 769.

(26) Juan, H. F.; Chang, S. C.; Huang, H. C.; Chen, S. T. *Proteomics* **2005**, *5*, 840.
(27) Hua, L.; Low, T. Y.; Sze, S. K. *Proteomics* **2006**, *6*, 586.
(28) Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 1791.
(29) Ehring, H.; Stromberg, S.; Tjernberg, A.; Noren, B. *Rapid Communications in Mass Spectrometry : RCM* **1997**, *11*, 1867.
(30) Jeannot, M. A.; Zheng, J.; Li, L. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 512.
(31) Galvani, M.; Bordini, E.; Piubelli, C.; Hamdan, M. *Rapid Communications in Mass spectrometry : RCM* **2000**, *14*, 18.
(32) Mirza, U. A.; Liu, Y. H.; Tang, J. T.; Porter, F.; Bondoc, L.; Chen, G.; Pramanik, B. N.; Nagabhushan, T. L. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 356.
(33) Cohen, S. L.; Chait, B. T. *Anal Biochem* **1997**, *247*, 257.
(34) Lahm, H. W.; Langen, H. *Electrophoresis* **2000**, *21*, 2105.
(35) Eckerskorn, C.; Strupat, K.; Karas, M.; Hillenkamp, F.; Lottspeich, F. *Electrophoresis* **1992**, *13*, 664.
(36) Ploug, M.; Stoffer, B.; Jensen, A. L. *Electrophoresis* **1992**, *13*, 148.
(37) Nesatyy, V. J.; Ross, N. W. *The Analyst* **2002**, *127*, 1180.
(38) Fabris, D.; Vestling, M. M.; Cordero, M. M.; Doroshenko, V. M.; Cotter, R. J.; Fenselau, C. *Rapid communications in mass spectrometry : RCM* **1995**, *9*, 1051.
(39) Schuhmacher, M.; Glocker, M. O.; Wunderlin, M.; Przybylski, M. *Electrophoresis* **1996**, *17*, 848.
(40) Lei, Z.; Anand, A.; Mysore, K. S.; Sumner, L. W. *Methods Mol Biol* **2007**, *355*, 353.
(41) Hossein, A. N.; Paknejad, M.; Mohammadian, T.; Farshad, F.; Shadisadat, S. E. *Clin Biochem* **2011**, *44*, S316.
(42) Dunn, M. J. *Methods Mol Biol* **2004**, *244*, 339.
(43) Seelert, H.; Krause, F. *Electrophoresis* **2008**, *29*, 2617.
(44) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Anal Chem* **2009**, *81*, 1049.
(45) Campbell, R.; Winkler, M. A.; Wu, H. *Anal Biochem* **2004**, *335*, 98.
(46) Domínguez, A.; Fernández, A.; González, N.; Iglesias, E.; Montenegro, L. *J. Chem. Educ* **1997**, *74*, 1227.
(47) Umlong, I. M.; Ismail, K. *Colloids and Surfaces A: Physicochem. Eng. Aspects* **2007**, *299*, 8.
(48) Jacquier, J. C.; DesbLne, P. L. *J. Chromatogr. A.* **1996**, *743*, 307.
(49) Wessel, D.; Fugge, U. I. *Anal Biochem* **1984**, *138*, 141.
(50) Wu, F.; Sun, D.; Wang, N.; Gong, Y.; Li, L. *Anal Chim Acta* **2011**, *698*, 36.

Chapter 6

Microwave-Assisted Acid Hydrolysis of Proteins Combined with Peptide Fractionation and Mass Spectrometry Analysis for Characterizing Protein Terminal Sequences*

6.1 Introduction

Protein terminal truncation and other modifications are common post-translational modifications that can alter protein structures and functions.[1-4] In addition, protein production and storage of protein-based pharmaceutical preparations may also cause the terminal degradation, resulting in changes of drug efficacy.[5] Thus determining any changes of the protein termini is very important for biological studies as well as for quality control during the production of protein-based therapeuticals and vaccines. Edman degradation has been traditionally used for protein N-terminal sequencing.[6] However, this method only applies to the proteins whose N termini are unblocked or unmodified at the N-terminal amino acid and the process is time-consuming. For protein C-terminal sequencing, Schlack-Kump degradation with (iso)thiocyanate[7,8] is a reaction similar to Edman degradation. The disadvantages of this method are low efficiency and low reproducibility and generally less than ten residues of protein C-terminus can be obtained. In

---

*A form of this chapter has been submitted for publication as: L. Chen, N. Wang and L. Li, "Microwave-Assisted Acid Hydrolysis of Proteins Combined with Peptide Fractionation and Mass Spectrometry Analysis for Characterizing Protein Terminal Sequences", Journal of Proteomics, submitted.

comparison, MS-based methods are fast, sensitive and can identify protein modifications. In many areas of applications, MS has become the preferred method for protein sequencing. To generate the terminal peptide sequence information, proteins are digested by exopeptidases such as CPY and CPB[9,10] or chemicals such as cyanogen bromide (CNBr)[11]. The cleaved peptides can be analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) or electrospray ionization (ESI) tandem mass spectrometry (MS/MS). However, protein digestion using an enzyme such as trypsin or a combination of multiple enzymes may fail to generate the terminal peptides with suitable sizes or at a sufficient abundance for successful MS sequencing.[12] In some cases, the quality of the MS/MS spectrum may not be sufficiently high for deducing information on individual amino acids or modified amino acids in the terminal peptides. Top-down proteomics provides another way for protein sequence analysis. MALDI in-source decay (ISD),[13] ESI electron capture dissociation (ECD) and electron transfer dissociation (ETD)[14,15] have been proven to be efficient to identify terminal peptides of mid-size proteins. However, these methods usually cannot provide the individual amino acid sequence information on the terminal peptides, particularly for relatively large proteins. New strategies for selective capturing of the terminal peptides from a proteomic digest that are subsequently sequenced by LC-MS/MS have been developed for relatively high throughput analysis of protein termini.[2,16-19]

One alternative approach that is particularly suitable for detailed characterization of a protein sequence is based on the use of microwave-assisted acid hydrolysis (MAAH) to degrade a protein into many short peptides, followed by MS and

MS/MS analysis of the resultant peptides.[20-23] The sequences of these short peptide ladders often overlap. Thereby a complete picture of the amino acid linkage or any modifications therein can be deduced.[24] This shotgun protein sequencing method can cover all or a large portion of the individual amino acids in a sequence. However, in some cases, the terminal peptide sequences may be missing due to relatively low abundance of these peptides generated, compared to the internal peptides, and ion suppression during the detection. For internal sequence, if a small set of peptide ladders containing a stretch of amino acids are not detected by MS due to their low abundances or ion suppression, other sets of more detectable overlapping peptide ladders can cover these amino acids. However, for terminal peptides, if a set of terminal peptides are not detected by MS, no other more detectable peptide ladders can cover this region.

In this chapter, we describe an improved shotgun sequencing method for characterizing terminal peptides. To increase the likelihood of detecting the terminal peptides, we developed a rapid HPLC fractionation method to fractionate the hydrolysate of a protein generated by MAAH using 3 M HCl. The low molecular weight peptides collected contain many terminal peptides that could be sequenced using LC-MS/MS. This method was developed using bovine serum albumen (BSA) as a model protein. Both MALDI-TOF and LC-MS/MS techniques were examined for detecting the terminal peptides. This method was then applied to the analysis of a recombinant protein (a red fluorescent protein mCherry with a short sequence tag added to the N-termini) and its N-terminal truncated form (mCherry minus the tag) to illustrate how this method can be used to characterize a protein or protein product in a real world situation.

6.2 Experimental

6.2.1 Chemicals and Reagents

Dithiothreitol (DTT), ammonium bicarbaonate (NH_4HCO_3), trifluoroacetic acid (TFA), LC-MS grade formic acid (FA), and bovine serum albumin were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). LC-MS grade water and acetonitrile (ACN) were from Fisher Scientific Canada (Edmonton, AB, Canada). ACS grade 37% HCl was from Merck (KGaA, Darmstadt, Germany).

6.2.2 mCherry Protein Preparation

Electrocompetent *E. coli* strain DH10B (Invitrogen) was transformed with plasmid containing the gene of the target protein (pBAD-mCherry) and plated on Lysogeny Broth (LB) medium (1% tryptone (BD 211705), 0.5% yeast extract (BD 212750), and 1% NaCl (Caledon 7560-1 from Sigma-Aldrich) agar plates supplemented with 0.1 mg/mL ampicillin (Fisher BP1760-25) and 0.02% L-arabinose (Sigma A3256). Then the plates were incubated for overnight at 37 ℃.

The cell culture and preparation of the cell lysates of *E. coli* were similar to those reported [25] with some modifications. Five mL LB culture of mCherry cells from the LB was incubated at 37 ℃ with shaking at 225 rpm overnight. The culture was centrifuged at 3901 *g* for 10 min. The pellets were resuspended in 50 mL of LB and incubated at 37 ℃ with shaking at 225 rpm overnight. Then the culture was centrifuged at 3901 *g* for 10 min. The pellets were resuspended in 500 mL of LB. Cells were incubated at 37 ℃ with shaking at 225 rpm for overnight. After

cooling the medium to 4 ℃, cells were centrifuged at 11300 g for 20 min at 4 ℃. The pellets were resuspended in 10 mL lysis buffer (50 mM sodium phosphate of pH 8.0, 300 mM NaCl 10 mM imidazole, 0.05 mg/mL DNase I (Roche 10 104 159 001), 0.01 mg/mL RNase A (Sigma R5000), 1 mg/mL lysozyme (Sigma L-7651), 1 tablet Complete Mini EDTA-Free Protease Inhibitor Cocktail (Roche 1183617000). The suspension was allowed to sit on ice for 15 min and then sonicated for 6 × 10 s. The lysate was centrifuged at 10900 *g* for 30 min at 4 ℃ to pellet unbroken cells. The supernatant was incubated with 1 mL Ni-NTA agarose beads (Qiagen) equilibrated in lysis buffer (minus enzymes and protease inhibitors) with gentle agitation for 1 h at 4 ℃. The beads were collected in a disposable column and washed with wash buffer (50 mM sodium phosphate of pH 8.0, 300 mM NaCl 20 mM imidazole) until no red colour was visible in the wash. The mCherry protein was eluted with an elution buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 250 mM imidazole) until all the red colour was removed from the column. The eluted protein solution was dialyzed to an exchange buffer (50 mM sodium phosphate of pH 8.0, 50 mM NaCl, 2 mM $CaCl_2$), and then precipitated with acetone at -20 ℃. The mCherry protein pellets were stored at -80 ℃.

The N-terminal truncated mCherry protein was prepared by adding 0.063 μg Enterokinase Light Chain (NEB) to the intact mCherry protein solution after dialysis. Then the solution was incubated at room temperature overnight. The digestion was passed over Ni-NTA-agarose and the effluent was collected. After acetone precipitation, the N-terminal truncated mCherry protein pellets were stored at -80 ℃.

The cDNA sequences for mCherry were confirmed by dye terminator cycle sequencing using the BigDye Terminator v3.1 Cycle SequencingKit (Applied Biosystems). Sequence analysis was performed at the University of Alberta Molecular Biology Service Unit.

6.2.3 Microwave-Assisted Acid Hydrolysis

For MAAH of proteins, 40 µL of 0.25 mg/mL protein solution was mixed with 0.5 µL of 500 mM DTT (dithiothreitol) and 40 µL of 6 M HCl in a 1.5 mL polypropylene vial. The vial was capped and then placed inside a household microwave oven (Panasonic, London Drugs, Edmonton) with 1200 W output at 2,450 MHz. A container with 100 mL of water was also placed in the microwave oven to absorb the excess microwave energy. The sample was subjected to 60 s microwave irritation, then cooled and dried in a SpeedVac vacuum centrifuge to remove all of the acid. To the vial, 30 µL of 250 mM NH_4HCO_3 was added to reconstitute the sample. Then 1 µL of 500 mM DTT was added and reduction of the disulfide was performed by incubation for 60 min at 37 ℃. After reduction, peptide solution was acidified to pH 2 with 50% TFA and diluted to 105 µL with 0.1% TFA.

6.2.4 LC Fractionation of Hydrolysates

LC fractionation was carried out on a 3 mm i.d. × 15 cm Zorbax 300 - SB C_3 column with a particle size of 3.5 µm and 300 Å pore (Agilent, Mississauga, ON, Canada) in an Agilent 1100 HPLC system (Palo Alto, CA, USA). After loading of

100 µL peptide sample, the column was flushed with 97.5% mobile phase A (0.1% TFA in water) and 2.5% mobile phase B (0.1% TFA in ACN) at a flow rate of 0.4 mL/min for 5 min at room temperature and the salts were effectively removed. Subsequently, the concentration of phase B in the mobile phase was step-wise increased to 30% for 9 min to elute the low molecular weight peptides from the column, followed by 15 min re-equilibration with 97.5% mobile phase A.

6.2.5 MALDI-TOF MS Analysis

The collected fractions were concentrated in SpeedVac, then mixed with the matrix solution of saturated α-cyano-4-hydrocynnamic acid (CHCA) and spotted on a MALDI plate for MALDI analysis by an AB SCIEX 4800 Plus MALDI TOF/TOF Analyzer (AB SCIEX, Foster City, CA, USA) equipped with a diode-pumped Nd:YAG laser at 355 nm. Either a linear or reflector mode was operated for MALDI-TOF mass spectra collection. The peak picking software in the 4800 Plus system was used for automatic peptide ion peak picking and mass assignment.

6.2.6 LC-ESI MS/MS Analysis

The collected fractions were dried in a SpeedVac, reconstituted to 0.2 with 0.1% formic acid and analyzed using a Premier quadrupole time-of-flight (QTOF) mass spectrometer equipped with a nanoACQUITY Ultra Performance LC system (Waters, Milford, MA, USA) [26]. Five µL of peptide solution was injected onto a 75 µm i.d. × 150 mm Atlantis dC_{18} column with 3 µm particle size (Waters). Solvent A consisted of 0.1% formic acid in water, and Solvent B consisted of

0.1% formic acid in ACN. Peptides were first separated at 35 ℃ using 15 min gradients (2%–7% Solvent B for 1 min, 7%–25% Solvent B for 3 min, 25%–50% Solvent B for 2 min, 50%–90% Solvent B for 1 min, 90%–95% Solvent B for 8 min; the column was pre-equilibrated at 2% Solvent B for 20 min) or 30 min gradients (2%–7% Solvent B for 1 min, 7%–25% Solvent B for 19 min, 25%–50% Solvent B for 1 min, 50%–90% Solvent B for 2 min, 90%–90% Solvent B for 6 min, 90%–95% Solvent B for 2 min; the column was pre-equilibrated at 2% Solvent B for 20 min) and electrosprayed into the mass spectrometer (fitted with a nanoLockSpray source) at a flow rate of 350 nL/min. MS scan was performed from m/z 400 to 1600 for 0.8 s, followed by six data-dependent MS/MS scans from m/z 50–1900 for 0.8 s each. The collision energy for MS/MS fragmentation was varied according to the mass and charge state of the eluting peptide. Leucine enkephalin and (Glu1)-fibrinopeptide B, was infused at a rate of 300 nL/min for mass calibration (i.e., lock-mass), and an MS scan was acquired for 1 s every 1 min throughout the run.

6.2.7 Protein Database Search

Raw MS and MS/MS data were corrected by lock-mass, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.3 (Waters). Peptide sequences were identified through database search with peak list files using the MASCOT search program (http://www.matrixscience.com). Database searches were restricted to the BSA protein sequence downloaded from the SwissProt database or mCherry protein sequence translated from its DNA sequence. The following search parameters were selected for all database searching: enzyme,

nonspecified; missed cleavages, 0; peptide tolerance, ±30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, (1+, 2+, and 3+); variable modifications, deamidation of asparagine and glutamine. The search results, including unique peptide sequences, ion score, MASCOT threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses were extracted to Excel files. All the identified peptides with scores lower than the MASCOT identity threshold scores for identity were then deleted from the list.

6.3 Results and Discussion

Figure 6.1 shows the workflow of the MAAH MS method tailored to characterizing the terminal peptides of proteins. A protein sample is subjected to MAAH using 3 M HCl for one minute and the hydrolysate containing many peptides is fractionated using reversed phase liquid chromatography (RPLC). The low molecular weight peptides are eluted at a relatively short retention time and collected for analysis by LC-ESI MS/MS (or MALDI-TOF MS, see below). The acquired mass spectral data are processed to generate the amino acid sequence information on the terminal peptides of a protein sample. During the course of developing this method, each step was examined carefully to generate the optimal results and some of the key findings are highlighted below, followed by an example of the application of this method for characterizing a recombinant protein and its truncated form.

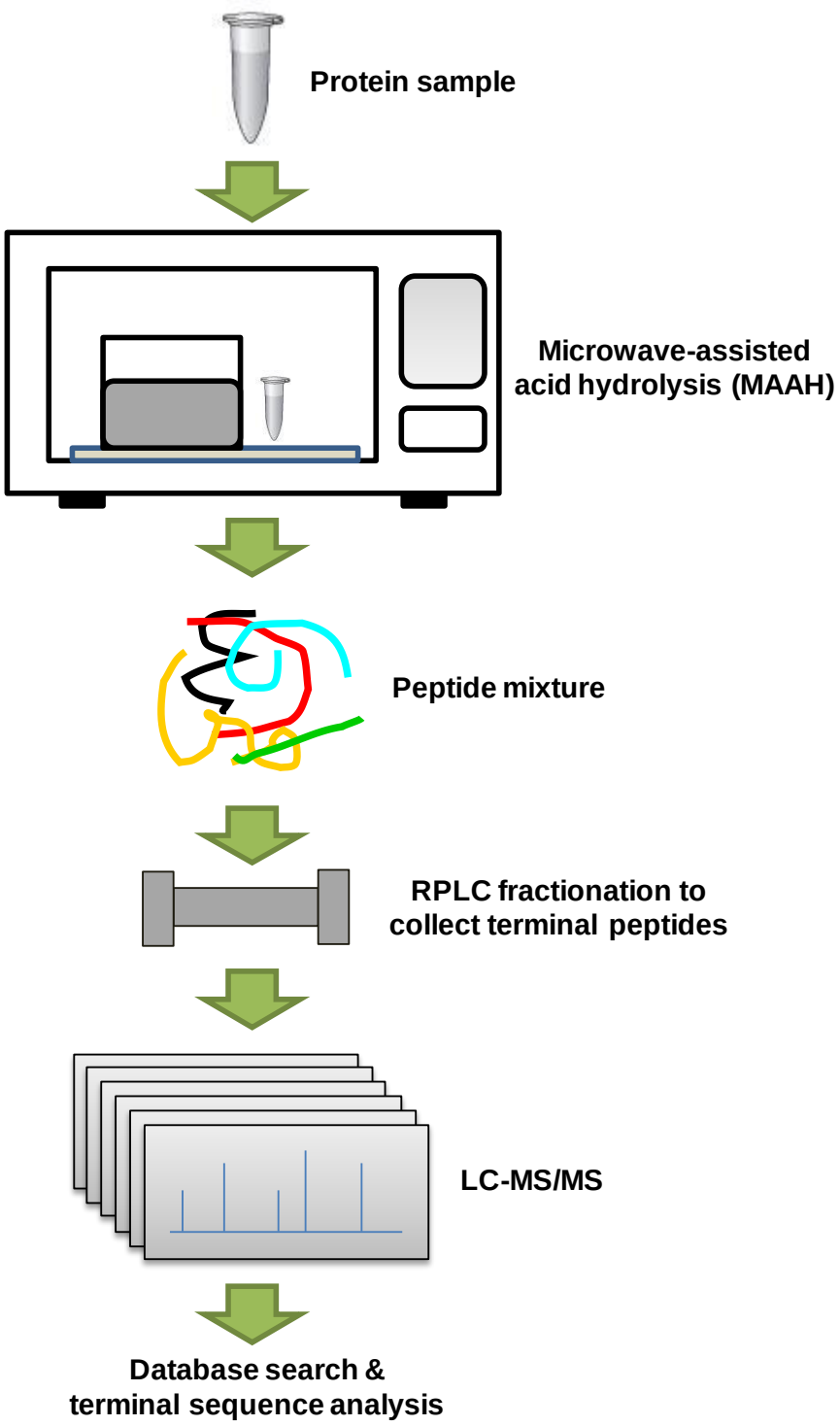


Figure 6.1 Workflow of the HCl MAAH method with isocratic RPLC fractionation to collect the low-molecular-weight peptides containing mainly terminal peptides that are sequenced by LC-ESI MS/MS.

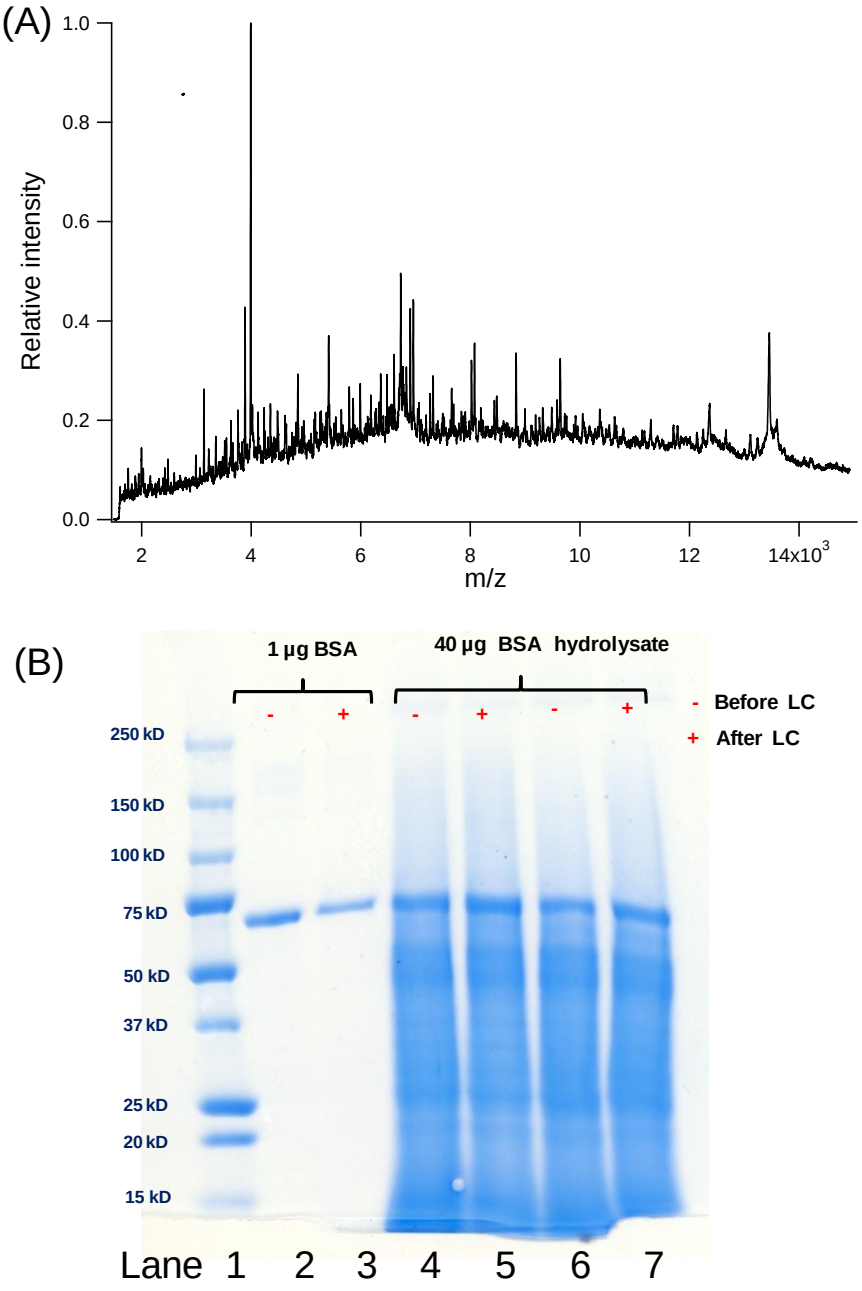


Figure 6.2 (A) MALDI-TOF mass spectrum of BSA after MAAH in 3 M HCl for 1 min irradiation followed by reduction with DTT. (B) SDS-PAGE image of protein standard molecular weight markers (lane 1), 1 µg of BSA before (lane 2) and after (lane 3) passing through the C_3 LC column, 40 µg of the BSA hydrolysate before (lanes 4 and 6) and after (lanes 5 and 7) passing through the C_3 LC column. The proteins and peptides were collected and loaded to the gel after passing through the column.

6.3.1 MALDI-TOF and SDS-PAGE Analysis of Hydrolysates

BSA was used as a model protein for the method development. Figure 6.2A shows the MALDI-TOF mass spectrum of the BSA hydrolysate. About 150 peaks were found in the spectrum. Among them, 85 peptides (57%) could be unambiguously matched with the terminal peptides of BSA within the relative mass error of $\pm 0.05\%$. Because of the presence of internal peptides, it is difficult to read out the protein sequence directly by the mass differences of adjacent peaks. This is in contrast with the smaller proteins (<20,000 Da) where the entire sequence of a protein can be read directly from the N- and C-terminal peptides that usually dominant the MALDI mass spectrum. As the protein size increases, more internal peptides are generated that have similar masses as the terminal peptides. Moreover, MALDI detection of larger polypeptides in the presence of many smaller peptides in the hydrolysate becomes less sensitive and accurate. As a result, only a few terminal peptides, if any, can be accurately determined from the hydrolysate of larger proteins.

The MALDI mass spectrum of BSA hydrolysate does not show many peaks at the high mass range (m/z >14,000), due to ion suppression and reduced sensitivity for detecting larger polypeptides by MALDI-TOF. However, SDS-PAGE analysis of the hydrolysate (see Figure 6.2B, lanes 4-7) shows a wide range of polypeptides with the molecular weights of up to ~67 kDa. Since we are interested in sequencing the terminal peptides, we decided to use HPLC to fractionate the low mass peptides to improve their detectability.

6.3.2 Column Selection for Peptide Fractionation

As the BSA hydrolysate contains many polypeptides with a wide range of molecular weights, a suitable HPLC fractionation method is needed to effectively remove the high molecular weight components while collecting the low molecular weight peptides with high efficiency and recovery. At first, a Varian Polaris C_8 column (1 mm i.d. × 50 mm with a particle size of 3 μm diameter and 180 Å pores) was used. It was found that the pressure applied to the column increased after running the BSA hydrolysate, even after extensive washing. SDS-PAGE analysis of the BSA hydrolysate before and after running this C_8 column clearly showed the loss of the BSA and large peptides in this column (data not shown). It is likely that the small pore size and long carbon chain of the stationary phase caused strong retention of the protein and large peptides. We then tested the Agilent Zorbax 300SB C_3 column (3 mm i.d. × 150 mm with a particle size of 3.5 μm diameter and 300 Å pores). Lane 4 in Figure 6.2B shows the SDS-PAGE image of the BSA hydrolysate before LC separation, while Lane 5 shows the image of the hydrolysate collected after running it through the column. Lanes 6 and 7 are the duplicate results. The four lanes are similar, indicating that there was little loss of BSA hydrolysates in the column. For BSA, sample loss, if any, was found to be mainly from the adsorption to the sample vial during the sample preparation step. For example, when 1 μg BSA was used, the SDS-PAGE image in Figure 6.2B (lanes 2 and 3) shows some losses after the C_3 column separation. However, when more than 5 μg BSA was used, the recovery rate of BSA protein and BSA hydrolysates in this column was found to be 100%, which was

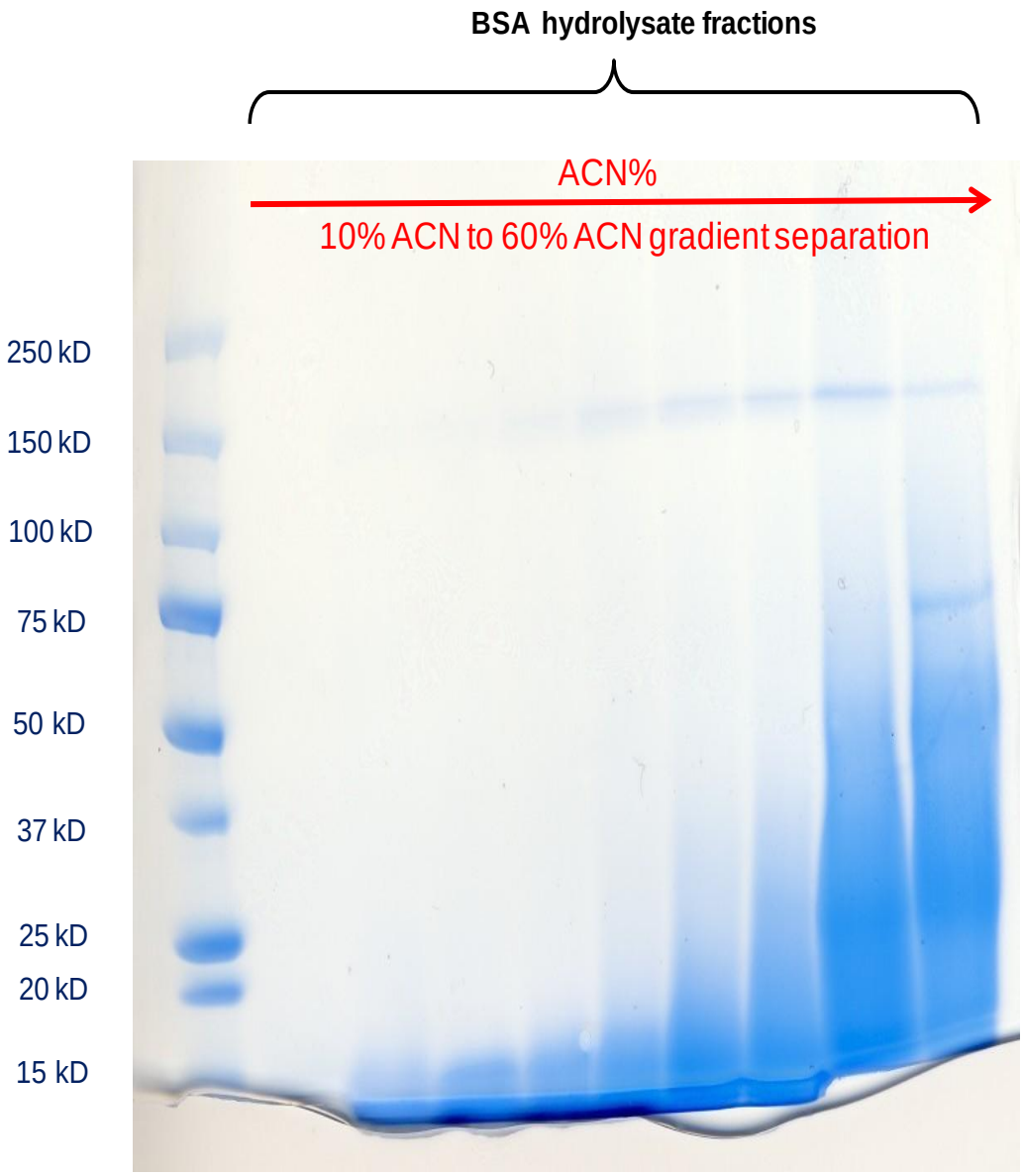


Figure 6.3 SDS-PAGE image of eight peptide fractions collected from the LC separation of 160 µg of the BSA hydrolysate using a gradient elution with varying organic solvent composition.

confirmed by both LC-UV and SDS-PAGE analysis.

6.3.3 LC Fractionation

To check the separation efficiency, eight fractions were collected from the LC experiment using the C_3 column with a gradient gradually increased from low percentage of organic solvent to high organic. Figure 6.3 shows the SDS-PAGE results of the eight fractions collected. It is clear that the low-molecular-weight peptides can be separated from the rest of the polypeptides in the low organic solvent fractions. To speed up the fractionation process, isocratic separation was then used to collect only the low molecular weight peptides. Four different levels of organic solvent (25% B_1, 30% B_1, 35% B_1 and 40% B_1) were tested for isocratic separations and the collected fractions were analyzed by MALDI-TOF. Figure 6.4 shows the MALDI mass spectra of these fractions. It is apparent that the use of 35% or 40% B_1 could elute many large peptides with molecular weights of >4 kDa. The use of 25% B_1 and 30% B_1 eluted many smaller peptides. Based on the LC-UV results, a larger amount of peptides were collected from the 30% B_1 run than the 25% B_1 run. Thus, 30% B_1 isocratic separation was used in the subsequent experiments to collect the small peptides from the hydrolysate for MS analysis.

Figure 6.5 shows the reflectron MALDI-TOF mass spectrum of the 30% B_1 fraction of the BSA hydrolysate. The spectrum contains only the peaks from m/z 1000 to 3000. The larger peptides were not efficiently detected in the reflectron mode and the overall spectrum looks much cleaner than that shown in Figure 6.4B. There are 14 peaks matched with the terminal peptides of BSA within m/z ± 0.3

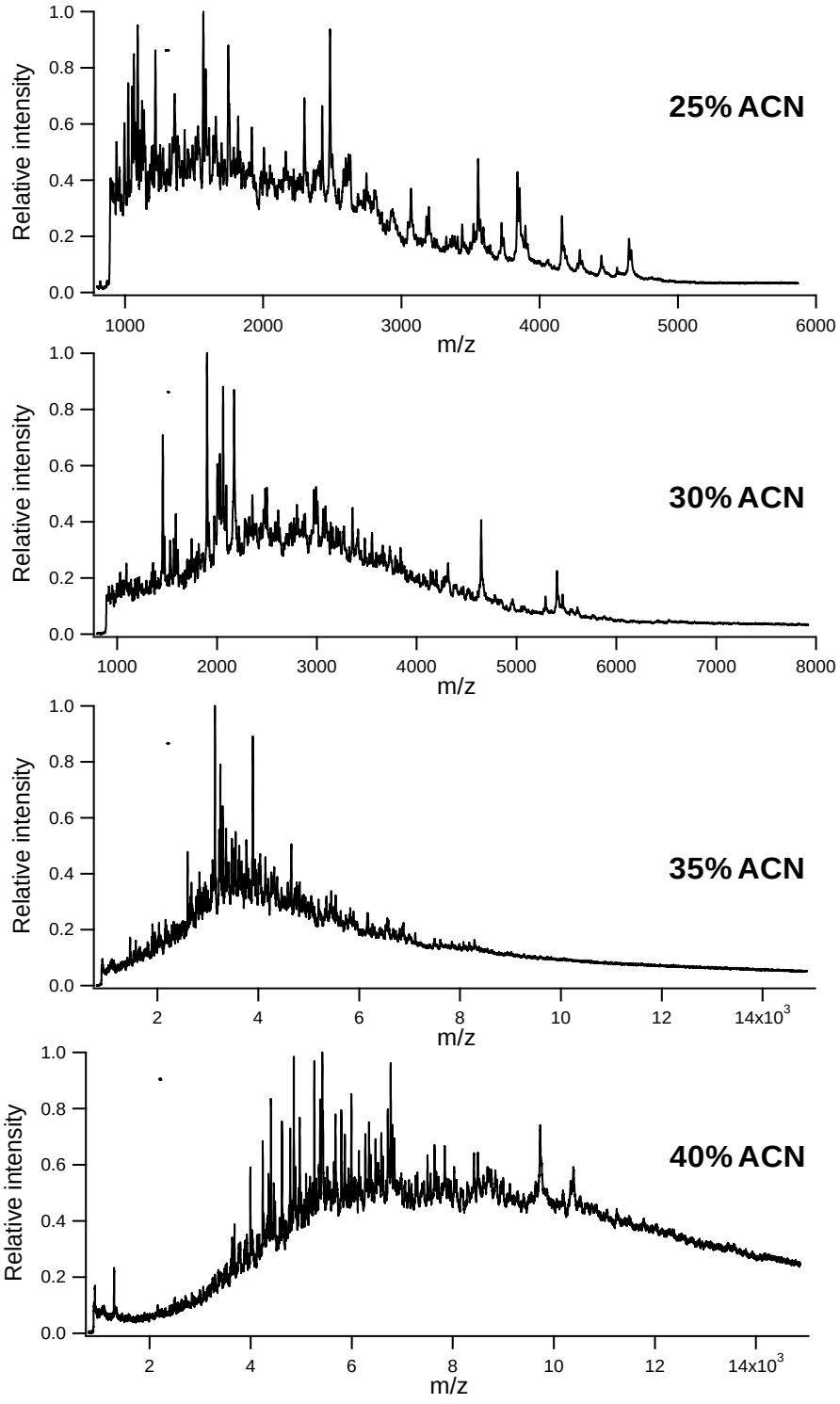


Figure 6.4 MALDI-TOF mass spectra of the BSA hydrolysate eluted at (A) 25% B_1, (B) 30% B_1, (C) 35% B_1 and (D) 40% B_1 from the C_3 column (A_1: 0.1% TFA in water and B_1: 0.1% TFA in ACN).

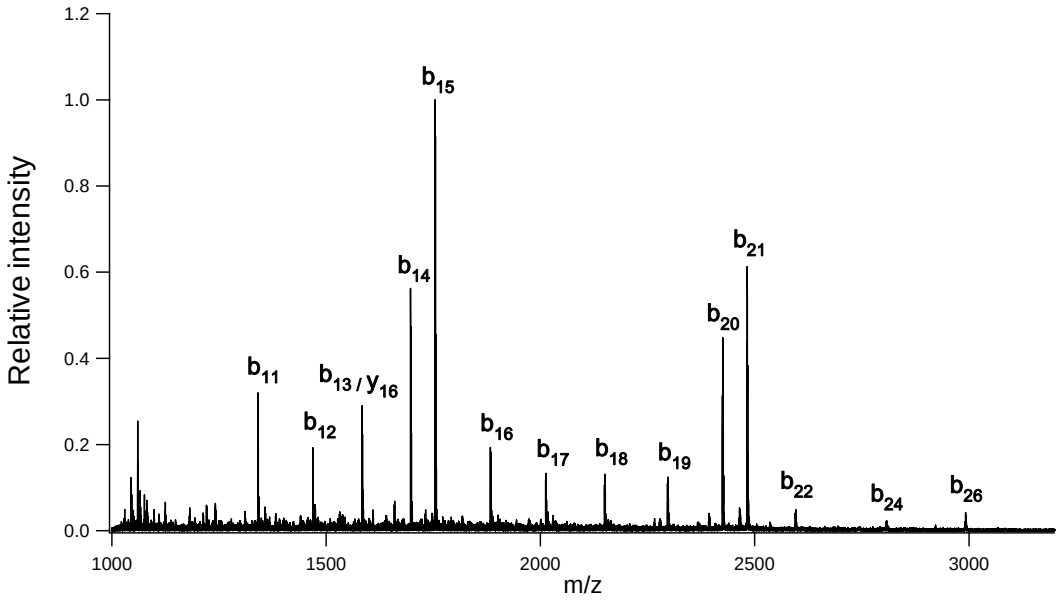


Figure 6.5 Reflectron mode MALDI-TOF mass spectrum of the low-molecular-weight peptide fraction collected from 30% B_1 isocratic elution of the BSA hydrolysate injected onto the C_3 column.

Da. The intensities of all these matched peaks were much higher than their adjacent peaks, making them easy to distinguish from other noise or peptide peaks in the spectrum. The peak of m/z 1584.02 matched with y_{16} or b_{13} within ± 0.2 Da or ± 150 ppm (b and y notations are used to refer to the terminal peptide ions from the N-termini and C-termini, respectively, although they are not the usual MS/MS fragment ions). The mass differences of peptides y_{16} and b_{13} are only 0.1 Da (63 ppm), so we cannot tell which one is correct by MS only. However, since all the other ions are from the b-ion series, the m/z 1584.02 peak was most likely from peptide b_{13}. Thus, the peptide ladder from b_{11} to b_{22} was complete and we could read out the amino acid composition directly by the mass differences of the two adjacent peaks. The fact that almost all the matched peaks belong to the N-

terminal peptides suggests that the C-terminal peptides of BSA were severely suppressed by the N-terminal peptides. This detection bias towards the N-terminal peptides is very likely due to different ionization efficiencies for the terminal peptides. Thus, we examined an alternative method of using LC-ESI MS/MS to see if we could detect more terminal peptides with less ion suppression or detection bias.

6.3.4 Optimization of LC-ESI MS/MS

For LC-ESI MS/MS analysis, the LC gradient was optimized for analysis of the BSA hydrolysate fraction with a goal of detecting as many terminal peptides as possible in the shortest time. Figure 6.6A compares 15 min and 30 min gradient experiments. No detection bias towards one-end of the protein was observed and many more terminal peptides were detected, compared to MALDI MS. A longer LC gradient resulted in more terminal peptides detected as well as more internal peptides identified. In the 15 min LC gradient experiment, over 60% of the peptides detected were terminal peptides, while only around 20% peptides were terminal peptides for the 30 min LC gradient. Although internal peptides were detected, the average matching score of the internal peptides was much lower than that of the terminal peptides, which also suggests that the terminal peptides were dominant in the RPLC fraction of the hydrolysate. An even longer gradient (e.g., 50 min) generated similar results as the 30 min gradient experiment. The small peptide fraction was a relatively simple mixture and there was no issue of under-sampling for the 30 min run. Thus, it was concluded that a 30-min gradient LC-ESI MS/MS run was sufficient for detecting the terminal peptides.

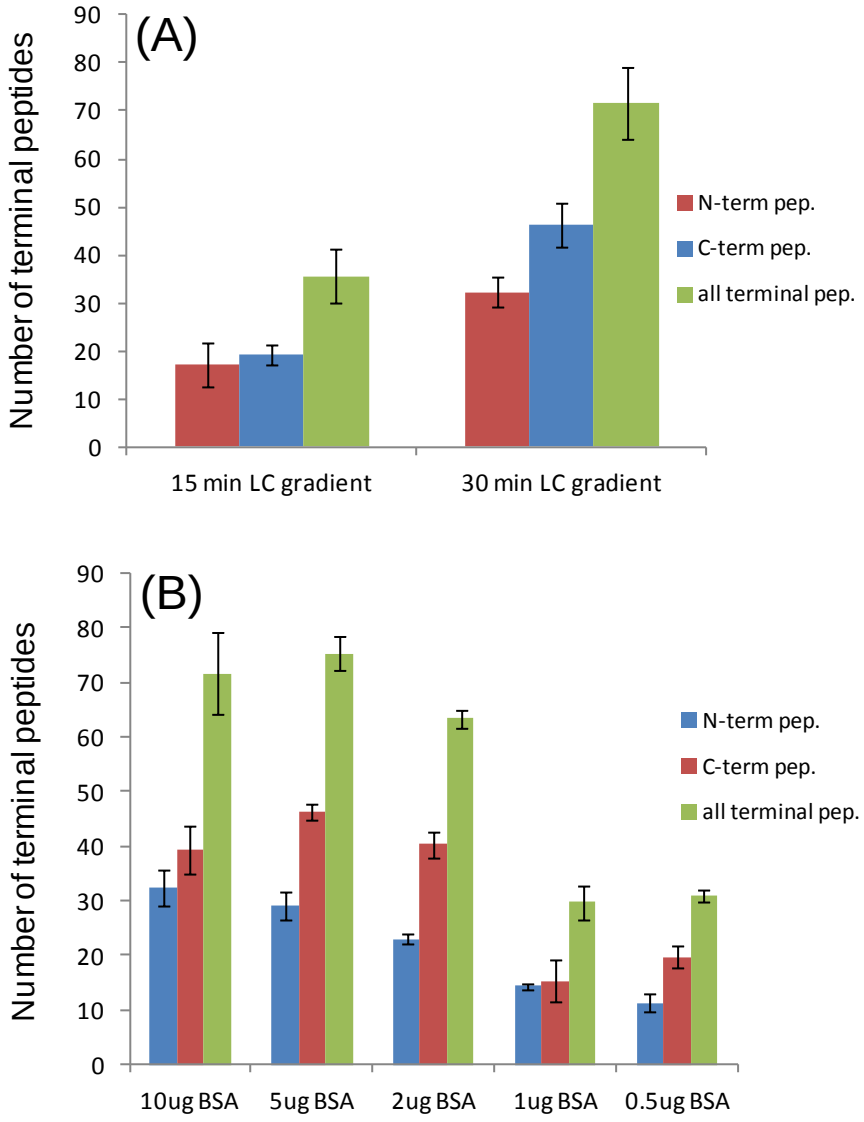


Figure 6.6 The number of terminal peptides identified from (A) 15- and 30-min gradient LC-ESI MS/MS (n=3) analysis of the low mass fractions from HCl MAAH of 10 µg of BSA and (B) using different amount of BSA with 30-min gradient LC-ESI MS/MS (n=3).

6.3.5 Effect of Sample Amount

Figure 6.6B shows the number of terminal peptides identified from the RPLC fraction of the BSA hydrolysate using a starting protein amount of 10 μg, 5 μg, 2 μg, 1 μg or 0.5 μg. In this case, the results of 10, 5 and 2 μg BSA were similar.

Although 1 μg and 0.5 μg BSA only identified about half of the number of the terminal peptides of the 10 μg sample, more than 10 peptides for each terminus of the protein could still be identified.

Finally, it should be noted that the total experiment time for the entire workflow is about 7 hours: 1 min for MAAH, 1 h for DTT reduction, 1 h for evaporating the acid using a SpeedVac, 30 min for RPLC fractionation, 4 h for drying the collected fraction using the SpeedVac, 30 min for LC-ESI MS/MS analysis and about 15 min for data analysis. The sample drying steps took the most time (about 5 hours). All of the experiment procedures including data analysis are quite straightforward and do not require any special skills.

6.3.6 Applications

The application of the above described method is illustrated in an example of characterizing a recombinant protein and its truncated form. For a recombinant protein, sequence confirmation is important to ensure that the product is the intended protein. In this case, the MAAH shotgun protein sequencing method was applied to characterize mCherry protein and its N-terminal truncated form. Figure 6.7A shows the protein sequence of mCherry provided to us. After protein expression and affinity purification using a His-tag column, the protein sample

(A) Original sequence

MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSRMVSKGEEDNMAIIK
EFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDIL
SPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSL
QDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRL
KLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAE
GRHSTGGMDELYK

(B) Corrected sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPMVSKGEEDNMAIIKEFM
RFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQ
FMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDG
EFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLK
DGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRH
STGGMDELYK

(C) Truncated protein sequence

DPMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAK
LKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMN
FEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSER
MYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITS
HNEDYTIVEQYERAEGRHSTGGMDELYK

(D) Unexpected by-product sequence

DLYDDDDKDPMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPY
EGTQTAKLKVTKGGPLPFAWDILSPQF||MYGSKAYVKHPADIPDYLKLSFPEGF
KWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMG
WEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNV
NIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK

Figure 6.7 (A) Original and (B) corrected sequence of the mCherry protein showing the sequence coverage from the peptides generated using TFA MAAH LC-ESI MS/MS. (C) Corrected sequence and sequence coverage of the truncated mCherry protein obtained using TFA MAAH LC-ESI MS/MS; the missing MRG sequence was covered by an N-terminal peptide identified using HCl MAAH, RPLC fractionation and LC-ESI MS/MS. (D) Sequence of an unexpected by-product in the truncated mCherry protein sample. The DLYDDDK sequence was determined using HCl MAAH, RPLC fractionation and LC-ESI MS/MS. The "||" indicates where the cleavage of the peptide bond can occur to form two large peptides (see text).

was subjected to HCl MAAH, followed by RPLC fractionation and LC-ESI MS/MS analysis. When we searched the MS/MS data against the database containing the mCherry protein sequence provided, more than 50 C-terminal peptides were detected. However, no N-terminal peptides were found. Repeated experiments gave the same results. Because all the other proteins tested including several proprietary proteins from Sanofi Pasteur yielded both terminal peptides,we suspected that the protein sequence given to us was incorrect.

To generate further evidence on the possibility of incorrect sequence, the mCherry protein was analyzed using TFA MAAH which could hydrolyze a protein to small peptides with molecular weights of generally less than 3000 Da using 10 min microwave irradiation. This method could potentially cover almost 100% sequence of proteins.[24] However, TFA MAAH generates a hydrolysate that is much more complex than that of HCl MAAH and contains terminal peptides as well as many internal peptides. In some cases, the terminal peptides cannot be identified due to ion suppression or low abundance. In comparison, HCl MAAH generates terminal peptides predominately which is good for terminal peptide analysis.

Figure 6.7A shows the sequence coverage obtained from the peptides generated using TFA MAAH of the mCherry protein with a two-hour-gradient LC-ESI MS/MS run. The first three amino acids, “MGG”, the 33rd to 37th amino acids, “PSSRM”, and the 107th to 108th amino acids, “MY”, could not be mapped by the peptides in the TFA MAAH hydrolysate. The “MYG” sequence involves in the formation of a five-member ring to form a chromophore that gives the

characteristic color of mCherry and the "Y" in this sequence was dehydrated [27-29]. If dehydration of Y was added as the modification in the database search, "MY" could be covered by several peptides. Thus, there appeared to be no problem with this region of the protein sequence.

In the meantime, ESI-TOF analysis was performed to determine the molecular masses of the protein and its truncated form. It was found that the measured molecular mass of the intact protein was 253 Da less than the calculated mass according to the sequence provided. The measured mass of the truncated protein was 351 Da less than the calculated mass. Thus, there must be some sequence errors in "MGG" and/or "PSSRM".

In order to confirm this finding, the plasmid containing the mCherry gene used in the culture was sent to the University of Alberta DNA sequencing facility for sequencing. It was found that "MGG" is actually "MRG" in the protein sequence and "PSSRM" should be "PM". The corrected sequence is shown in Figure 6.7B. The calculated mass of this new protein sequence matched with the measured mass of the intact protein. The initial sequence of mCherry was then replaced with the correct sequence in the database. All the TFA MAAH MS/MS data were re-searched against the new database. Now the peptides of the intact protein hydrolyzed by TFA only missed the "MRG" part in the sequence, not counting for the "MY" chromophore part.

Since MRG was not covered by the TFA MAAH LC-MS/MS method, we applied the HCl MAAH method to the intact mCherry protein. Only one N-terminal peptide, MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKD, was found

with high confidence (Mascot matching score of ~100). This peptide confirmed the N-terminal sequence. Apparently, the formation of this peptide from the hydrolysis of a very weak D-P amide bond of the protein dominated the HCl hydrolysis process that prevented the formation of other N-terminal peptides. Note that this peptide is so long that it would be expected to be hydrolyzed to smaller peptides in TFA MAAH. We tried to collect the low organic fractions of TFA hydrolysates by the same C_3 column. Generally for the low organic fractions of TFA hydrolysates, most of the peptides identified were the internal peptides. In this particular case, no N-terminal peptide was found. Thus, HCl MAAH is complementary to TFA MAAH for generating the whole protein sequence information. Regarding the issue of sequence errors, it was suspected that the plasmid might have altered the sequence during many runs of culturing since it was produced many years ago.

The characterization of the truncated form of mCherry also illustrates the importance of HCl MAAH for protein sequencing. In this case, enterokinase was used to cut after the sequence DDDDK [30] to remove the His tags and other amino acids at the N-terminal of the intact protein (see Figure 6.7C for the sequence). There were 13 N-terminal peptides detected using HCl MAAH with RPLC fractionation and LC-ESI MS/MS (lists of identified peptides from the intact and truncated proteins are shown in Tables 6.1 and 6.2), confirming the N-terminal sequence. However, in addition to these terminal peptides, there were three other peptides containing "DLYDDDDK" in their sequences, which is located before the cutting site of enterokinase. These peptides were not identified in the TFA hydrolysates. At first we thought they might come from the hydrolysates of the

Table 6.1 List of unique terminal peptides identified from the intact mCherry protein.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
1 - 32	MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKD	3659.49	0.0141	3.85	100.18
265 - 269	DELYK	666.3224	0.0032	4.8	26.87
264 - 269	MDELYK	797.3629	0.0072	9.0	29.65
263 - 269	GMDELYK	854.3844	-0.0143	-16.7	33.25
262 - 269	GGMDELYK	911.4059	0.011	12	46.92
261 - 269	TGGMDELYK	1012.454	0.0181	17.9	63.29
260 - 269	STGGMDELYK	1099.486	0.0144	13.1	58.94
259 - 269	HSTGGMDELYK	1236.545	0.0078	6.3	84.53
258 - 269	RHSTGGMDELYK	1392.646	0.0043	3.1	34.94
257 - 269	GRHSTGGMDELYK	1449.667	0.0126	8.69	53.28
256 - 269	EGRHSTGGMDELYK	1578.71	0.0118	7.47	70.89
255 - 269	AEGRHSTGGMDELYK	1649.747	0.0203	12.3	73.01
253 - 269	ERAEGRHSTGGMDELYK	1934.89	0.0028	1.4	64.26
252 - 269	YERAEGRHSTGGMDELYK	2097.954	0.0108	5.15	51.61
251 - 269	QYERAEGRHSTGGMDELYK	2226.012	0.0168	7.55	30.45
248 - 269	IVEQYERAEGRHSTGGMDELYK	2567.207	-0.0101	-3.93	43.69
247 - 269	TIVEQYERAEGRHSTGGMDELYK	2669.239	0.0054	2.0	62.7

246 - 269	YTIVEQYERAEGRHSTGGMDELYK	2831.318	0.0337	11.9	83.01
245 - 269	DYTIVEQYERAEGRHSTGGMDELYK	2946.345	0.0462	15.7	65.87
244 - 269	EDYTIVEQYERAEGRHSTGGMDELYK	3075.388	0.0393	12.8	67.42
243 - 269	NEDYTIVEQYERAEGRHSTGGMDELYK	3190.415	0.0336	10.5	97.54
239 - 269	ITSHNEDYTIVEQYERAEGRHSTGGMDELYK	3627.654	0.0477	13.1	29.93

Table 6.2 List of unique terminal peptides identified from the truncated mCherry protein.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
32 - 42	DPMVSKGEEDN	1220.4867	-0.0016	-1.3	33.45
32 - 43	DPMVSKGEEDNM	1350.5432	0.0077	5.7	78.6
32 - 44	DPMVSKGEEDNMA	1421.5803	0.0078	5.5	57.73
32 - 47	DPMVSKGEEDNMAIIK	1775.8433	0.0165	9.29	116.35
32 - 48	DPMVSKGEEDNMAIIKE	1904.8859	0.0037	1.9	70.71
32 - 49	DPMVSKGEEDNMAIIKEF	2051.9543	0.0155	7.55	23.8
32 - 51	DPMVSKGEEDNMAIIKEFMR	2339.0959	0.0065	2.8	169.22
265 - 269	DELYK	666.3224	0.0048	7.2	30.14
264 - 269	MDELYK	797.3629	0.0088	11	30.21
263 - 269	GMDELYK	854.3844	-0.0251	-29.4	26.39

262 - 269	GGMDELYK	911.4059	-0.014	-15	42.22
261 - 269	TGGMDELYK	1012.454	-0.0061	-6.0	63.81
260 - 269	STGGMDELYK	1099.486	0.0159	14.5	64.23
259 - 269	HSTGGMDELYK	1236.545	0.005	4	89.71
258 - 269	RHSTGGMDELYK	1392.646	0.0071	5.1	77.09
257 - 269	GRHSTGGMDELYK	1449.667	-0.0247	-17.0	95
256 - 269	EGRHSTGGMDELYK	1578.71	0.0078	4.9	84.77
255 - 269	AEGRHSTGGMDELYK	1649.747	-0.0137	-8.30	83.87
254 - 269	RAEGRHSTGGMDELYK	1805.848	-0.0086	-4.8	28.49
253 - 269	ERAEGRHSTGGMDELYK	1934.89	-0.0287	-14.8	63.75
252 - 269	YERAEGRHSTGGMDELYK	2097.954	0.0006	0.3	63.1
251 - 269	QYERAEGRHSTGGMDELYK	2226.012	0.0072	3.2	36.52
248 - 269	IVEQYERAEGRHSTGGMDELYK	2567.207	-0.0185	-7.21	22.45
247 - 269	TIVEQYERAEGRHSTGGMDELYK	2668.255	0.0146	5.47	86.84
246 - 269	YTIVEQYERAEGRHSTGGMDELYK	2831.318	0.0013	0.46	79.59
245 - 269	DYTIVEQYERAEGRHSTGGMDELYK	2946.345	0.0035	1.2	29.98
244 - 269	EDYTIVEQYERAEGRHSTGGMDELYK	3075.388	0.0025	0.81	33.91

intact proteins which were not removed completely by the Ni-NTA-agarose when the flow-through of the truncated protein was collected. After re-purifying the flow-through by Ni-NTA-agarose, the DLYDDDDK-containing peptides were still detected. We then proposed that there might be some other by-product present in the sample that was generated by the enzyme. This product did not contain the his-tag and thus it would not be removed by Ni-NTA-agarose.

We used MALDI MS and SDS-PAGE to support our hypothesis. Figure 6.7D shows the sequence of the proposed protein by-product. In acidic or basic condition or being heated, the peptide bond near the chromophore of the mCherry protein would hydrolyze to two peptides,[29] as indicated in Figure 6.7D. For the N-terminal truncated mCherry protein, the molecular weights of the two peptides should be 8070 and 18862, while for the protein by-product, the molecular weights of the two peptides are 9050 and 18862. Figure 6.8 shows the SDS-PAGE image of the intact and N-terminal truncated proteins. When the loading amount was larger than 2 μg, the unique hydrolyzed peptide of the protein by-product could be seen. Figure 6.9 shows the MALDI spectrum of the N-terminal truncated mCherry protein sample in the m/z range of 7000 to 9500. The peak intensity of the unique peptide of the protein by-product was rather low likely due to the ion suppression of the major protein. Both MALDI and SDS-PAGE results supported the presence of the impurity protein starting with "DLYDDDK". Thus, we concluded that there was a by-product formed from the enzyme digestion with a sequence shown in Figure 6.7D.

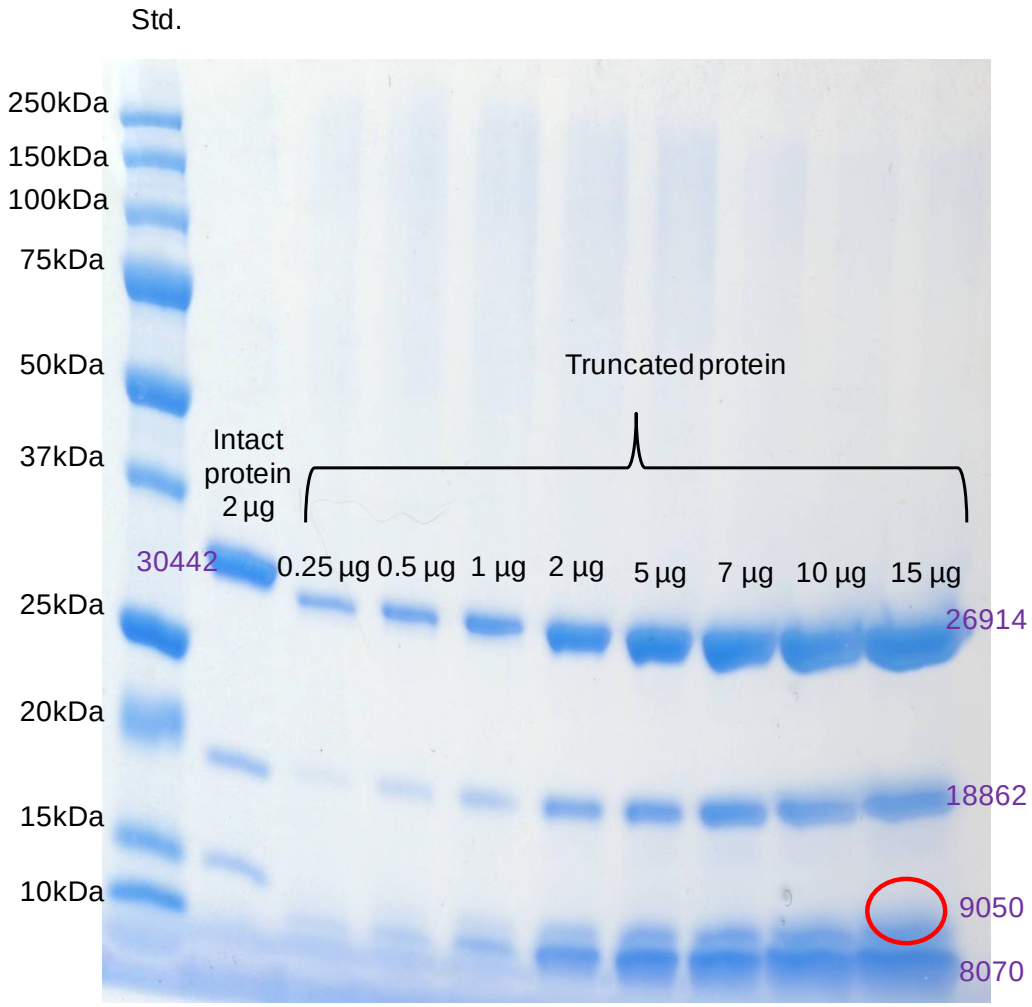


Figure 6.8 SDS-PAGE images. The calculated mass of the protein or peptide is indicated for the corresponding gel band.

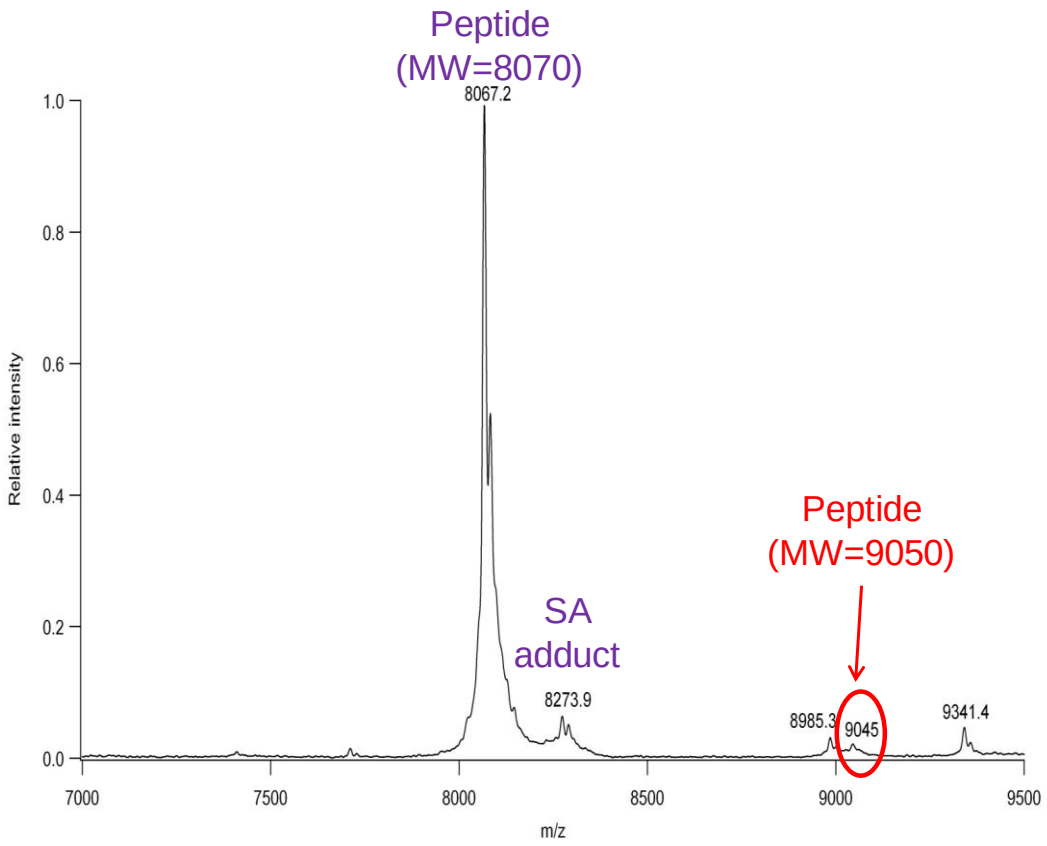


Figure 6.9 MALDI-TOF mass spectrum of the truncated protein sample. Sinapinic acid (SA) was used as the matrix. The peak at m/z 8985.3 is likely from the triply charged ion of the intact molecular ion (26914+Na+Na+H). The peak at m/z 9341.4 cannot be assigned; it may be from an impurity in the sample or adduct ions from the MALDI analysis.

The above results on the characterization of a recombinant protein and its truncated form indicate that the HCl MAAH method with RPLC fractionation and LC-ESI MS/MS analysis is a powerful tool for characterizing terminal peptides of

a protein. By combining HCl and TFA MAAH, the entire sequence of the mCherry protein as well as the truncated protein was confirmed. In the case of the truncated protein produced by enterokinase treatment of the intact mCherry protein, the HCl MAAH method allowed the identification of a protein by-product.

While it is not shown herein, there were several other examples of successful characterization of terminal peptides of proprietary proteins with intended or unintended modifications of terminal sequences that were provided to us by our collaborators at Sanofi Pasteur. We believe that the HCl MAAH method described in this work is a useful tool for detailed analysis of amino acid sequences and modifications of the termini of a protein or a simple mixture of proteins.

6.4 Conclusions

We have developed a method for sequencing N- and C-terminus of a protein. BSA was used as a model protein to optimize the process and demonstrate the performance of the method. It was shown that RPLC fractionation simplified the protein hydrolysate generated by HCl MAAH and the collected low-molecular-weight peptides were suitable for LC-ESI MS/MS analysis. Compared to direct MALDI-TOF MS analysis, LC-ESI MS/MS is more appropriate for terminal peptide sequencing due to its less biased detection and improved identification of the peptides with MS/MS. The total analysis time is less than 7 hours, including 5 hours of sample drying steps. The analysis can be done without the need of a

highly skilled researcher for sample handling and data analysis, as the experimental protocol and data interpretation are straightforward. This method should be useful for rapid characterization of terminal sequences of proteins, such as for quality control of proteins during the production process and storage as well as development of new products where modifications of terminal amino acids may be used for enhancing chemophysical or therapeutic properties.

6.5 Literature Cited

(1) Chevigne, A.; Fievez, V.; Schmit, J. C.; Deroo, S. *Biochem. Pharmacol.* **2011**, *82*, 1438.
(2) Lange, P. F.; Overall, C. M. *Curr. Opin. Chem. Biol.* **2013**, *17*, 73.
(3) Peng, D. G.; Satterlee, J. D.; Ma, L. H.; Dallas, J. L.; Smith, K. M.; Zhang, X. H.; Sato, M.; La Mar, G. N. *Biochemistry* **2011**, *50*, 8823.
(4) Kim, J. Y.; Lee, J. R.; Choi, S.; Kim, E. M.; Jung, N. K.; Kim, Y. H.; Yoo, J. S.; Lee, S. W. *Rapid Communicantion in Mass Spectrometry* **2013**, *27*, 521.
(5) Christiansen, C.; Skotland, T. *Biotechnol. Appl. Biochem.* **2010**, *55*, 121.
(6) Edman, P.; Begg, G. *Eur J Biochem* **1967**, *1*, 80.
(7) Stark, G. R. *Biochemistry-Us* **1968**, *7*, 1796.
(8) Boyd, V. L.; Bozzini, M.; Guga, P. J.; Defranco, R. J.; Yuan, P. M.; Loudon, G. M.; Nguyen, D. *J Org Chem* **1995**, *60*, 2581.
(9) Schar, M.; Bornsen, K. O.; Gassmann, E. *Rapid Communications in Mass Spectrometry* **1991**, *5*, 319.
(10) Patterson, D. H.; Tarr, G. E.; Regnier, F. E.; Martin, S. A. *Anal Chem* **1995**, *67*, 3971.
(11) Samyn, B.; Sergeant, K.; Castanheira, P.; Faro, C.; Van Beeumen, J. *Nat Methods* **2005**, *2*, 193.
(12) Switzar, L.; Giera, M.; Niessen, W. M. A. *J. Proteome Res.* **2013**, *12*, 1067.
(13) Resemann, A.; Wunderlich, D.; Rothbauer, U.; Warscheid, B.; Leonhardt, H.; Fuchser, J.; Kuhlmann, K.; Suckau, D. *Anal Chem* **2010**, *82*, 3283.
(14) Calligaris, D.; Villard, C.; Lafitte, D. *J. Proteomics* **2011**, *74*, 920.
(15) Zhang, H.; Ge, Y. *Circ.-Cardiovasc. Genet.* **2011**, *4*, 711.
(16) Dormeyer, W.; Mohammed, S.; van Breukelen, B.; Krijgsveld, J.; Heck, A. J. R. *J. Proteome Res.* **2007**, *6*, 4634.
(17) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. *Nature Biotechnology* **2003**, *21*, 566.
(18) Kleifeld, O.; Doucet, A.; Keller, U. A. D.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. *Nature Biotechnology* **2010**, *28*, 281.

(19) Schilling, O.; Barre, O.; Huesgen, P. F.; Overall, C. M. *Nat Methods* **2010**, *7*, 508.
(20) Zhong, H. Y.; Zhang, Y.; Wen, Z. H.; Li, L. *Nature Biotechnology* **2004**, *22*, 1291.
(21) Zhong, H. Y.; Marcus, S. L.; Li, L. *J Am Soc Mass Spectr* **2005**, *16*, 471.
(22) Sandoval, W. N.; Pham, V.; Ingle, E. S.; Liu, P. S.; Lill, J. R. *Comb. Chem. High Throughput Screen* **2007**, *10*, 751.
(23) Lill, J. R.; Ingle, E. S.; Liu, P. S.; Pham, V.; Sandoval, W. N. *Mass Spectrom Rev* **2007**, *26*, 657.
(24) Wang, N.; Li, L. *J Am Soc Mass Spectr* **2010**, *21*, 1573.
(25) Wu, F.; Sun, D. F.; Wang, N.; Gong, Y.; Li, L. *Anal Chim Acta* **2011**, *698*, 36.
(26) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Anal Chem* **2009**, *81*, 1049.
(27) Yarbrough, D.; Wachter, R. M.; Kallio, K.; Matz, M. V.; Remington, S. J. *P Natl Acad Sci USA* **2001**, *98*, 462.
(28) Pletnev, S.; Subach, F. V.; Dauter, Z.; Wlodawer, A.; Verkhusha, V. V. *J Am Chem Soc* **2010**, *132*, 2243.
(29) Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldridge, K. K.; Tsien, R. Y. *P Natl Acad Sci USA* **2000**, *97*, 11990.
(30) Shahravan, S. H.; Qu, X. L.; Chan, I. S.; Shin, J. A. *Protein Expres Purif* **2008**, *59*, 314.

Chapter 7

Comparison of Household and Commercial Microwave Oven in Microwave-Assisted Acid Hydrolysis (MAAH)

7.1 Introduction

Microwave-assisted acid hydrolysis (MAAH) of proteins combined with mass spectrometry (MS) has been developed as an alternative method to the traditional top-down or bottom-up proteomics apporach for protein sequence analysis. Within several minutes, the proteins are digested to peptides with the help of acid and microwave irradiation and peptides are analyzed by MS. Zhong *et al.*[1] first introduced a microwave-assisted acid hydrolysis (MAAH) method to identify the termini of proteins using hydrochloric acid (HCl). A protein of interest was hydrolyzed to terminal peptides in 6 M HCl by microwave irradiated for a short time (i.e., 1 min). By analyzing these polypeptide ladders, the complete sequence information as well as the post-translational modification (PTM) was obtained. In addition to HCl, trifluroacetic acid (TFA) is also used for protein degradation. Wang *et al.*[2] microwave-irridated the protein BSA in 25% TFA for about 10 min, and then analyzed the hydrolyzed peptides by LC-ESI QTOF MS/MS. 100% sequence coverage was achieved. They successfully identified the phosphorylation positions of bovine α_{S1}-casein variant B protein using this method. It is worthwhile to mention that both HCl and TFA MAAH hydrolyze peptide bonds nonspecifically, while microwave combined with diluted formic

acid produce aspartyl-specific cleavages.[3-7]

Until now, MAAH for whole sequence analysis has been mainly done with a household microwave oven. For reasons such as regulatory approval and good laboratory practices, a commercial microwave device tailored to perform microwave-assisted chemistry is a preferred choice. Currently, commercial systems are used for other microwave-assisted techniques such as trypsin digestion[8-10], but have not been evaluated extensively for MAAH for whole protein sequence analysis.

In this chapter, we compare the performance of a commercial microwave device with an optimized household MAAH method, and then apply both methods for analyzing a recombinant protein to illustrate the applicability of MAAH MS for protein sequence analysis.

7.2 Experimental

7.2.1 Chemicals and Reagents

Dithiothreitol (DTT), ammonium bicarbaonate (NH_4HCO_3), trifluoroacetic acid (TFA), LC-MS grade formic acid (FA), bovine serum albumin (BSA), bovine hemoglobin and human hemoglobin S were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). LC-MS grade water and acetonitrile (ACN) were from Fisher Scientific Canada (Edmonton, AB, Canada). ACS grade 37% HCl was from Merck (KGaA, Darmstadt, Germany).

7.2.2 Normal Human Hemoglobin Protein Preparation

Fresh blood was collected from a healthy candidate and stored in a tube with anticoagulant reagent (EDTA). Red blood cells were separated from plasma and other blood components by centrifuging at 1958 × *g* for 15 min as commonly used in hospital and clinical research institutes. After centrifugation, plasma and red blood cells were separated into two layers and the plasma was transferred into a 1.5 mL micro-centrifuge tube. The red blood cell solution was lysed by adding LC/MS grade water and vortexing. Then the lysed solution was centrifuged at 20800 × *g* for 15 min. The supernatant containing the hemoglobins was transferred to a 1.5 mL micro-centrifuge tube for future research and the cell deris was discarded.

7.2.3 Microwave-Assisted Acid Hydrolysis of Proteins

For TFA MAAH, 10 μL (1 μg/μL) of the protein solution was mixed with an equal volume of 20 mM DTT in a 1.5 mL polypropylene centrifuge vial and incubated at 60 °C for 20 min. 20 μL 50% (vol/vol) TFA was added to the sample solution after incubation. The vial was then capped, sealed with Teflon tape, and placed in a domestic 1200 W (2450 MHz) microwave oven (Panasonic, London Drugs, Edmonton, Canada). The sample vial was placed on a Scienceware round bubble rack (Fisher Scientific, Edmonton, Canada) and floated in a plastic beaker, which contained 100 mL of water. The beaker was placed in the center of the rotating plate in the microwave oven. After microwave irradiation for 7.5 min or 10 min, the sample vial was taken from the microwave oven and the solution was dried in a SpeedVac vacuum centrifuge (Thermo Savant, Milford, MA) to remove

the acid.[2]

For the commercial microwave oven, the CEM Discover single-mode microwave reactor equipped with a fiber-optic probe microwave apparatus (CEM Corporation, Matthews, NC) was used. MAAH was carried out in the CEM microwave system with a water bath of boiled water for 7.5 min or 10 min with 200 or 300 W power applied, followed by drying down the sample in the SpeedVac. The dried hydrolysates were then reconstituted with 105 μL 0.1% TFA.

For HCl MAAH, 40 μL of 0.25 mg mL^{-1} protein solution was mixed with 0.5 μL of 500 mM DTT and 40 μL of 6 M HCl in a 1.5 mL polypropylene vial. The vial was capped, sealed with Teflon tape, and then placed inside the household microwave oven. A container with 100 ml of water was placed beside the sample vial to absorb the excess microwave energy. The sample was subjected to 60 s microwave irritation. For the improved HCl MAAH in the household microwave oven, the vial was placed on a Scienceware round bubble rack and floated in a plastic beaker containing 100 mL of boiled water. The beaker was placed in the center of the rotating plate in the microwave oven. The sample was microwave irradiated for a period of time from 30 s to 75 s in a water bath of boiled water with 240 W power applied.

For the commercial microwave oven method, MAAH was carried out in the CEM Discover microwave system with a water bath of boiled water for a period of time from 30 to 90 s with 200 or 300 W power applied. The acid lysed samples were cooled and dried in a SpeedVac to remove all of the acid. 30 μL of 250 mM NH_4HCO_3 was used to reconstitute the sample. 9 μL of 500 mM DTT was added

and reduction of the disulfide was performed by incubation for 60 min at 37 °C. After reduction, peptide solution was acidified to pH 2 with 50% TFA and diluted to 105 μL with 0.1% TFA.

7.2.4 LC Desalting or Fractionation of Hydrolysates

The desalting of the hydrolysates generated by TFA MAAH was carried out in an Agilent 1100 HPLC system (Palo Alto, CA, USA). In brief, desalting of the hydrolysates was performed on a 4.6 mm i.d. x 5 cm Polaris C18 A column with a particle size of 3 μm and 300 Å pore (Varian, MA, USA). After loading of 100 μL polypeptide sample, the column was flushed with 97.5% mobile phase A (0.1% TFA in water) and 2.5% mobile phase B (0.1% TFA in ACN) at a flow rate of 1mL/min at room temperature and the salts were effectively removed. Subsequently, the concentration of phase B in the mobile phase was step-wise increased to 85% to ensure complete elution of the polypeptides from the column, followed by 15 min re-equilibration with mobile phase A.[11]

The fractionation of the hydrolysates generated by HCl MAAH was carried out on a 3 mm i.d. × 15 cm Zorbax 300 - SB C_3 column with a particle size of 3.5 μm and 300 Å pore (Agilent, Mississauga, ON, Canada) in the Agilent 1100 HPLC system. After loading of 100 μL acidified polypeptide sample, the column was flushed with 97.5% mobile phase A (0.1% TFA in water) and 2.5% mobile phase B (0.1% TFA in ACN) at a flow rate of 0.4 mL/min for 5 min at room temperature and the salts were effectively removed. Subsequently, the concentration of phase B in the mobile phase was step-wise increased to 30% for

9 min to elute the low molecular weight polypeptides from the column, followed by 15 min re-equilibration with mobile phase A.

7.2.5 LC ESI MS/MS Analysis

The collected fractions were dried in the SpeedVac vacuum centrifuge to remove all the solvents and acid, reconstituted with 0.1 % formic acid and analyzed by a quadrupole time-of-flight (QTOF) Premier mass spectrometer (Waters, Manchester, U.K.) equipped with a nanoACQUITY Ultra Performance LC system (Waters, Milford, MA, USA). In brief, 5 µL of peptide solution was injected onto a 75 µm i.d. × 150 mm Atlantis dC_{18} column with 3 µm particle size (Waters). Solvent A consisted of 0.1% formic acid in water, and Solvent B consisted of 0.1% formic acid in ACN. The peptides generated by TFA MAAH were first separated using 120 min gradients (2%–6% Solvent B for 2 min, 6%–25% Solvent B for 95 min, 30%–50% Solvent B for 10 min, 50%–90% Solvent B for 10 min, 90%–5% Solvent B for 5 min at 35 ℃; the column was pre-equilibrated at 2% Solvent B for 20 min) and electrosprayed into the mass spectrometer (fitted with a nanoLockSpray source) at a flow rate of 300 nL/min. Mass spectra were acquired from m/z 300 to 1600 for 0.8 s, followed by four data-dependent MS/MS scans from m/z 50–1900 for 0.8 s each. The peptides generated by HCl MAAH were separated using a 30 min gradient (2%–7% Solvent B for 1 min, 7%–25% Solvent B for 19 min, 25%–50% Solvent B for 1 min, 50%–90% Solvent B for 2 min, 90%–90% Solvent B for 6 min, 90%–95% Solvent B for 2 min; the column was pre-equilibrated at 2% Solvent B for 20 min) and electrosprayed into the mass spectrometer (fitted with a nanoLockSpray source) at a flow rate of 350

nL/min. Mass spectra were acquired from m/z 400 to 1600 for 0.8 s, followed by six data-dependent MS/MS scans from m/z 50–1900 for 0.8 s each. The collision energy used to perform MS/MS was varied according to the mass and charge state of the eluting peptide. Leucine enkephalin and (Glu1)-fibrinopeptide B, a mixed mass calibrant (i.e., lock-mass), was infused at a rate of 300nL/min, and an MS scan was acquired for 1 s every 1 min throughout the run.[12]

7.2.6 Protein Database Search

Raw MS and MS/MS data were lock-mass-corrected, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.3 (Waters). Peptide sequences were identified via automated database searching of peak list files using the MASCOT search program (http://www.matrixscience.com). Database searches were restricted to the protein sequences of the corresponding protein as well as its variants downloaded from the SwissProt database and the HbVar database.[13,14] The following search parameters were selected for all database searching: enzyme, nonspecified; missed cleavages, 0; peptide tolerance, ±30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, (1+, 2+, and 3+); variable modifications, deamidation of asparagine and glutamine. The search results, including unique peptide sequences, ion score, MASCOT threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses were extracted to Excel files. All the identified peptides with scores lower than the MASCOT identity threshold scores for identity were then deleted from the list.

7.3 Results and Discussion

7.3.1 Method Development

For the household microwave oven, microwave irradiation is continuously applied to the samples at the specified energy level, but in the standard mode of CEM Discover microwave system which control the temperature and pressure at the same time, power would be adjusted or even off to maintain a certain temperature or pressure. The reproducibility of MAAH in the CEM Discovery microwave system operated in this mode was poor and much more time was needed to hydrolyze the samples to the same extent as the household microwave oven. To mimic the MAAH condition in the household microwave oven, the power-time mode of CEM was chosen. It applies a pre-set power to the samples all the time unless the temperature or pressure exceeds the set limit. The household microwave oven has a 1200 w maximum power, while CEM only has 300 W. To compensate the power difference, a water bath of boiled water was used for MAAH. Before samples were put in the container, 100 mL water was heated to 100 °C in the container by microwave irradiation. Then sample vials were placed into the container, and microwave irradiation was applied for a short period time with 200 W or 300 W powe. The temperature limit was set to 150 °C and the pressure limit was set to 147 psi. The fiber-optic probe was immersed in the water bath to monitor the environment temperature outside the sample vials. The pressure and temperature were impossible to exceed the limits, so the power would be applied all the time. The actual temperature and pressure inside the tubes in CEM were unknown; this was also true for the household microwave

Table 7.1 The number of identified peptides and sequence coverage of BSA hydrolysates generated by TFA MAAH.

	Identified peptides	Sequence coverage
[a]C, [c]7.5 min, [e]200 W	676 ± 25	97% ± 0%
[a]C, [d]10 min, [e]200 W	1363 ± 57	99% ± 1%
[a]C, [c]7.5 min, [f]300 W	1334 ± 74	99% ± 1%
[a]C, [d]10 min, [f]300 W	1593 ± 49	100% ± 0%
[b]H, [c]7.5 min	1622 ± 77	100% ± 0%
[b]H, [d]10 min	1629 ± 45	100% ± 0%

[a]Commercial microwave oven (CEM discover microwave system)
[b]Household microwave oven
[c]7.5 min: 7.5 min MAAH time
[d]10 min: 10 min MAAH time
[e]200 W: apply 200 W microwave power
[f]300 W: apply 300 W microwave power

oven.

7.3.1.1 TFA MAAH

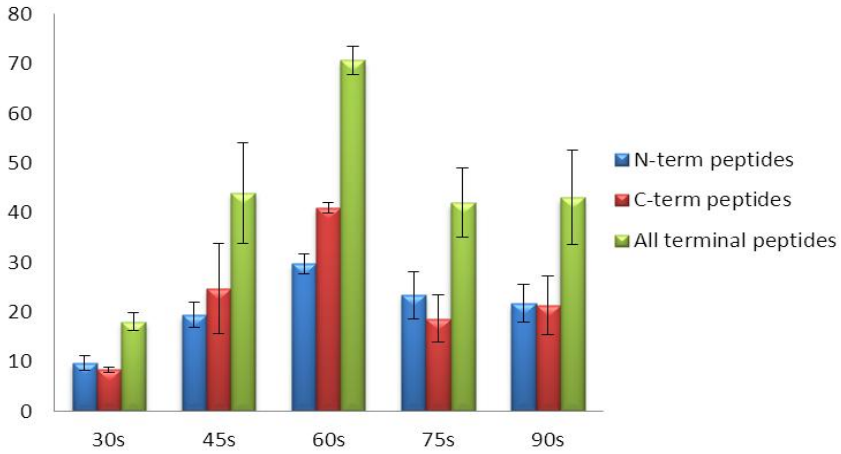
The BSA standard protein purchased from Sigma is a variant substituting A (alanine) to T (threonine) at position 214. If we searched against the database containing the original sequence, we always obtained a 99% sequence coverage with the position 214 missing, while 100% sequence coverage could be obtained

for the variant sequence. This result is in accordance with what Wang *et al.* found using the same method.[2] For the TFA MAAH in the CEM Discover microwave system, similar results were generated. Nearly 100% sequence coverage was achieved for the different conditions tested. Table 7.1 lists the number of identified peptides and sequence coverage of BSA hydrolysates generated by TFA MAAH. Around 1500 peptides were identified with 100% sequence coverage of BSA for microwave irradiated for 7.5 min or 10 min in the household microwave oven, while for the CEM Discover microwave system, comparable results were obtained for 10-min microwave irradiation at the maximum power, i.e., 300 W.

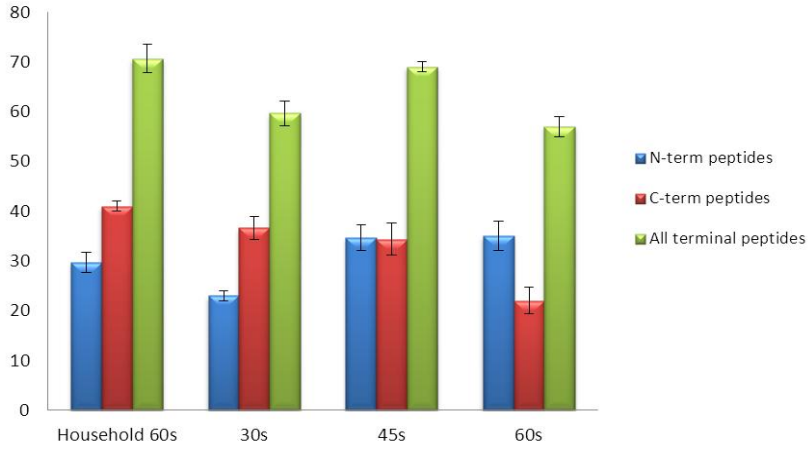
7.3.1.2 HCl MMAH

If HCl MAAH is used for terminal peptide analysis, the microwave irradiation process should be controlled carefully to avoid the further hydrolysis of terminal peptides into internal peptides. We tried to do HCl MAAH of BSA protein for 30, 45, 60, 75 and 90 s in the household microwave oven first. Figure 7.1 shows the number of terminal peptides of BSA identified in the hydrolysates generated through HCl MAAH, analyzed by LC-ESI MS/MS (n=3). For BSA HCl MAAH in the household microwave oven, 60 s seems the best.

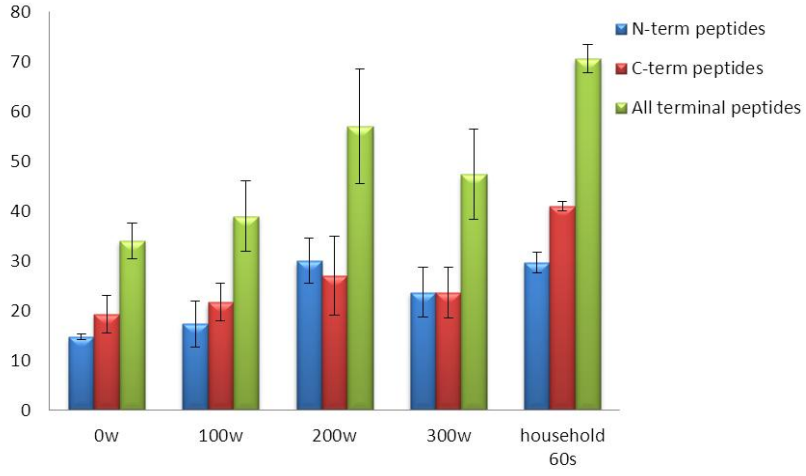
However, household microwave ovens are known to have “hot” spots due to unevenly distributed microwave radiation.[2] If we tried to use another domestic 1200 W (2450 MHz) microwave oven to perform the HCl MAAH, almost no hydrolysis occurred when we put the sample vials at the same position in the microwave oven. Then we tried to place the sample vials at other positions during the HCl MAAH. The proteins hydrolyzed to different extents in the different


80
70
60
50
40
30
20
10
0
30s
45s
60s
75s
90s
N-term peptides
C-term peptides
All terminal peptides

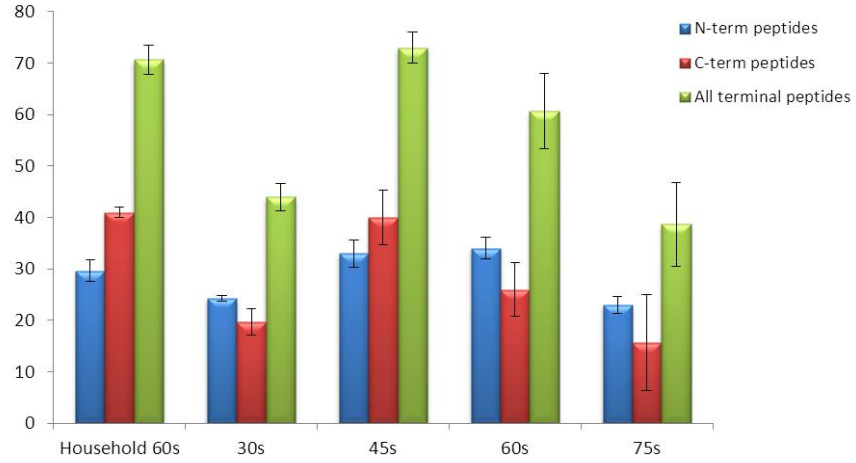
(a)


80
70
60
50
40
30
20
10
0
Household 60s
30s
45s
60s
N-term peptides
C-term peptides
All terminal peptides

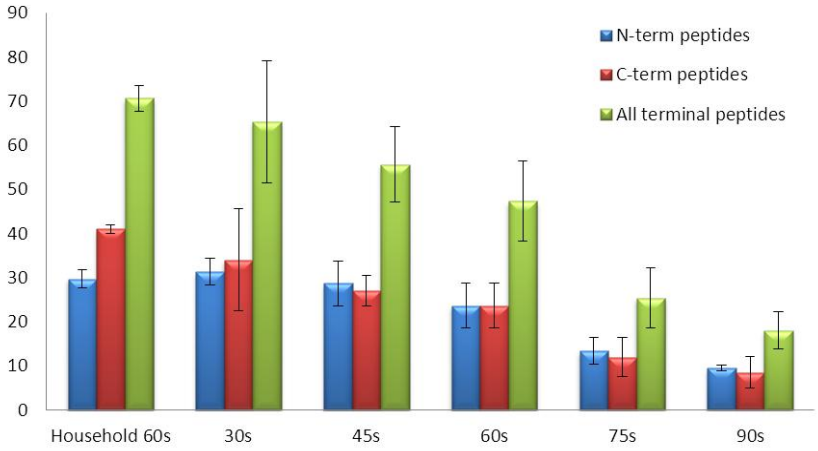
(b)


80
70
60
50
40
30
20
10
0
0w
100w
200w
300w
household 60s
N-term peptides
C-term peptides
All terminal peptides

(c)



(d)



(e)

Figure 7.1 The number of terminal peptides of BSA identified in the hydrolysates generated through HCl MAAH, analyzed by LC-ESI MS/MS (n=3). (a) BSA HCl MAAH for 30, 45, 60, 75 and 90 s in the household microwave oven. (b) BSA HCl MAAH for 30, 45 and 60 s in a water bath of boiled water in the household microwave oven. (c) BSA HCl MAAH for 60 s at 0, 100, 200, and 300 W in the CEM. (d) BSA HCl MAAH for 30, 45, 60, and 75 s at 200 W power in the CEM. (e) BSA HCl MAAH for 30, 45, 60, 75, and 90 s at 300 W power in the CEM.

positions and it would take a rather long time to find a right " hot" spot to get a similar result as the previous domestic microwave oven did.

Wang *et al.* demonstrated reproducible TFA MAAH in a household microwave oven using a water bath. The sample vials were placed in a rack floated in a plastic beaker with 100 mL water inside. They got similar results no matter where the vials were placed in the beaker and no matter where the beaker was placed in the household microwave oven.[2] Thus we considered operating the HCl MAAH in the same apparatus used in their experiment to generate reproducible hydrolysis results. Because no stir bar could be used in the water bath, we used a water bath of boiled water, instead of cold water, in the start to avoid uneven heating in the water bath. The 100 mL water in the beaker could be microwave irradiated to boiling within 2 min in the domestic 1200 W microwave oven. Then we tested different hydrolysis times for HCl MAAH of the proteins at 1200 W power. All the proteins hydrolyzed to peptides, and internal peptides dominated even after 20 s microwave irradiation. To slow down the hydrolysis process, lower power (240 W) was used to test different hydrolysis times from 30 s to 60 s. Figure 7.1b shows that HCl MAAH in the water bath of boiled water for 45 s at 240 W power could generate comparable results with the previous "dry" method in a different domestic microwave oven. Furthermore, the results were also reproducible at different positions in the beaker and different positions in the household microwave oven.

For HCl MAAH in the commercial microwave oven, a water bath of boiled water was also employed to avoid the uneven heating before the water temperature

reached 100 °C. First we tested BSA HCl MAAH for 60 s at 0, 100, 200, and 300 W in the CEM Discover microwave system. Figure 7.1c shows that the number of identified peptides gradually increased as the power increased, reached the maximum at 200 W power, then decreased at 300 W. Only heating (i.e., no microwave irradiation was applied) could hydrolyze the proteins, but with the assistance of microwave irradiation to agitate the molecules, the proteins would generate more terminal peptides in a period of time as short as 60 s. Then we tested HCl MAAH of BSA for 30 to 75 s at 200 W power in the CEM. Figure 7.1d demonstrates that HCl MAAH for 45 s at 200 W in CEM generated comparable results to the household microwave oven. At last, BSA HCl MAAH at 300 W power in the CEM was also tested. Figure 7.1e shows that HCl MAAH for 30 s at 300 w in CEM generated similar results as the household microwave oven. It is not surprising that higher power microwave irradiation requires less time in MAAH.

The above results indicate that similar results could be obtained in the CEM as in the household microwave oven for both TFA and HCl MAAH for the standard

protein, BSA. We then applied the method of TFA and HCl MAAH performed in a commercial microwave oven to clinically relevant samples.

7.3.2 Sequencing of Hemoglobin Variants by TFA MAAH

Human hemoglobin is a tetrameric protein composed of two different globin subunits, each of which non-covalently binds a heme molecule.[15] The subunits are α- and β-globins consisting of 141 and 146 with amino acid residues, respectively,

for the major hemoglobin Hb A (or Hb A_0).[15,16] Over 1000 variants of human hemoglobins have been discovered to date[17,18] and this number keeps increasing. Most human hemoglobin variants have a point mutation substituting a single amino acid for another.[19] Hb S, sickle hemoglobin, is one of them. It substitutes the glutamic acid (G) to valine (V) at position 6 of the β-chain.[20,21] This mutation causes polymerization of Hb S, leading to a decreased deformability of the red blood cells, the so-called sickle cells, which cannot pass through small blood vessels, leading to a lack of oxygen supply of the surrounding tissue.[21-23]

Since the first analysis of a tryptic peptide mixture of globin chains from human hemoglobin (Hb) using field desorption mass spectrometry reported by Matsuo *et al.*, mass spectrometry (MS) has been rapidly applied to the characterization of Hb proteins.[24] However, it was not until the introduction of electrospray ionization mass spectrometry (ESI–MS) in 1988 that it became practicable to determine the molecular weights of intact globin chains with sufficient accuracy to allow MS to be used as part of the procedure to identify abnormal Hbs.[25] Recently, ESI-MS and matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS have become the two main analysis tools for the characterization of hemoglobin variants following detection using either HPLC or electrophoresis.[24,26,27] The presence of a variant and the nature of the mutation may be indicated by the accurate mass measurement of globin polypeptides. But the mutation site cannot be predicted and the specificity is lacking. For example, the sickle hemoglobin (β6 Glu>Val) is 30 Da lighter the normal protein and such a mass shift is also observed for other amino acid substitutions (Trp>Arg,

Thr>Ala or Glu>Val at other positions), hence a conventional bottom up proteomics approach is the solution for mass spectrometry based variant characterization.[17,28] TFA MAAH combined with LC-ESI MS/MS could cover the entire protein sequence. Thus it could become an ideal technique in the analysis of proteins with point-mutations such as hemoglobin variants.

(a)

Sequence Coverage: **99%**

Matched peptides shown in **Bold Red**

```
     1 MVLSAADKGN VKAAWGKVGG HAAEYGAEAL ERMFLSFPTT KTYFPHFDLS
    51 HGSAQVKGHG AKVAAALTKA VEHLDDLPGA LSELSDLHAH KLRVDPVNFK
   101 LLSHSLLVTL ASHLPSDFTP AVHASLDKFL ANVSTVLTSK YR
```

(b)

Sequence Coverage: **100%**

Matched peptides shown in **Bold Red**

```
     1 MLTAEEKAAV TAFWGKVKVD EVGGEALGRL LVVYPWTQRF FESFGDLSTA
    51 DAVMNNPKVK AHGKKVLDSF SNGMKHLDDL KGTFAALSEL HCDKLHVDPE
   101 NFKLLGNVLV VVLARNFGKE FTPVLQADFQ KVVAGVANAL AHRYH
```

(c)

Sequence Coverage: **99%**

Matched peptides shown in **Bold Red**

```
     1 MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS
    51 HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK
   101 LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR
```

(d)

Sequence Coverage: **99%**

Matched peptides shown in **Bold Red**

```
  1 MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS
 51 TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD
101 PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
```

(e)

Sequence Coverage: **99%**

Matched peptides shown in **Bold Red**

```
  1 MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS
 51 HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK
101 LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR
```

(f)

Sequence Coverage: **99%**

Matched peptides shown in **Bold Red**

```
  1 MVHLTPVEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS
 51 TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD
101 PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
```

Figure 7.2 The sequence coverage of (a) alpha subunit of bovine hemoglobin, (b) beta subunit of bovine hemoglobin, (c) alpha subunit of normal human hemoglobin, (d) beta subunit of normal human hemoglobin, (e) alpha subunit of sickle hemoglobin, and (f) beta subunit of sickle hemoglobin, digested by TFA MAAH.

In our work, we digested the normal human hemoglobin, sickle hemoglobin and bovine hemoglobin by TFA MAAH, analyzed the digests using LC-ESI MS/MS and searched the data against the database containing the protein sequences of all the human and bovine hemoglobin variants. The alpha and beta subunits of bovine hemoglobin were identified with 99% and 92% sequence coverage, respectively. If the database only contained all the bovine hemoglobin variants, 99% and 100% sequence coverage could be obtained, as Figure 7.2a and 7.2b show. Generally, the Mascot identity threshold decreased with the database size. Thus a smaller database results in the identification of more peptides and larger sequence coverage. The experimental results demonstrate the first methionine of the alpha subunit is removed after translation while the beta unit is not, matched with the theoretic protein sequence. Including the corrected matched alpha and beta subunits of bovine hemoglobin, Mascot search results list 61 variants in total, but only the two corrected matched variants had the highest sequence coverage, as well as a variant which had a lysine (K) mutated to glutamine (Q) at the position 131 of the beta subunit were also present in the search results. The mass difference between K and Q is 0.5 Da, so we could distinguish the two variants by mass spectrometry.

For human hemoglobins, because too many variants are present, the correct alpha and beta units cannot be identified by searching against the whole hemoglobin database. To address this issue, we performed the LC-ESIMS analysis of intact hemoglobins with an Agilent Technologies 6220 oaTOF (Agilent, Mississauga, ON, Canada) to determine the molecular weights of the hemoglobins. Then we picked out the variants with molecular weights within ± 2 Da of the experimental

molecular weight from the database to construct a specific database for each variant of hemoglobin. For the normal human hemoglobin, 62 variants were present in the search results when we searched against its specific database containing 89 variants. For the sickle hemoglobin, 36 variants were present in the search results when we searched against its specific database containing 55 variants. These variants had point mutations of Asp>Asn, Asn>Asp, Glu>Gln, Gln>Glu, Lys>Gln, Gln>Lys, Lys>Glu, Glu>Lysc, or Asn>Ile and the mutated beta subunits of sickle cell had mutations of Trp>Arg, Thr>Ala, Arg>Lys, Val>Ala, Gln>pro or Glu>Val at other positions. TFA MAAH cuts every peptide bond of proteins nonspecifically, so the tandem mass spectra are much more complicated than the ones generated from the specific enzyme digestions. It is difficult to distinguish these mutations whose molecular weight differences are within 1 Da, especially when deamidation at Gln and Asn is rather common in microwave-assisted acid hydrolysis.

Kleinert *et al.* also pointed out the limitations of mass spectrometry in hemoglobin variants analysis and stated that "Two important drawbacks of the MS methods should be mentioned. First, its insufficient resolution prevents the detection of Hb mutations with small mass differences of the globin chains. The precision of normal low-resolution mass measurements was insufficient to distinguish the wild-type β-chain from several β-chain variants such as HbC, D, or E. Second, MS as described here is only a qualitative technique, and in particular, minor Hb fractions such as HbA1c or HbA2, which are important for diagnosis of diabetes mellitus or thalassemias, respectively, cannot be quantified."[26] Brennan also commented similarly by stating that, whereas

traditional methods readily detect the majority of common variants, such as HbC, HbD, or HbE, "the substitutions involved in these, and similar charge variants (Glu> Lys, Glu> Gln, Asp> Asn, Lys> Gln, and Leu> Ile) involve mass changes of 1 Da or less, and are not detectable by mass spectrometry."[29] The specific trypsin digestion[18,21,30-32] or IEF[33] could be combined with mass spectrometry to further distinguish these variants with similar molecular weights.

For the corrected identified subunits of human hemoglobins, Figure 7.2c, 7.2d, 7.2d and 7.2f demonstrates that 99% sequence coverage was obtained for all of them, indicating all have the first methionine removed.

The technique we used here clearly has its limitations, as described above. However, for point mutations that have larger molecular weight differences, TFA MAAH combined with molecular weight analysis should be able to determine the site of the mutations by checking the sequence coverage. We will need to examine this issue by using suitable hemoglobin variants that we cannot access to at this moment. Future work in collaboration with clinical researchers having various hemoglobin variants should benefit the development of this TFA HCl MAAH method for variant characterization.

7.3.3 Terminal Peptides Analysis of Hemoglobin Variants by HCl MAAH

The digests of the normal human hemoglobin, sickle hemoglobin and bovine hemoglobin by HCl MAAH were analyzed by LC-ESI MS/MS and searched against the whole hemoglobin database containing the protein sequences of all the human and bovine hemoglobin variants. Several terminal peptides of the alpha

Table 7.2 List of identified unique terminal peptides of alpha subunit of bovine hemoglobin.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
2 - 15	VLSAADKGNVKAA W	715.3989	0.0108	7.56	63.46
2 - 16	VLSAADKGNVKAA WG	743.9005	-0.0075	-5.0	60.94
2 - 17	VLSAADKGNVKAA WGK	538.9725	0.0068	4.2	69.39
2 - 18	VLSAADKGNVKAA WGKV	571.9933	0.0007	0.4	66.89
2 - 19	VLSAADKGNVKAA WGKVG	443.4934	-0.0343	-19.4	27.41
2 - 20	VLSAADKGNVKAA WGKVGG	457.7482	-0.0366	-20.0	54.94
2 - 21	VLSAADKGNVKAA WGKVGGH	492.0184	-0.0147	-7.48	40.19
2 - 22	VLSAADKGNVKAA WGKVGGHA	509.7728	-0.0342	-16.8	105.67
2 - 23	VLSAADKGNVKAA WGKVGGHAA	703.0545	0.0083	3.9	77.11
2 - 25	VLSAADKGNVKAA WGKVGGHAAEY	600.5677	0.0024	1.0	26.16
2 - 26	VLSAADKGNVKAA WGKVGGHAAEYG	819.4282	0.002	0.8	67.56
2 - 27	VLSAADKGNVKAA WGKVGGHAAEYGA	632.5798	-0.0078	-3.1	27.5
2 - 28	VLSAADKGNVKAA WGKVGGHAAEYGA E	886.1238	0.0091	3.4	36.27
2 - 29	VLSAADKGNVKAA WGKVGGHAAEYGA EA	909.8045	0.0141	5.17	51.46

2 - 30	VLSAADKGNVKAAWGKVGGHAAEYGAEAL	947.496	0.0045	1.6	25.91
2 - 31	VLSAADKGNVKAAWGKVGGHAAEYGAEALE	990.5109	0.0066	2.2	17.23
2 - 32	VLSAADKGNVKAAWGKVGGHAAEYGAEALER	782.1603	0.0067	2.1	88.68
2 - 34	VLSAADKGNVKAAWGKVGGHAAEYGAEALERMF	851.6856	-0.001	-0.3	107.09
2 - 16	VLSAADKGNVKAAWG	743.9005	-0.0075	-5.0	60.94
2 - 17	VLSAADKGNVKAAWGK	538.9725	0.0068	4.2	69.39
2 - 18	VLSAADKGNVKAAWGKV	571.9933	0.0007	0.4	66.89
2 - 19	VLSAADKGNVKAAWGKVG	443.4934	-0.0343	-19.4	27.41
137 - 142	LTSKYR	767.4455	0.0045	5.9	13.3
135 - 142	TVLTSKYR	484.284	0.0036	3.7	25.41
134 - 142	STVLTSKYR	527.7994	0.0024	2.3	52.5
133 - 142	VSTVLTSKYR	385.2183	-0.0172	-44.7	52.36
132 - 142	NVSTVLTSKYR	423.2288	-0.0286	-22.6	47.71
131 - 142	ANVSTVLTSKYR	669.8755	0.0061	4.6	86.53
130 - 142	LANVSTVLTSKYR	484.6068	-0.0158	-10.9	80.41
129 - 142	FLANVSTVLTSKYR	533.6264	-0.0254	-15.9	91.29
128 - 142	KFLANVSTVLTSKYR	576.3308	-0.0072	-4.2	41.54
127 - 142	DKFLANVSTVLTSKYR	614.6764	0.0027	1.5	34.66

126 - 142	LDKFLANVSTVLTS KYR	489.527	-0.0098	-5.0	45.25
125 - 142	SLDKFLANVSTVLTS KYR	681.3812	0.001	0.5	42.87
122 - 142	VHASLDKFLANVST VLTSKYR	588.0751	-0.0139	-5.92	13.04

Table 7.3 List of identified unique terminal peptides of beta subunit of bovine hemoglobin.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
1 - 14	MLTAEEKAAVTAF W	784.4006	0.0115	7.34	27.9
1 - 15	MLTAEEKAAVTAF WG	812.9114	0.0116	7.14	62.6
1 - 20	MLTAEEKAAVTAF WGKVKVD	732.0607	0.0099	4.5	70.37
1 - 21	MLTAEEKAAVTAF WGKVKVDE	775.0741	0.0075	3.2	35.91
1 - 23	MLTAEEKAAVTAF WGKVKVDEVG	620.5761	-0.0075	-3.0	18.51
1 - 24	MLTAEEKAAVTAF WGKVKVDEVGG	846.1174	0.0261	10.3	16.12
137 - 145	ANALAHRYH	526.7777	0.0097	9.2	24.47
134 - 145	AGVANALAHRYH	640.3382	0.0037	2.9	25.34
133 - 145	VAGVANALAHRYH	689.874	0.0069	5.0	27.28
132 - 145	VVAGVANALAHRY H	739.4096	0.0097	6.6	79.59
129 - 145	FQKVVAGVANALA HRYH	471.0035	-0.032	-17	96.07

Table 7.4. List of identified unique terminal peptides of alpha subunit of normal human hemoglobin.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
2 - 13	VLSPADKTNVK A	622.3519	0.0073	5.9	57.1
2 - 14	VLSPADKTNVK AA	657.8722	0.0108	8.22	42.14
2 - 15	VLSPADKTNVK AAW	750.9115	0.0101	6.73	48.98
2 - 16	VLSPADKTNVK AAWG	778.9294	0.0084	5.4	44.19
2 - 23	VLSPADKTNVK AAWGKVGAHA G	726.4034	0.0131	6.02	74.97
2 - 26	VLSPADKTNVK AAWGKVGAHA GEYG	632.3359	0.0118	4.67	36.16
137 - 142	LTSKYR	767.4458	0.0048	6.3	20.75
136 - 142	VLTSKYR	866.5117	0.0023	2.7	17.98
134 - 142	STVLTSKYR	527.7935	-0.0094	-8.9	63.37
132 - 142	SVSTVLTSKYR	620.8499	0.003	2	74.14
131 - 142	ASVSTVLTSKYR	437.9125	-0.0037	-2.8	75.03
126 - 142	LDKFLASVSTVL TSKYR	643.3698	0.0097	5.0	107.55
125 - 142	SLDKFLASVSTV LTSKYR	1008.0728	0.0212	10.5	39.81
119 - 142	TPAVHASLDKFL ASVSTVLTSKYR	648.6121	0.0074	2.9	32.64

Table 7.5 List of identified unique terminal peptides of beta subunit of normal human hemoglobin.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
2 - 12	VHLTPEEKSAV	605.3302	0.0058	4.8	49.72
2 - 14	VHLTPEEKSAVTA	691.3753	0.0112	8.11	39.8
2 - 15	VHLTPEEKSAVTAL	747.9191	0.0147	9.84	35.47
2 - 17	VHLTPEEKSAVTALWG	869.4655	0.0067	3.9	54.54
143 - 147	AHKYH	655.3354	0.0043	6.6	24.37
137 - 147	GVANALAHKYH	591.3073	0.0012	1.0	100.67
136 - 147	AGVANALAHKYH	626.3358	0.0051	4.1	96.09
134 - 147	VVAGVANALAHKYH	725.4048	0.0062	4.3	80.25
133 - 147	KVVAGVANALAHKYH	526.9627	-0.0015	-1.0	54.18
131 - 147	YQKVVAGVANALAHKYH	624.3347	0.0086	4.6	86.05
130 - 147	AYQKVVAGVANALAHKYH	648.0114	0.0016	0.8	100.37
129 - 147	AAYQKVVAGVANALAHKYH	671.3638	0.0057	2.8	40.74

Table 7.6 List of identified unique terminal peptides of alpha subunit of sickle hemoglobin.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
2 - 13	VLSPADKTNVKA	622.3524	0.0083	6.7	56.11
2 - 14	VLSPADKTNVKAA	438.9098	-0.0115	-8.75	43.69
2 - 15	VLSPADKTNVKAA W	750.911	0.0091	6.1	64.9
2 - 16	VLSPADKTNVKAA WG	778.9288	0.0072	4.6	31.33
136 - 142	VLTSKYR	866.5153	0.0059	6.8	23.87
135 - 142	TVLTSKYR	484.2835	0.0026	2.7	40.59
134 - 142	STVLTSKYR	527.7863	-0.0238	-22.6	65.53
133 - 142	VSTVLTSKYR	577.3356	0.0064	5.6	64.74
132 - 142	SVSTVLTSKYR	620.8524	0.008	6	80.52
131 - 142	ASVSTVLTSKYR	437.9104	-0.01	-8	80.42
130 - 142	LASVSTVLTSKYR	475.6047	-0.0112	-7.87	64.15
129 - 142	FLASVSTVLTSKYR	524.634	0.0083	5.3	22.98
128 - 142	KFLASVSTVLTSKYR	567.3313	0.0052	3.1	111.01
127 - 142	DKFLASVSTVLTSKY R	605.6718	-0.0002	-0.1	101.83
125 - 142	SLDKFLASVSTVLTS KYR	672.3805	0.0098	4.9	27.38
119 - 142	TPAVHASLDKFLAS VSTVLTSKYR	648.6112	0.0038	1.5	37.87

Table 7.7 List of identified unique terminal peptides of beta subunit of sickle hemoglobin.

position	peptide sequence	m/z (Da)	Δm	Error (ppm)	peptide score
2 - 8	VHLTPVE	794.4468	0.0061	7.7	45.64
2 - 13	VHLTPVEKSAVT	640.8672	0.0062	4.8	35.73
2 - 14	VHLTPVEKSAVTA	676.3885	0.0117	8.66	53.48
2 - 15	VHLTPVEKSAVTAL	732.9312	0.0131	8.95	36.57
2 - 18	VHLTPVEKSAVTALWGK	612.6785	-0.0168	-9.16	25.94
143 - 147	AHKYH	655.3381	0.007	11	23.93
139 - 147	ANALAHKYH	513.2514	-0.0207	-20.2	76.44
137 - 147	GVANALAHKYH	591.3131	0.0128	10.8	96.92
136 - 147	AGVANALAHKYH	626.3348	0.0031	2.5	96.34
134 - 147	VVAGVANALAHKYH	725.4064	0.0094	6.5	91.12
133 - 147	KVVAGVANALAHKYH	526.6348	-0.0012	-0.8	113.42
132 - 147	QKVVAGVANALAHKYH	569.6484	-0.003	-2	88.61
131 - 147	YQKVVAGVANALAHKYH	624.0083	0.0134	7.17	111.89
130 - 147	AYQKVVAGVANALAHKYH	647.3572	0.007	4	68.53
129 - 147	AAYQKVVAGVANALAHKYH	671.0385	0.0138	6.87	48.55

and beta subunits of bovine hemoglobin were identified and no other variants were identified except the variant (β131 Lys>Glu), as it was the case in TFA MAAH. More terminal peptides were found if we searched against the database containing the bovine hemoglobin variants only. Tables 7.2 and 7.3 list the identified unique terminal peptides of the alpha or beta subunit of bovine hemoglobin. They clearly show the first methionine of the beta subunit is removed while the alpha subunit is not.

For human hemoglobins, the correct alpha and beta units could not be identified if we searched against the whole hemoglobin database. Again the specific hemoglobin databases only containing the protein sequences matched with the experimental molecular weight were used for the search. 60 variants were found in the Mascot search results for normal human hemoglobin, while 21 variants were present for sickle hemoglobin, slightly less than the TFA MAAH search results. Most peptides generated by HCl MAAH are terminal peptides, and the spectra are relatively less complicated than the ones generated by TFA MAAH, and thus the false identified hemoglobin variants are reduced. Tables 7.4, 7.5, 7.6 and 7.7 list the identified unique terminal peptides of the alpha and beta subunit of human hemoglobins, indicating all had the first methionine removed.

7.4 Conclusions

We have improved the in-solution HCl MAAH method for terminal peptides analysis using the household microwave oven with a water bath of boiled water, overcoming the hot spot problem associated with microwave ovens. The results were more reproducible than the “dry” method. The two MAAH methods using

HCl or TFA developed in the household microwave oven were successfully implemented into a commercial microwave device. Similar numbers of terminal peptides were identified by LC-ESI MS/MS analysis. We found the results were comparable for TFA MAAH of BSA at 300 W for 10 min in the commercial microwave oven in a water bath of boiled water with 1200 W for about 10 min in the household microwave oven. For HCl MAAH, microwave irradiation at 200 W for 45 s in the commercial microwave oven was similar that 240 W for 45 s in the household microwave oven with a water bath of boiled water. High temperature and microwave irradiation are critic for fast acid hydrolysis. The microwave power was lowered in HCl MAAH to control the hydrolysis speed so that the most peptides generated were terminal peptides. We applied the two MAAH methods to analyze the variants of human and bovine hemoglobins, but MAAH methods were difficult to distinguish the variants with mutations of 1 Da mass differences, especially due to the more frequent deamidation in MAAH. For mutations with larger mass differences, MAAH should be very useful in determining the nature and the position of the mutation.

7.5 Literature Cited

(1) Zhong, H. Y.; Zhang, Y.; Wen, Z. H.; Li, L. *Nat Biotechnol* **2004**, *22*, 1291.
(2) Wang, N.; Li, L. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1573.
(3) Hua, L.; Low, T. Y.; Sze, S. K. *Proteomics* **2006**, *6*, 586.
(4) Hauser, N. J.; Han, H. L.; McLuckey, S. A.; Basile, F. *J Proteome Res* **2008**, *7*, 1867.
(5) Li, A. Q.; Sowder, R. C.; Henderson, L. E.; Moore, S. P.; Garfinkel, D. J.; Fisher, R. J. *Anal Chem* **2001**, *73*, 5395.

(6) Swatkoski, S.; Russell, S. C.; Edwards, N.; Fenselau, C. *Anal Chem* **2006**, *78*, 181.
(7) Swatkoski, S.; Gutierrez, P.; Wynne, C.; Petrov, A.; Dinman, J. D.; Edwards, N.; Fenselau, C. *J Proteome Res* **2008**, *7*, 579.
(8) Sun, W.; Gao, S.; Wang, L.; Chen, Y.; Wu, S.; Wang, X.; Zheng, D.; Gao, Y. *Mol Cell Proteomics* **2006**, *5*, 769.
(9) Vesper, H. W.; Mi, L.; Enada, A.; Myers, G. L. *Rapid communications in mass spectrometry : RCM* **2005**, *19*, 2865.
(10) Ha, N. Y.; Kim, S. H.; Lee, T. G.; Han, S. Y. *Langmuir* **2011**, *27*, 10098.
(11) Reiz, B.; Li, L. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1596.
(12) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Anal Chem* **2009**, *81*, 1049.
(13) Giardine, B.; van Baal, S.; Kaimakis, P.; Riemer, C.; Miller, W.; Samara, M.; Kollia, P.; Anagnou, N. P.; Chui, D. H. K.; Wajcman, H.; Hardison, R. C.; Patrinos, G. P. *Hum Mutat* **2007**, *28*.
(14) Hardison, R. C.; Chui, D. H. K.; Giardine, B.; Riemer, C.; Patrinos, G. P.; Anagnou, N.; Miller, W.; Wajcman, H. *Hum Mutat* **2002**, *19*, 225.
(15) Wada, Y. *J Chromatogr B* **2002**, *781*, 291.
(16) Zurbriggen, K.; Schmugge, M.; Schmid, M.; Durka, S.; Kleinert, P.; Kuster, T.; Heizmann, C. W.; Troxler, H. *Clin Chem* **2005**, *51*, 989.
(17) Mathew, B.; Bhat, V.; Mandal, A. K. *Anal Biochem* **2011**, *416*, 135.
(18) Griffith, W. P.; Kaltashov, I. A. *Int J Mass Spectrom* **2008**, *278*, 114.
(19) Williams, J. P.; Giles, K.; Green, B. N.; Scrivens, J. H.; Bateman, R. H. *Rapid Communications in Mass Spectrometry* **2008**, *22*, 3179.
(20) Falick, A. M.; Shackleton, C. H. L.; Green, B. N.; Witkowska, H. E. *Rapid Commun Mass Sp* **1990**, *4*, 396.
(21) Huttenhain, R.; Hess, S. *Proteomics* **2010**, *10*, 3657.
(22) Rees, D. C.; Williams, T. N.; Gladwin, M. T. *Lancet* **2010**, *376*, 2018.
(23) Strasser, B. J. *Science* **1999**, *286*, 1488.
(24) Guo, N.; Higgins, T. N. *Clin Biochem* **2009**, *42*, 99.
(25) Wild, B. J.; Green, B. N.; Cooper, E. K.; Lalloz, M. R. A.; Erten, S.; Stephens, A. D.; Layton, D. M. *Blood Cell Mol Dis* **2001**, *27*, 691.
(26) Kleinert, P.; Schmid, M.; Zurbriggen, K.; Speer, O.; Schmugge, M.; Roschitzki, B.; Durka, S. S.; Leopold, U.; Kuster, T.; Heizmann, C. W.; Frischknecht, H.; Troxler, H. *Clin Chem* **2008**, *54*, 69.
(27) Zanella-Cleon, I.; Joly, P.; Becchi, M.; Francina, A. *Clin Biochem* **2009**, *42*, 1807.
(28) Boemer, F.; Ketelslegers, O.; Minon, J. M.; Bours, V.; Schoos, R. *Clin Chem* **2008**, *54*, 2036.
(29) Brennan, S. O. *Clin Chem* **2008**, *54*, 8.
(30) Guo, J. S.; Uppal, S.; Easthon, L. M.; Mueser, T. C.; Griffith, W. P. *Int J Mass Spectrom* **2012**, *312*, 70.
(31) Henning, S.; Mormann, M.; Peter-Katalinic, J.; Pohlentz, G. *Amino Acids* **2011**, *41*, 343.
(32) Houston, C. T.; Reilly, J. P. *Anal Chem* **1999**, *71*, 3397.
(33) Wajcman, H.; Prehu, C.; Bardakdjian-Michau, J.; Prome, D.; Riou, J.; Godart, C.; Mathis, M.; Hurtrel, D.; Galacteros, F. *Hemoglobin* **2001**, *25*, 169.

Chapter 8

In-Gel Microwave-Assisted Acid Hyrolysis (MAAH) of Proteins Separated by Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

8.1 Introduction

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is a widely used technique for the separation and characterization of proteins.[1] In particular, two-dimensional (2-D) SDS-PAGE, with higher separation ability than other techniques, has been applied for the comprehensive analysis of protein composition in a given biological sample.[2] Protein separation by SDS-PAGE combined with sensitive mass spectrometry analysis is also widely used in proteome analysis.

To analyze the protein using MS, electroelution and passive extraction are the traditional ways to recover gel-separated proteins.[3] Electroelution[4] has high recovery rates for proteins from gels, but the elution must be purified to remove the salts and SDS accompanied with the proteins prior to MS analysis. Passive extraction[5-10] involves direct protein extraction with a solvent such as an acidic organic solvent. The extractions are salts and SDS free, but the extraction efficiencies are rather low.

In addition to directly extract proteins out of gel, in-gel digestion of SDS-PAGE separated proteins is efficient to generate peptides for MS analysis.[11] Proteins can be digested in-gel enzymatically or chemically. In-gel trypsin digestion[4,12-25] is the most commonly used technique. While chemical CNBr is extensively applied for methionine cleavage throughout the years[26], in-gel CNBr[20,27-30] digestion is also widely used for proteome analysis. In addition to trypsin and CNBr, dilute formic acid was reported to cleave proteins at the C-terminal or sometimes N-terminal of the aspartyl (Asp) residues of a protein.[31-33] In-gel digestion using formic acid is also reported.[33] Xiang *et al.*[34] reported a method, in-gel partial protein hydrolysis by hydrochloric acid, for N- and C-termini identification of proteins isolated by gel. Proteins were digested in-gel by 3 M HCl overnight at room temperature. Then terminal peptide ladders were identified by MALDI-MS analysis.

Protein digestion can be sped up with the help of microwave energy. More and more researchers now are trying to incorporate microwave irradiation into different reactions so that much less time is needed for the sample preparation steps. Sun *et al.*[35] developed in-gel microwave-assisted protein enzymatic digestion. They could digest protein mixtures in gel in 25 min, and its peptide yield efficiency was better than the present standard method (16 h or overnight). Juan *et al.* [36] also shortened the in-gel protein digestion time from 16 h to 5 min in microwave-assisted trypsin digestion for successful protein identifications. Hua *et al.*[31] developed a microwave-assisted formic acid digestion technique. Dilute formic acid cleaves proteins either in-solution or in-gel specifically at the C-terminal of aspartyl (Asp) residues within 10 min of exposure to microwave irradiation.

Microwave-assisted acid hydrolysis (MAAH) was first introduced by Zhong *et al.*[37] Proteins were hydrolyzed to terminal peptides in less than 2 min by 3M HCl with microwave irradiation. In this work, we applied this MAAH method to proteins separated by SDS-PAGE for terminal peptide analysis. In-gel HCl MAAH was performed for the standard protein BSA. The hydrolyzed terminal peptides were identified by LC-ESI MS/MS analysis. We tested different gel types and staining methods, optimized the MAAH time, and analyzed different amounts of BSA in-gel. While the number of identified terminal peptides decreased with the reduced loading amounts of BSA in-gel, the terminal peptides of BSA could still be identified with a sample loading of as low as 0.5 μg (7.5 pmol).

8.2 Experimental

8.2.1 Chemicals and Reagents

Dithiothreitol (DTT), ammonium bicarbaonate (NH_4HCO_3), β-mercaptoethnanol, sodium dodecyl sulfate (SDS), glycerol, bromophenol blue, tris(hydroxymethyl)aminomethane (tris base), glacial acetic acid, trifluoroacetic acid (TFA), methanol, LC-MS grade formic acid (FA), and bovine serum albumin (BSA) protein standard were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). LC-MS grade water and acetonitrile (ACN) were from Fisher Scientific Canada (Edmonton, AB, Canada). ACS grade 37% HCl was from Merck (KGaA, Darmstadt, Germany).

8.2.2 SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE was carried out in a Bio-Rad mini-PROTEAN 3 system using 1 mm thick 12% Tris-HCl and Mini-PROTEAN® TGX™ precast gel of 10 wells of 50 µL from Bio-rad (Mississauga, ON, Canada) were used. The sample was mixed with the same volume of glycine loading buffer (62.5 mM tris base, pH 6.8, 5% β-mercaptoethnanol (v/v), 2% SDS, trace amount bromphenol blue, 12.5% glycerol), heated at 95 ℃ for 5 min, cooled to room temperature, and loaded onto the gel. For the molecular weight markers, the Precision Plus Protein™ All Blue Standard of 10 – 25 kDa from Bio-Rad (Mississauga, ON, Canada) was used. The electrophoresis was performed at 50 mA. The gel was fixed for 30 min in 10% acetic acid /40% Methanol and then stained with Coomassie Brilliant Blue G-250 (Bio-Rad, Mississauga, ON, Canada) for one hour. The gel was destained in water for two hours. For copper stain, the gel was washed in fresh distilled, deionized water (DDI water) for 5 min, stained in diluted Copper Stain from Bio-Rad (Mississauga, ON, Canada) and washed in fresh DDI water for 3 min. The protein bands were excised, transferred to a 1.5 mL polyethylene tube with 1 mL water and stored in -80 ℃ freezer prior to in-gel digestion.

8.2.3 Microwave-Assisted Acid Hydrolysis of Proteins

The protein band of interest was defrost at room temperature, washed with fresh DDI water for 10 min, cut into small pieces, dehydrated in 100% ACN, and completely dried in SpeedVac. Then the gel pieces were covered with 100 µL 9 mM DTT in 100 mM NH_4HCO_3 and reduced at 65 ℃ for 30 min, followed by washing with fresh DDI water, dehydration with ACN, drying in SpeedVac, and

rehydration with 80 μL 3 mM DTT in 3M HCl for 15 min on ice. After rehydration, the sample vial was sealed with Teflon tape, and put in a domestic 1200 W (2450 MHz) microwave oven (Panasonic, London Drugs, Edmonton, Canada). For the original "dry" method, a container with 100 mL of water was placed beside the sample vial to absorb the excess microwave energy. The sample was subjected to 60 s microwave irritation, and then cooled on ice. For the improved water bath method, the sample vial was placed on a Scienceware round bubble rack (Fisher Scientific, Edmonton, Canada) and floated in a plastic beaker, which contained 100 mL of boiled water. The beaker was placed in the center of the rotating plate in the microwave oven. The sample was microwave irradiated for a short period of time in a water bath of boiled water with 240 W power applied.

After in-gel HCl MAAH, the supernatant was removed and transferred to a 1.5 mL polyethylene sample vial. Then the gel pieces were washed with fresh DDI water. The remaining peptides in the gel pieces were extracted twice using 0.25 % TFA in 50% ACN for 30 min with vortex, sonication or rotation by the Labquake® Shaker Tube Rotator (Lab Industries Inc. Berkeley, Ca). At last, the gel pieces were dehydrated in 100% ACN. The extractions and washed solutions as well as the last dehydrated solution were pooled with the original supernatant and dried completely in a SpeedVac vacuum centrifuge (Thermo Savant, Milford, MA) to remove all the acid.

8.2.4 LC Desalting and Fractionation of Hydrolysates

The dried hydrolysates were reconstituted with 105 μL 0.1% TFA, then fractionation was carried out on a 3 mm × 15 cm Zorbax 300 - SB C_3 column with a particle size of 3.5 μm and 300 Å pores (Agilent, Mississauga, ON, Canada) in an Agilent 1100 HPLC system (Palo Alto, CA, USA). After loading of 100 μL polypeptide sample, the column was flushed with 97.5% mobile phase A (0.1% TFA in water) and 2.5% mobile phase B (0.1% TFA in ACN) at a flow rate of 0.4 mL/min for 5 min at room temperature and the salts were effectively removed. Subsequently, the concentration of phase B in the mobile phase was increased to 30% and holded for 9 min to elute the low molecular weight polypeptides from the column, followed by 15 min re-equilibration with mobile phase A.

8.2.5 LC ESI MS/MS Analysis

The collected fractions were dried in the SpeedVac vacuum centrifuge to remove all the solvents and acid, reconstituted with 0.1 % formic acid and analyzed by a quadrupole time-of-flight (QTOF) Premier mass spectrometer (Waters, Manchester, U.K.) equipped with a nanoACQUITY Ultra Performance LC system (Waters, Milford, MA, USA). In brief, 5 μL of peptide solution was injected onto a 75 μm × 150 mm Atlantis dC_{18} column with 3 μm particle size (Waters). Solvent A consisted of 0.1% formic acid in water, and Solvent B consisted of 0.1% formic acid in ACN. Peptides were separated using 30 min gradients (2%–7% Solvent B for 1 min, 7%–25% Solvent B for 19 min, 25%–50% Solvent B for 1 min, 50%–90% Solvent B for 2 min, 90%–90% Solvent B for 6 min, 90%–95% Solvent B for 2 min; the column was pre-equilibrated at 2% Solvent B for 20 min)

and electrosprayed into the mass spectrometer (fitted with a nanoLockSpray source) at a flow rate of 350 nL/min. Mass spectra were acquired from m/z 400 to 1600 for 0.8 s, followed by six data-dependent MS/MS scans from m/z 50–1900 for 0.8 s each. The collision energy used to perform MS/MS was varied according to the mass and charge state of the eluting peptide. Leucine enkephalin and (Glu1)-fibrinopeptide B, a mixed mass calibrant (i.e., lock-mass), was infused at a rate of 300nL/min, and an MS scan was acquired for 1 s every 1 min throughout the run.[38]

8.2.6 Protein Database Search

Raw MS and MS/MS data were lock-mass-corrected, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.3 (Waters). Peptide sequences were identified via automated database searching of peak list files using the MASCOT search program (http://www.matrixscience.com). Database searches were restricted to the protein sequence of the corresponding protein downloaded from the SwissProt database. The following search parameters were selected for all database searching: enzyme, nonspecified; missed cleavages, 0; peptide tolerance, ±30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, (1+, 2+, and 3+); variable modifications, deamidation of asparagine and glutamine. The search results, including unique peptide sequences, ion score, MASCOT threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses were extracted to Excel files. All the identified peptides with scores lower than the MASCOT identity threshold scores for identity were then deleted from the list.

8.2.7 MALDI-TOF MS Analysis

The collected LC elution was completely dried down in SpeedVac. The samples were then reconstituted with 0.1% TFA, mixed with the matrix solution of α-cyano-4-hydrocynnamic acid (CHCA) and spotted on a MALDI plate for matrix-assisted laser desorption ionization (MALDI) analysis.

The MALDI-TOF mass spectra were obtained on an Applied Biosystems/MDS SCIEX 4800 Plus MALDI TOF/TOF Analyzer (AB Sciex, Foster City, CA, USA). Ionization was performed with a diode-pumped Nd: YAG laser at 355 nm. The analyzer was used in a linear or reflector mode of operation. The peptide ion peak picking and mass assignment were done automatically using the peak picking software in the 4800 Plus system. [39]

8.3 Results and Discussion

8.3.1 Method Development

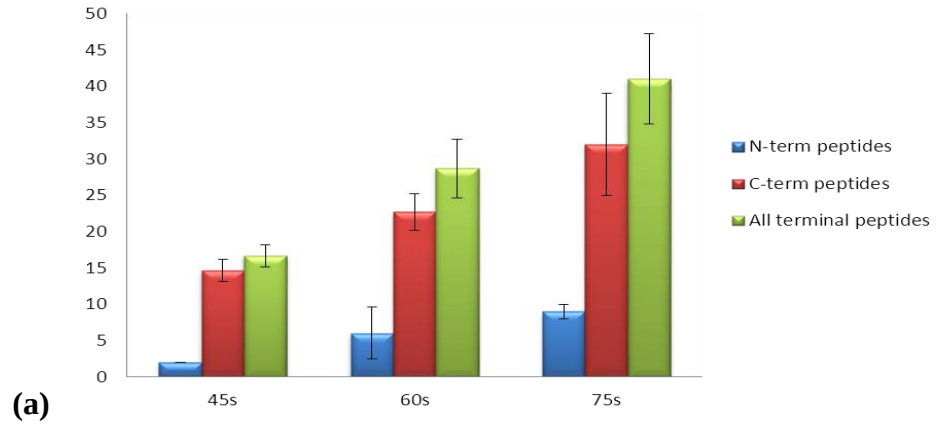
For the gel electrophoresis, no de-staining procedures were applied. The residue copper ions could be removed in desalting steps, while the Coomassie blue would be gradually removed in the in-gel digestion process. Even if a few Coomassie blue went into the gel extract, they precipitated out in 0.1% TFA solution which was used to reconstitute the extracted peptides before HPLC purification and could be removed by centrifugation. At last, HPLC desalting and fractionation step could also remove the Coomassie blue dye from the peptides, if there were any residue dye present. Thus de-staining is unnecessary for in-gel HCl MAAH.

If no reduction was performed before in-gel HCl MAAH, a terminal peptide generated by the acid hydrolysis might still connect to the peptide of the other protein terminus through the disulfide bonds. It is difficult to extract these connected terminal peptides out from the gel, so the following reduction has to be done in-gel after acid evaporation using the SpeedVac. However, it would take forever to evaporate all the acids in the sample vial at the presence of the gel. Furthermore, the terminal peptides would gradually hydrolyze to internal peptides in the drying process. Instead of evaporating the acid, if we chose to neutralize the acidic solution, lots of salts would be generated after the neutralization of 80 μL 3M HCl. The reduction and extraction might also be affected by the salts present. The task is much easier if the reduction is done first. So we performed the reduction before in-gel HCl MAAH.

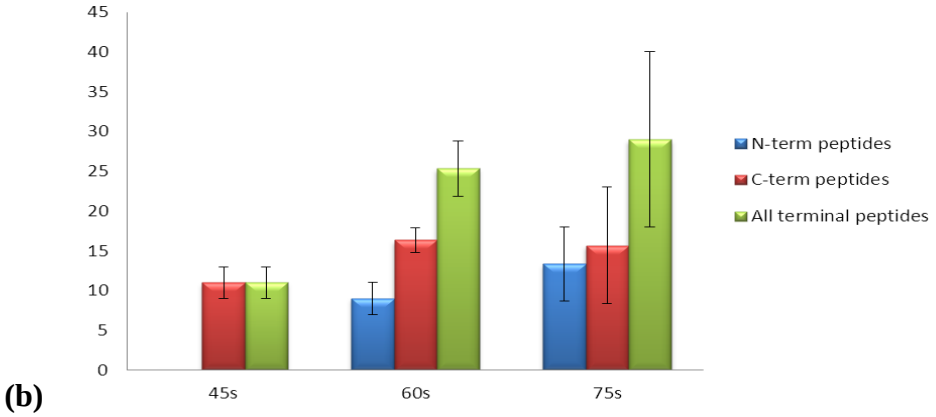
After reduction, gels were washed with DDI water to remove the residue NH_4HCO_3 so that the pH of 3 M HCl used for MAAH would not be affected. The gel was washed after in-gel HCl MAAH to remove the residue HCl in the gel so that the extraction solution would not be too acidic and further hydrolysis of terminal peptides could be reduced. HPLC desalting and fractionation was used to remove the salts present after MAAH and the large peptides whose molecular weights were larger than 4 kDa. Large peptides might clog the pore of the C_{18} column used in LC ESI MS/MS analysis and are difficult to fragment well in MS/MS.

8.3.1.1 Extraction Method from the Gel

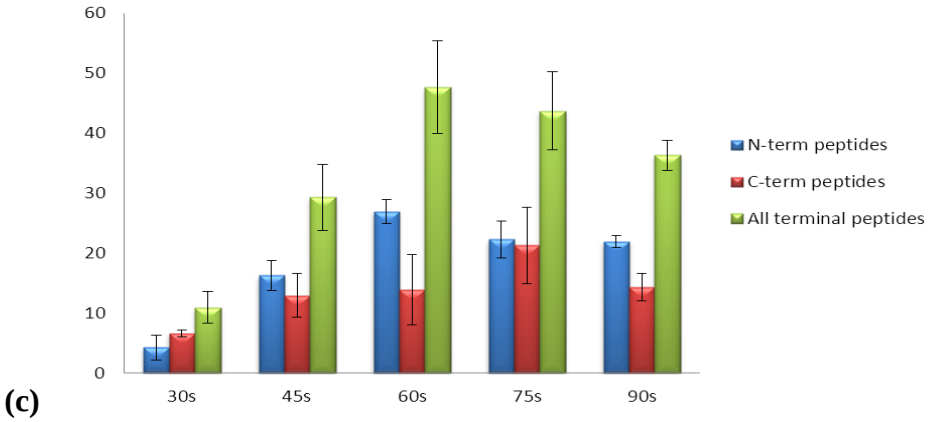
For the in-gel HCl MAAH, BSA was used as a model protein. 10 μg BSA was


50
45
40
35
30
25
20
15
10
5
0
45s
60s
75s
N-term peptides
C-term peptides
All terminal peptides

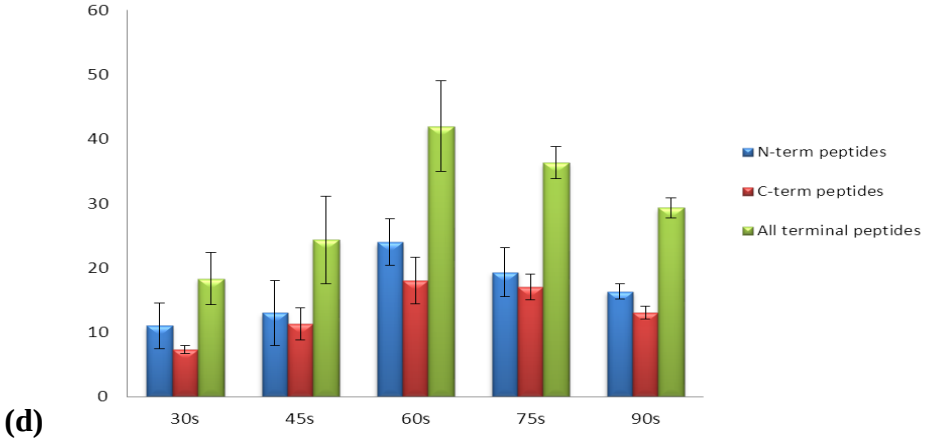
(a)


45
40
35
30
25
20
15
10
5
0
45s
60s
75s
N-term peptides
C-term peptides
All terminal peptides

(b)

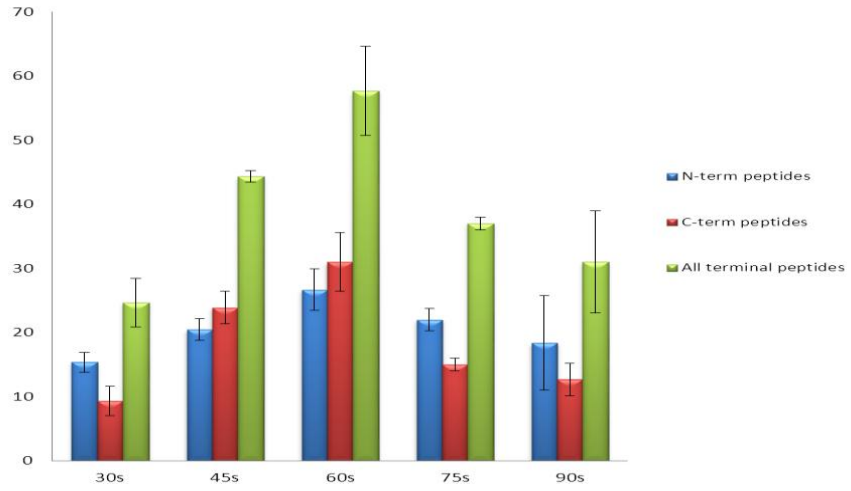

60
50
40
30
20
10
0
30s
45s
60s
75s
90s
N-term peptides
C-term peptides
All terminal peptides

(c)


60
50
40
30
20
10
0
30s
45s
60s
75s
90s
N-term peptides
C-term peptides
All terminal peptides

(d)

(e)



(f)

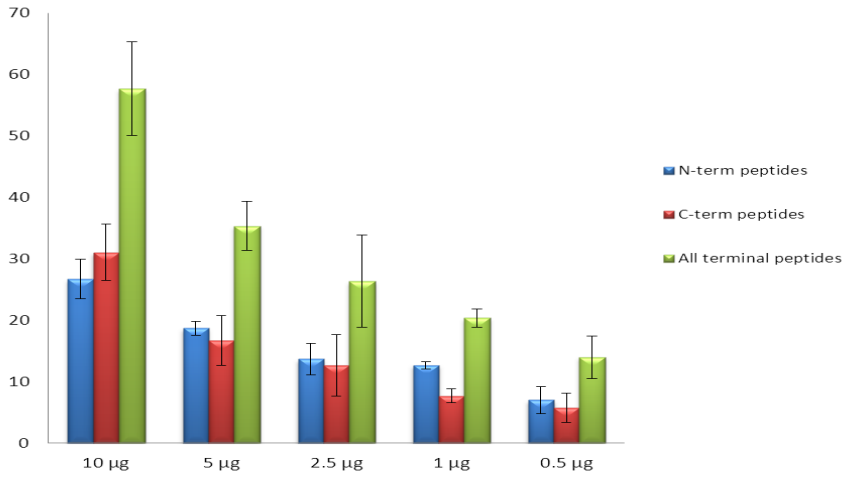


Figure 8.1 The number of terminal peptides of BSA identified in the hydrolysates generated through in-gel HCl MAAH, analyzed by LC ESI MS/MS (n=3). (a) 12% tris-HCl gel and Coomassie Brilliant Blue G-250 stain were used. BSA HCl "dry" MAAH for 45, 60, and 75 s. (b) 12% tris-HCl gel and copper stain were used. BSA HCl "dry" MAAH for 45, 60, and 75 s. (c) 12% tris-HCl gel and Coomassie Brilliant Blue G-250 were used. BSA HCl MAAH for 30, 45, 60, 75 and 90 s in a water bath of boiled water. (d) 12% tris-HCl gel and copper stain were used. BSA HCl MAAH for 30, 45, 60, 75 and 90 s in a water bath of boiled water. (e) Mini-PROTEAN® TGX™ precast gel and Coomassie Brilliant Blue G-250 were used. BSA HCl MAAH for 30, 45, 60, 75 and 90 s in a water bath of boiled water. (f) Mini-PROTEAN® TGX™ precast gel and Coomassie Brilliant Blue G-250 were used. The amounts of BSA loaded into the gel were 10, 5, 2.5, 1 and 0.5 µg, respectively.

used for all the method development. First we followed the original "dry" method for the in-solution MAAH described by Zhong *et al*[37] to test the effects of different peptide extraction methods after MAAH. The 12% tris-HCl gel pieces were vortexed or rotated in a 4 °C cool room, or sonicated in an iced-water bath for 60 min in total. The temperature was strictly controlled to prevent the further hydrolysis of terminal peptides in acidic conditions of the extraction process. Table 8.1 lists the number of identified terminal peptides from BSA by the three extraction methods from the gel. The extraction by vortexing, sonication or rotation identified 24 ± 8, 29 ± 4 and 26 ± 6 peptides, respectively. The numbers were not significantly different at the 95% confidence level by t-test. We chose sonication extraction for all the subsequent experiments, considering it gave the largest number of peptides identified and had the least variations.

8.3.1.2 MAAH Methods: "Dry" vs. Water Bath

Following the "dry" MAAH method developed by Zhong *et al.* ,[37] the 12% tris-HCl gel pieces with Coomassie blue G-250 stain or copper stain were microwave irradiated for 45, 60 and 75 s separately. Figures 8.1a and 8.1b demonstrate that the number of identified peptides of BSA increased with the MAAH time. 41 ± 6 and 29 ± 11 terminal peptides were identified for the Coomassie blue G-250 stained and copper stained gels, respectively. For the terminal peptides analysis, generally the number of identified terminal peptides would increase with the MAAH time at first, then reach the maximum at a point, and decrease with the longer hydrolysis time after that point. A 75 s MAAH time might not be the optimal and thus longer hydrolysis time might be needed for the "dry" method.

Table 8.1 The number of identified terminal peptides of BSA by different extraction methods from the gel after in-gel HCl MAAH.

Extraction ways	N-terminal peptides	C-terminal peptides	All terminal peptides
Vortexing	4 ±2	20 ±7	24 ±8
Sonication	6 ±4	23 ±3	29 ±4
Rotation	3 ±3	23 ±4	26 ±6

The hydrolysis speed is slower, compared to the in-solution HCl MAAH method whose optimal time is 60 s. Furthermore, “hot” spots are located at different postions of different microwave ovens. The “dry” method requires careful locatization of the “hot” spots first, which is a long and tedious process.

We tried to improve the “dry” MAAH method with a water bath apparatus described by Wang *et al*.[40] Reproducible hydrolysis results of BSA protein were obtained for the in-slution HCl MAAH, as described in Chapter 7. The optimal hydrolysis condition for in-slution HCl MAAH was found to be microwave irradiated for 45 s at 240 W in a water bath of boiled water and 69 ± 1 terminal peptides could be identified by LC-ESI MS/MS analysis. We used this condition to do the in-gel HCl MAAH, except the microwave irradiation time. Different

periods of time from 30 to 90 s were tested for the in-gel HCl MAAH using a water bath. Figures 8.1c and 1d show that 60 s MAAH could identify the largest number of peptides. 48 ± 8 and 42 ± 7 terminal peptides were identified for the Coomassie blue G-250 stained and copper stained tris-HCl gels, respectively. The identified terminal peptides in-gel were less than in-solution, but it was justified considering the low extraction efficiency of peptides from the gel. The longest peptide identified for in-gel MAAH was about 2.5 kDa, while the peptide from in-solution MAAH was nearly 3.4 kDa. In-gel HCl MAAH required 15 s longer microwave irradiation time than the in-solution method. Because of the hindrance of the gel in the hydrolysis process, the protein in-gel takes a longer time to hydrolyze to the same extent as a protein in-solution which moves freely without the gel.

8.3.1.3 Different Staining Methods: Coomassie Blue G-250 Stain vs. Copper Stain

Compared to the widely used Coomassie blue stain, copper stain has a fast staining process and the stain is reversible. Many researchers used copper stain to facilitate downstream MS analysis, even though it is a negative stain. Although no significant difference of terminal peptides identified for the Coomassie blue G-250 stained and copper stained tris-HCl gels by t-test at 95% confidence level, the former had a slightly higher average number than the latter for the water bath method. Overall, the staining method did not affect the results significantly.

8.3.1.4 Different Gel Types: 12% Tris-HCl vs. Mini-PROTEAN® TGX™ Precast Gel

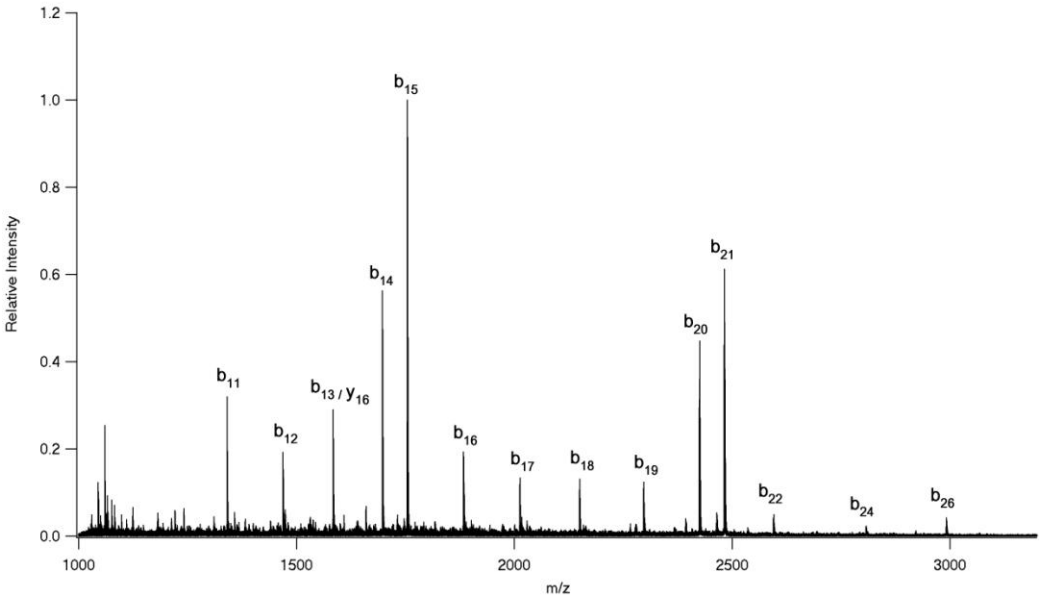
Mini-PROTEAN® TGX™ precast gel is an upgraded precast gel for Tris-HCl gel from Bio-Rad. It has better resolution and higher loading capacities. Copper stain does not work on this kind of gel due to the different gel chemistry. Thus only Coomassie blue G-250 stained gels were tested. In-gel HCl MAAH was performed for 30, 45, 60, 75 and 90 s separately in a water bath

of boiled water. Figure 8.1e demonstrates that the optimal hydrolysis time for Mini-PROTEAN® TGX™ precast gel was also 60 s. 58± 8 terminal peptides were identified, more than the Tris-HCl gel but still less than the in-solution HCl MAAH results. Different gel types somehow affect the hydrolysis processes differently, though the differences are not significant. Overall, in-gel HCl MAAH for 60 s at 240 W in a water bath of boiled water is applicable for the two types of gels tested. Terminal peptide ladders could be identified by this method.

8.3.2 Sensitivity Test

We loaded 10, 5, 2.5, 1 and 0.5 μg BSA into a Mini-PROTEAN® TGXTM precast gel and stained the gel with Coomassie Brilliant Blue G-250. In-gel HCl MAAH was performed for 60 s at 240 W in a water bath of boiled water for all these samples. Figure 8.1f shows that the identified terminal peptides decreased with the decreasing loading amounts of BSA in-gel. When the BSA amount dropped to 0.5 μg (7.5 pmol), only about 10 terminal peptides in total were

(a)



(b)

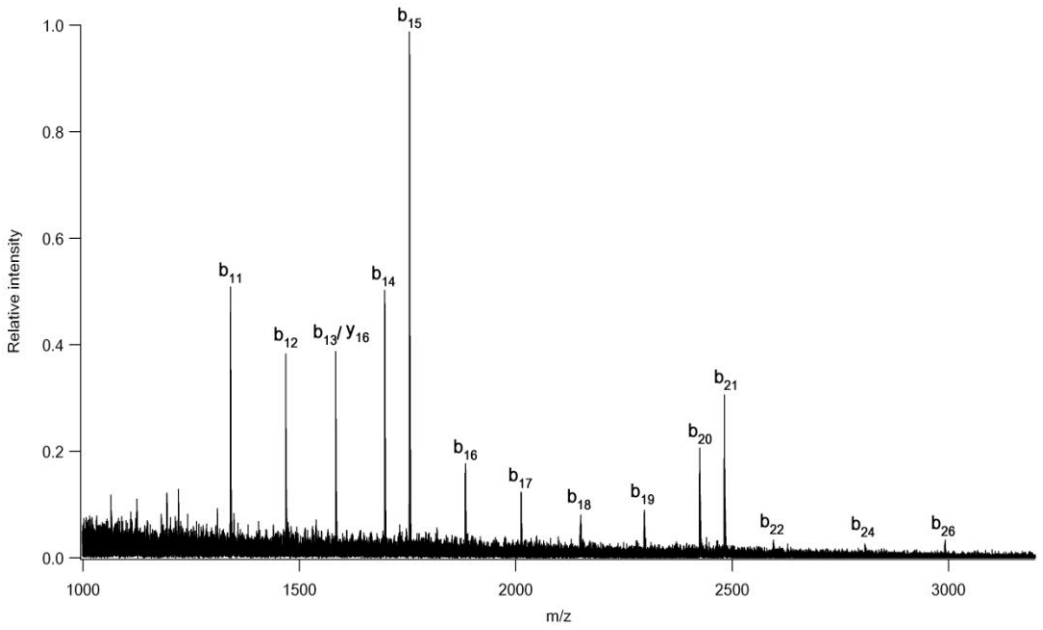


Figure 8.2 The spectra of low molecular weight peptides in BSA hydrolysates generated by (a) in-solution HCl MAAH digestion, and (b) in-gel HCl MAAH digestion.

identified. If the amount increased to 2.5 μg (37.5 pmol), over 10 terminal peptides were found for each term of the protein. Although we could identify terminal peptides with 0.5 μg protein, we preferred to load more sample, at least 2.5 μg, into the gel for better results.

8.3.3 MALDI-TOF MS Analysis of Low-molecular-Weight Peptides

We also used MALDI-TOF MS to analyze collected low molecular weight peptides in BSA hydrolysates generated by in-gel HCl MAAH digestion. Figure 8.2 shows the results. Although the spectrum of the BSA hydrolysates digested in-gel was noisier than the one of BSA hydrolysates digested in-solution, we could identify the same terminal peptides as in-solution digestion. The intensities of terminal peptides were much higher than their adjacent peaks, making them easy to be distinguished from others in the spectrum. We found 14 peaks matched with the terminal peptides of BSA within m/z ± 0.3 Da. The mass differences of peptides y_{16} and b_{13} are only 0.1 Da, so in the spectrum we cannot tell which one is correct by MS only. The ion suppression in MALDI was so severe that almost all the C-terminal peptides of BSA were suppressed by N-terminal peptides. Thus for detecting the low molecular weight terminal peptides generated by in-gel HCl MAAH, LC ESI MS/MS analysis is a preferred method over MALDI-TOF MS.

8.4 Conclusions

We have developed an in-gel HCl MAAH method for terminal peptide analysis. We improved the “dry” MAAH method by using awater bath of boiled water. In

addition to faster hydrolysis process, more reproducible MAAH results were obtained using the water-bath method. Different methods of extracting peptides from the gel after MAAH were examined and it was found that vortexing, sonication and rotation, gave similar results. The method of gel staining on the detectability of peptides from in-gel HCl MAAH was also studies.Coomassie blue stain and copper stain did not show significant differences. With Coomassie blue stain (G-250 stain), mini-PROTEAN® TGXTM precast gel outperformed Tris-HCl gel due to different gel chemistries. The number of terminal peptides identified by in-gel MAAH was less than in-solution MAAH.It was shown that the identified terminal peptides decreased with the decreasing loading amounts of BSA in-gel. Although terminal peptides could be identified using 0.5 μg protein in-gel, more proteins, at least 2.5 μg, were preferred to load into the gel for producing better results.On the effect of MS detection method, it was found that ion suppression in MALDI was much more severe than ESI. For the analysis of terminal peptides generated by in-gel HCl MAAH, LC ESI MS/MS is more appropriate than MALDI-TOF MS.

8.5 Literature Cited

(1) Liang, X.; Bai, J.; Liu, Y. H.; Lubman, D. M. *Anal Chem* **1996**, *68*, 1012.
(2) Cooper, J. W.; Gao, J.; Lee, C. S. *Electrophoresis* **2004**, *25*, 1379.
(3) Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 1791.
(4) Mortz, E.; Sareneva, T.; Haebel, S.; Julkunen, I.; Roepstorff, P. *Electrophoresis* **1996**, *17*, 925.
(5) Ehring, H.; Stromberg, S.; Tjernberg, A.; Noren, B. *Rapid Communications in Mass Spectrometry : RCM* **1997**, *11*, 1867.

(6) Jeannot, M. A.; Zheng, J.; Li, L. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 512.
(7) Galvani, M.; Bordini, E.; Piubelli, C.; Hamdan, M. *Rapid Communications in Mass Spectrometry : RCM* **2000**, *14*, 18.
(8) Mirza, U. A.; Liu, Y. H.; Tang, J. T.; Porter, F.; Bondoc, L.; Chen, G.; Pramanik, B. N.; Nagabhushan, T. L. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 356.
(9) Cohen, S. L.; Chait, B. T. *Anal Biochem* **1997**, *247*, 257.
(10) Lahm, H. W.; Langen, H. *Electrophoresis* **2000**, *21*, 2105.
(11) van Montfort, B. A.; Canas, B.; Duurkens, R.; Godovac-Zimmermann, J.; Robillard, G. T. *J Mass Spectrom* **2002**, *37*, 322.
(12) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal Chem* **1996**, *68*, 850.
(13) Hellman, J. *Proteomics* **2007**, *7*, 3441.
(14) Zhou, J.; Hu, Y.; Lin, Y.; Liu, H.; Xie, P. *Journal of chromatography. B,* **2011**, *879*, 2957.
(15) Medzihradszky, K. F.; Leffler, H.; Baldwin, M. A.; Burlingame, A. L. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 215.
(16) Palviainen, M.; Raekallio, M.; Vainionpaa, M.; Kosonen, S.; Vainio, O. *Toxicon : official journal of the International Society on Toxinology* **2012**, *60*, 1228.
(17) Li, G.; Waltham, M.; Anderson, N. L.; Unsworth, E.; Treston, A.; Weinstein, J. N. *Electrophoresis* **1997**, *18*, 391.
(18) Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P. R. *Rapid Communications in Mass Spectrometry : RCM* **2000**, *14*, 496.
(19) Klarskov, K.; Roecklin, D.; Bouchon, B.; Sabatie, J.; Van Dorsselaer, A.; Bischoff, R. *Anal Biochem* **1994**, *216*, 127.
(20) van Montfort, B. A.; Doeven, M. K.; Canas, B.; Veenhoff, L. M.; Poolman, B.; Robillard, G. T. *Biochimica et Biophysica Acta* **2002**, *1555*, 111.
(21) Borchers, C.; Peter, J. F.; Hall, M. C.; Kunkel, T. A.; Tomer, K. B. *Anal Chem* **2000**, *72*, 1163.
(22) Bantscheff, M.; Weiss, V.; Glocker, M. O. *Biochemistry-Us* **1999**, *38*, 11012.
(23) Sutton, C. W.; Pemberton, K. S.; Cottrell, J. S.; Corbett, J. M.; Wheeler, C. H.; Dunn, M. J.; Pappin, D. J. *Electrophoresis* **1995**, *16*, 308.
(24) Sickmann, A.; Dormeyer, W.; Wortelkamp, S.; Woitalla, D.; Kuhn, W.; Meyer, H. E. *Electrophoresis* **2000**, *21*, 2721.
(25) Larsson, T.; Norbeck, J.; Karlsson, H.; Karlsson, K. A.; Blomberg, A. *Electrophoresis* **1997**, *18*, 418.
(26) Meyer, B.; Papasotiriou, D. G.; Karas, M. *Amino Acids* **2011**, *41*, 291.
(27) Ogorzalek Loo, R. R.; Stevenson, T. I.; Mitchell, C.; Loo, J. A.; Andrews, P. C. *Anal Chem* **1996**, *68*, 1910.
(28) Patterson, S. D. *Electrophoresis* **1995**, *16*, 1104.
(29) Ogorzalek Loo, R. R.; Mitchell, C.; Stevenson, T. I.; Martin, S. A.; Hines, W. M.; Juhasz, P.; Patterson, D. H.; Peltier, J. M.; Loo, J. A.; Andrews, P. C. *Electrophoresis* **1997**, *18*, 382.

(30) Quach, T. T.; Li, N.; Richards, D. P.; Zheng, J.; Keller, B. O.; Li, L. *J Proteome Res* **2003**, *2*, 543.
(31) Hua, L.; Low, T. Y.; Sze, S. K. *Proteomics* **2006**, *6*, 586.
(32) Inglis, A. S. *Methods in Enzymology* **1983**, *91*, 324.
(33) Li, A.; Sowder, R. C.; Henderson, L. E.; Moore, S. P.; Garfinkel, D. J.; Fisher, R. J. *Anal Chem* **2001**, *73*, 5395.
(34) Xiang, B.; Macisaac, S.; Lardizabal, K.; Li, B. *Rapid Communications in Mass Spectrometry : RCM* **2010**, *24*, 3447.
(35) Sun, W.; Gao, S.; Wang, L.; Chen, Y.; Wu, S.; Wang, X.; Zheng, D.; Gao, Y. *Mol Cell Proteomics* **2006**, *5*, 769.
(36) Juan, H. F.; Chang, S. C.; Huang, H. C.; Chen, S. T. *Proteomics* **2005**, *5*, 840.
(37) Zhong, H. Y.; Zhang, Y.; Wen, Z. H.; Li, L. *Nat Biotechnol* **2004**, *22*, 1291.
(38) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Anal Chem* **2009**, *81*, 1049.
(39) Reiz, B.; Li, L. *J Am Soc Mass Spectr* **2010**, *21*, 1596.
(40) Wang, N.; Li, L. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1573.

Chapter 9

Conclusions and Future Work

Sample preparation plays a very important part prior to MS analysis in proteomics. Generally speaking, it includes protein digestion by enzyme or chemicals, protein/peptide separation or enrichment, and sample cleanup. However, the current sample preparation methods do not meet all the needs of proteome analysis. The goal of my thesis work was to develop and apply new or improved protein sample preparation methods for MS analysis of various types of samples with different needs of protein characterization.

After introducing the related techniques to my thesis work in Chapter 1, I described the development of an inexpensive isotope labeling method (2-MEGA or dimethylation after guanidination), in Chapter 2, for quantitative proteome analysis which was based on a previously reported isotope labeling chemistry. A commercially available liquid handler was used for sample preparation to minimize variability from sample handling during the labeling reaction for high throughput applications. In addition, compatibility with various front end protein preparation methods was demonstrated. Using commonly used buffer and surfactant conditions, over 94% of tryptic peptides were correctly labeled; in the case of urea-containing buffers, 92% correct labeling was obtained. This work illustrates that the optimized 2-MEGA labeling method can be used to handle a wide range of protein samples for quantitative proteomics.

In Chaper 3, we developed a microbore LC-UV quantification method for determining the total amount of peptides generated from a proteomic sample preparation process, particularly suitable for few cell proteomics where the amount of the starting materials is limited. With a 1.0 mm i.d. column, the LOQ was found to be about 40 ng which is adequate for quantifying peptides generated from the proteins extracted from a few hundred cells. The relative standard deviation in signal response at different sample loading amounts on the same day were <10%, while the day-to-day relative standard deviations were <15%. This LC-UV method can also be used for desalting which improves the LC-MS performance for peptide identification. Thus, the microbore LC-UV method can be used to track the sample preparation process during the proteomic analysis of a few cells. We envisage this microbore LC-UV method will be useful in proteome profiling of small numbers of CTCs isolated in blood of patients with cancer or primary cells procured from tumor tissue samples by LCM for disease diagnosis and prognosis.

In order to find biomarkers of breast cancer, the 2-MEGA labeling method was applied to breast cancer tissues for relative proteomic comparison with the normal breast tissues in Chapter 4. Lipids and detergents were successfully removed by acetone precipitation and no interference from the acid-labile surfactant proteasMAX used for protein solubilization after acetone precipitation. The labeling efficiencies were ~94% for the breast tumor tissues, similar as that of other samples. Forward and reverse labeling greatly reduced the quantification errors. Three individual tumor samples were analyzed and ~4000 to 8000 peptides and 1500 to 2500 proteins were identified for each 2D LC MS/MS analysis. 606,

880 and 456 differentially expressed proteins with a quantification ratio of either less than 0.67 or larger than 1.5, compared to a pooled normal tissue sample, were found in cases CT0018, MT1275 and MT699, respectively. 119 proteins were found differentially expressed in all three cases. The common differentially expressed proteins are putative biomarkers. Future work will be needed to verify and validate the potential use of some of the proteins as biomarkers for improved breast cancer diagnosis and prognosis using a large set of tissues samples.

The work described in Chapters 5-8 was based on microwave-assisted acid hydrolysis (MAAH) of proteins combined with MS for protein sequence analysis. In Chapter 5, we developed a method that combines electroelution and LC ESI MS/MS analysis for sequencing proteins separated by SDS-PAGE. It was found that the electroelution apparatus of Bio-Rad concentrated SDS and salts in protein elution solution, bringing a major challenge to the downstream purification work. We applied equilibrium dialysis to remove NH_4HCO_3 first, and then used acetone precipitation to remove SDS and Coomassie blue at the same time. Recovery of proteins (BSA as a model protein) was found to be around 25% by LC-UV quantification. MALDI-TOF MS analysis shows the spectrum of the purified protein electroeluted from the gel was similar to that of the standard protein. LC ESI MS/MS analysis was applied to analyze the low molecular weight peptides in the MAAH hydrolysates of the electroeluted protein. 12 N-terminal peptides and 8 C-terminal peptides were found in a 15 min LC ESI MS/MS analysis. However, the problem of this method is that many purifying steps are needed for the MS analysis and thus the protein recovery rate is rather low; the recovery rate was only 25% for 40 μg BSA loaded onto the gel. The relatively large sample amounts

needed for analysis makes this method have limited application areas. Nevertheless, if a large amount of proteins is available for analysis, electroelution combined with MAAH and LC-ESI MS/MS can be used for sequencing proteins separated by SDS-PAGE.

In Chapter 6, we developed a relatively simple protocol for analyzing terminal sequences of a large protein. BSA was used as a model protein to optimize the process and demonstrate the performance of the protocol. RPLC fractionation simplifies the protein hydrolysates generated by MAAH and the collected low-molecular-weight peptides are suitable for LC ESI MS/MS analysis. Compared to MALDI-TOF, LC ESI MS/MS is more appropriate for terminal peptide sequencing due to its reduced ion suppression and unambiguous identification of the peptides based on CID MS/MS. The total analysis is less than 7 hr with minimum skill requirement for sample handling and data analysis. This protocol should be useful for rapid characterization of terminal sequences of proteins, such as for quality control of proteins during the protein production process and storage.

Chapter 7 focuses on the comparison of household microwave oven with a commercial microwave device for MAAH. We also improved the in-solution HCl MAAH method for terminal peptide analysis using the household microwave oven with a water bath of boiled water. The results were more reproducible than the "dry" method and there was no hotspot problem associated with the household microwave oven heating. The MAAH methods using HCl and TFA developed with the use of household microwave oven were successfully

transferred into the commercial device (CEM Discover). The peptide identification results from LC ESI MS/MS analysis of hydrolysates were almost the same. High temperature and high power microwave irradiation were found to be critical for fast acid hydrolysis. The microwave power was lowered in HCl MAAH to control the hydrolysis speed so that the most peptides generated were terminal peptides. We applied the two MAAH methods to analyze the variants of human and bovine hemoglobins. While we could readily differentiate the normal and sickle cell hemoglobins, MAAH methods were difficult to distinguish the variants with mutations of 1 Da mass differences, especially due to the more frequent deamidation in MAAH. For mutations with larger mass differences, MAAH is useful to determine the nature and the position of the mutation. We believe that HCl and TFA MAAH MS are the complementary methods to other protein characterization tools such as gel electrophoresis for detecting new hemoglobin variants. Future work should be focused on the applications of these methods for analyzing new hemoglobin variants. It is also worth exploring the feasibility of these methods for rapid characterization of certain hemoglobin variants such as sickle cell hemoglobin for clinical applications. Direct analysis of blood samples without much sample preparation prior to MAAH and MS analysis may prove to be attractive for clinical diagnosis of sickle cell or other hemoglobin variants.

To improve the HCl MAAH method for characterizing gel-separated proteins, in-gel HCl MAAH was developed and described in Chapter 8. Different extraction methods from the gel after MAAH, including vortexing, sonication and rotation, gave similar results, but sonication in an iced water bath allowed the detection of

slightly more terminal peptides. Proteins in-gel with Coomassie blue stain G-250 stain could identify a few more terminal peptides than copper stain, although the differences were not significant. With Coomassie blue stain G-250 stain, the mini-PROTEAN® TGXTM precast gel outperformed the Tris-HCl gel due to different gel chemistries. The number of terminal peptides identified by in-gel MAAH was less than in-solution MAAH, mainly because the hindrance of the gel in the hydrolysis process made the protein in-gel take a longer time to hydrolyze to the same extent and possible sample loss during the extraction process. The identified terminal peptides decreased with the decreasing loading amounts of BSA in-gel. Although terminal peptides could be identified using 0.5 μg protein in-gel, the use of at least 2.5 μg would be recommended to load onto the gel for better results. Compared to the electroelution method described in Chapter 5, in-gel MAAH appears to provide much higher sensitivity for terminal peptide analysis.

As stated above, I have developed a series of sample preparation methods for different purposes of protein characterization. Some work still needs to be done for improving or demonstrating real sample applications of the method or studying the problems encountered during the course of method development. For example, if we want to achieve the goal of proteome analysis of a single cell or a few cells, the sample procedures should be optimized to reduce the sample loss. Acetone precipitation should be avoided and thus any detergent should not be used for protein extraction. It may prove to be better if we digest the cellular proteins on-column *in situ,* followed by direct analysis of the peptides by LC MS/MS. The column used in LC MS analysis can also be replaced with even smaller i.d. to gain higher sensitivity.

In the electroelution work, it was surprising to find the salt and SDS were concentrated during the electroelution process. Unless a better electroelution apparatus is developed, the purifying procedures as described in this thesis are necessary to remove the salts and SDS present in the elution solution. Maybe changing to another electroelution apparatus can simplify the purification process, assuming the salts and SDS are not concentrated in the protein elution. With a simpler purification process, sample loss will be minimized and the overall detection sensitivity should be improved.

For MAAH, HCl MAAH is a good method for terminal peptide analysis and TFA MAAH is good for the whole sequence analysis. A mutant or modified sequence can be verified by combining the two MAAH methods with LC ESI MS analysis. The applications of these methods for real world applications should be explored to characterize the sequences of proteins present in various biological samples. Because MAAH requires the proteins to be relative pure, protein separation before MAAH is very important. Considering SDS-PAGE is a power protein separation technique, gel separation followed by MAAH should be a very useful technique for protein sequence analysis. In-gel HCl MAAH has higher sensitivity than electroelution with MAAH, in addition to simpler procedures and faster sample preparation process. The protein digestion only takes about 1 min instead of overnight for in-gel trypsin digestion and few gel-induced modifications are observed. The protein sequencing applicability of real samples by in-gel HCl MAAH should be illustrated in the future.

In summary, proteomics deals with many different types of samples and tries to answer many different questions on proteins. This research field requires many different analytical tools that should be complementary to each other. By combining a set of these tools, questions related to biological processes, medical and clinical applications, industrial production and quality controls can be better answered. My thesis work contributes to the development of these enabling tools for better protein or proteome characterization.