University of Alberta

Development and Applications of Quantitative Mass Spectrometric Methods for Proteome and Protein Analysis

by

Yanan Tang

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

©Yanan Tang

Spring 2013
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

Quantitative proteomics has been increasingly recognized as a key research topic in proteomics research, as it enumerates changes of proteins' expression level in a bio-system. Quantitative proteomics provides an important way to discover biomarkers for disease diagnosis and prognosis, and also to understand biological processes and mechanisms. This thesis describes the method optimization in shotgun proteomic sample preparation and 2-MEGA (dimethylation after guanidation) labeling, which is known as a stable isotopic labeling technique. In this work, a LC-UV quantification method was developed to evaluate sample integrity in shotgun proteome sample preparation. Based on the quantification results obtained with this method, MS analysis parameters could be optimized to ensure better proteome coverage in profiling a proteomic sample. Also, in shotgun proteomic sample preparation for quantification, the compatibility of 2-MEGA labeling chemistry with commonly used cell lysis buffer, salts and detergents was evaluated to ensure better than 95% correct labeling of the peptides. In all tested reaction conditions, the most common side reaction products were N-terminal guanidination (~2%) for glycine and alanine N-terminal peptides. Various protein sample preparation methods were found to be compatible with 2-MEGA labeling.

After the optimization of the sample preparation process, the 2-MEGA labeling method was applied to comparative analysis of HER2 normal and overexpressing MCF-7 human breast cancer cell lines, as well as plasma samples of pig model

collected at different stages and treatments of deep tissue injury (DTI). In the breast cancer cell experiment, 4 potential biomarkers with different expression level change in HER2 normal and HER2 overexpressing cell lines were identified. In the pig plasma sample analysis, a number of differentially expressed proteins were identified that may potentially be used as biomarkers for diagnosing and studying DTI. In the future, more biological evaluation work needs to be done to reveal more information for their prognostic values.

In addition to comparative proteomic quantification, my thesis work also included the development of a novel MS-based absolute quantification method for a mixture of N-truncated protein and intact protein. It was based on dansyl labeling reaction on N-terminal free amine of proteins. Microwave-assisted acid hydrolysis (MAAH) were performed after dansylation to release N-terminal amino acids with the dansyl label, which were quantified on MS with internal standards of dansyl labeled amino acids.

Taken together, the thesis work developed several new analytical methods to facilitate protein quantification and illustrated an increasingly important role of quantitative mass spectrometry in meeting the bioanalytical challenges in protein and proteome characterization.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2D	two dimensional
2MEGA	dimethylation after guanidination
AALS	anionic acid labile surfactant
AQUA	absolute quantification
CE	capillary electrophoresis
CID	collision induced dissociation
DMF	dimethylformamide
DTI	deep tissue injury
ECL	enhanced chemiluminescence
EIC	extracted ion chromatograms
ErbB-2	epidermal growth factor receptor 2
ERPLC	electrostatic repulsion-hydrophilic interaction chromatography
ESI	electrospray ionization
FTICR	Fourier transform ion-cyclotron resonance
HCCA	α-cyanohydroxycinnamic acid
HILIC	hydrophilic interaction chromatography
IEF	isoelectric focusing
LC	liquid chromatography
LCM	laser capture microdissection
LOD	limit of detection
LOQ	limit of quantification
IES	intermittent electrical stimulation
IHC	immunohistochemistry
m/z	mass to charge ratio
MAAH	microwave assisted acid hydrolysis
MALDI	matrix assisted laser desorption ionization
MCF-7	Michigan cancer foundation - 7
MRM	multiple reaction monitoring

MRI	magnetic resonance imaging
MS	mass spectrometry/mass spectrometric
MS/MS	tandem mass spectrometry
NPLC	normal phase liquid chromatography
QQQ	triple quadrupole
QconCAT	quantification concatamer
QTOF	quadrupole time of flight
RF	radiofrequency
ROC	receiver operating curve
RP	reverse phase
RPLC	reverse phase liquid chromatography
RSD	relative standard deviations
SAX	strong anion exchange
SCI	spinal cord injury
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labelling by amino acids in cell culture
TFA	trifluoroacetic acid
TCA	trichloroacetic acid
TOF	time of flight
UV	ultraviolet

Chapter 1 - Introduction

Unlike conventional ways of studying biological systems, where information of only one or a few proteins is studied at a given time, proteomics involves the simultaneous analysis of hundreds to thousands of proteins from a defined system. Depending on the type of biological question to be addressed, proteome analysis takes several different forms. In some studies, proteome analysis only requires the generation of a protein profile present in a bio-sample[1-4]. In other cases, quantitative proteome analysis is performed[5-7]. In recent decades, quantitative analysis of protein expression level changes at the proteome level is gaining more attentions, as it provides deeper information, compared to generating a protein list, for understanding a biological process such as protein signal pathways.

My thesis focuses on the quantification of proteins in a proteomic sample. Specifically, we wish to quantify proteins from cells, bio-fluids, or pharmaceutical protein products, with an ultimate goal of developing effective methods for searching potential protein biomarkers in pre-clinical study, or for quality control in pharmaceutical production. In my work, mass spectrometry (MS) is used for proteome analysis. All of the technical developments presented in this thesis are centered on the use of mass spectrometry for relative and absolute protein quantification. There are many excellent reviews on each topic of MS and proteome analysis and thus I do not intend to cover all areas in detail in this chapter. Rather I will concentrate on discussing the most relevant topics to my thesis work. In this chapter, protein sample preparation methods related to MS analysis will be described, followed by a brief introduction of MS and MS/MS, particularly quadrupole time-of-flight (QTOF) MS which I have used extensively for my theis work. The techniques of MS-based proteomic quantification will also be discussed. Finally, the scope of my thesis will be given.

1.1 Overview of MS-Based Shot Gun Proteome Analysis

1.1.1 Protein Purification

Most proteome samples are those from extracted cells, tissues or biofluids, in which samples derived from cultured cells are most commonly used in lab research. In proteome sample preparation, protein extraction is the first key step. Protein extraction from cells often involves rupturing the cellular membrane via physical action (e.g., sonication, pressure or freeze and thaw), the use of extraction buffers containing buffers that weaken cellular structure, or a combination of both. Proteins from other sources, like tissue samples, could require more elaborate sample preparation procedures to remove cellular components and extracellular material that may interfere with MS analysis. Removal of lipids, salts and other small organic molecules can be effectively achieved with protein precipitation using a solvent[8] such as acetone or trichloroacetic acid (TCA).

Once proteome samples are extracted and reasonably purified, a variety of analytical separation methods can be used to further separate proteins. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is one of the most widely used protein level separation methods. SDS-PAGE is noted for its high resolution but laborious work procedure and poor reproducibility. Other widely used separation methods for proteomics are solution-based, such as isoelectric focusing (IEF) and protein reversed phase liquid chromatography (RPLC). These methods offer the advantage that proteins are recovered in solution and can be subjected to additional sample work-up procedures at both protein and peptide level. These solution-based proteome separation methods can also be coupled with antibodies or other binding molecules to specifically enrich certain classes of proteins, like lectins and periodate oxidation/hydride resins used to select glycoproteins[9].

1.1.2 Protein Digestion and Peptide Level Separation

Proteins are usually subjected to digestion after proper purification. This step is aim to cut proteins of interest into smaller peptides, and can be achieved by

enzymatic or chemical methods. The most common protein digestion method utilizes trypsin, an enzyme which cleaves C-terminally to lysine and arginine in peptides, unless the next residue is proline[10]. Other commonly used enzymes include chymotrypsin, Lys-C, Lys-N and Glu-C/V8. Each of these enzymes has their own specific cutting positions. For enzymatic digestion, it's desirable to perform with proteins solubilized in their denatured forms so that many potential cleavage sites can be accessible for enzyme, thus eliminating the possibility of mis-cleavage.

Chemical methods are less frequently employed for proteomics experiments because of their non-specific cutting positions. But they could be used to address specific challenges. For example, microwave-assisted acid hydrolysis (MAAH) has been suggested for sequencing hydrophobic proteins insoluble in most aqueous digestion buffers[11].

Like proteins, peptides can also be separated using similar separation methods, such as isoelectric focussing[12] and capillary electrophoresis. In the case of peptides, they are relatively small in sizes, compared with proteins, and have less interaction among themselves. Therefore more varieties of separation techniques based on liquid chromatography (LC) could be applied to peptide separation, like ionic exchange chromatography, normal phase LC (NPLC) and reversed phase LC (RPLC). Considering the compatibility of most RPLC solvents and additives to electrospray ionization, RPLC separation is typically the last dimension of separation before mass spectrometry.

Strong cation exchange (SCX) at low pH is another widely used separation technique for peptide fractionation. In an acidic condition (pH ~2.7), most peptides are positively charged and they are retained on the SCX column by ionic interactions. Since SCX separates peptides according to positive charges on peptides, it has been used to selectively enrich for certain types of peptides[13-14]. For example, SCX is an enrichment technique for phosphopeptides, as the phosphorylation modification makes peptides less positively charged, thus retaining less on SCX column. Other less

commonly used peptides separation techniques include hydrophilic interaction chromatography (HILIC)[15], strong anion exchange (SAX)[16] and electrostatic repulsion-hydrophilic interaction chromatography (ERPLC)[17].

In case of complex peptide mixtures, such as whole cell lysates, the capability of a single separation method may not be sufficient to perform comprehensive proteome analysis. To this end, multidimensional separation techniques have been developed[18-19]. There are two criteria for setting up a multidimensional separation scheme. First, the separation methods should be orthogonal on their separation mechanism; second, effective interfacing between different separation dimensions should be in place, thus the separation power of the earlier dimension won't be lost when transferring to another separation dimension. At present, most of the shotgun proteome analysis methods are based on two-dimensional (2D) LC, in which SCX is the 1st separation dimension and RPLC serves as the 2nd dimension.

1.2 MS Instrumentation

1.2.1 Ionization

Before peptides are introduced into a mass spectrometer and analyzed, they need to become ionized. Electrospray ionization (ESI) is one of the major ionization methods in proteomic analysis (Figure 1.1). Electrospray ionization begins when a solution is sprayed through a narrow bore steel or conducting capillary under a high voltage, typically between 2 to 3 kV. Under the influence of the high electric field at the capillary tip, a Taylor cone is formed at a threshold voltage and the solution is sprayed to form an aerosol of charged droplets. In the heated source region, the solvent of the charged droplets evaporates until it has attained a sufficient charge density to allow sample ions to be ejected from the surface of the droplet[20]. There are two main theoretical models rationalizing the formation of gas phase analyte ions: the charged residue model[21] and the ion evaporation model.[22]

In the charged residue model, with solvent evaporation, smaller fission droplets

are formed from the main droplet once the charge to volume rate of the main droplet exceeds the Rayleigh limit. Compared to the main droplet, the smaller fission products have higher charge to volume rates. If solution is continuing evaporating, the smaller fission droplet can keep producing even smaller fission droplets, till the production of single charged droplets. If there is an analyte molecule within this single charged droplet, continuing desolvation would eventually lead to charge transfer to the analyte and formation of the gas phase ion. Alternatively, the ion evaporation model suggests that gas phase ions are formed directly from the droplets, without the production of the smaller fission products. Depending on the nature of an analyte, either model can be applied to explain the ionization process. For large molecules, like peptides and proteins, the charged residue model is considered to be a better approximation for the formation of gas phase ions, whereas the ion evaporation model is suggested to be more accurate in modeling ion formation for smaller analytes.

1.2.2 MS Analyzer

After ionization, ions in gas phase are transported to a mass analyzer and are separated according to their m/z values. There are a variety of mass analyzers in different configurations available for proteomic analysis. The main instruments used for my thesis work were quadrupole time-of-flight instruments (QTOFs). It combines a quadrupole mass analyzer and a time-of-flight tube to form a tandem mass spectrometer. Figure 1.2 displays a generalized schematic of a QTOF premier system from Waters. Briefly, the system consists of an ESI source, a quadrupole unit, a collision cell and an orthogonal acceleration TOF tube.

The quadrupole is composed of four cylindrical metal rods that are arranged in a square or near-square configuration. The diametrically opposed rods are paired; the same (DC) voltage and a radiofrequency (RF) voltage are applied for each paired rods to form an oscillating electric field. At a giving ratio of DC and RF voltages, ions of a particular *m/z* ratio will be able transit down the entire length of the rods; other

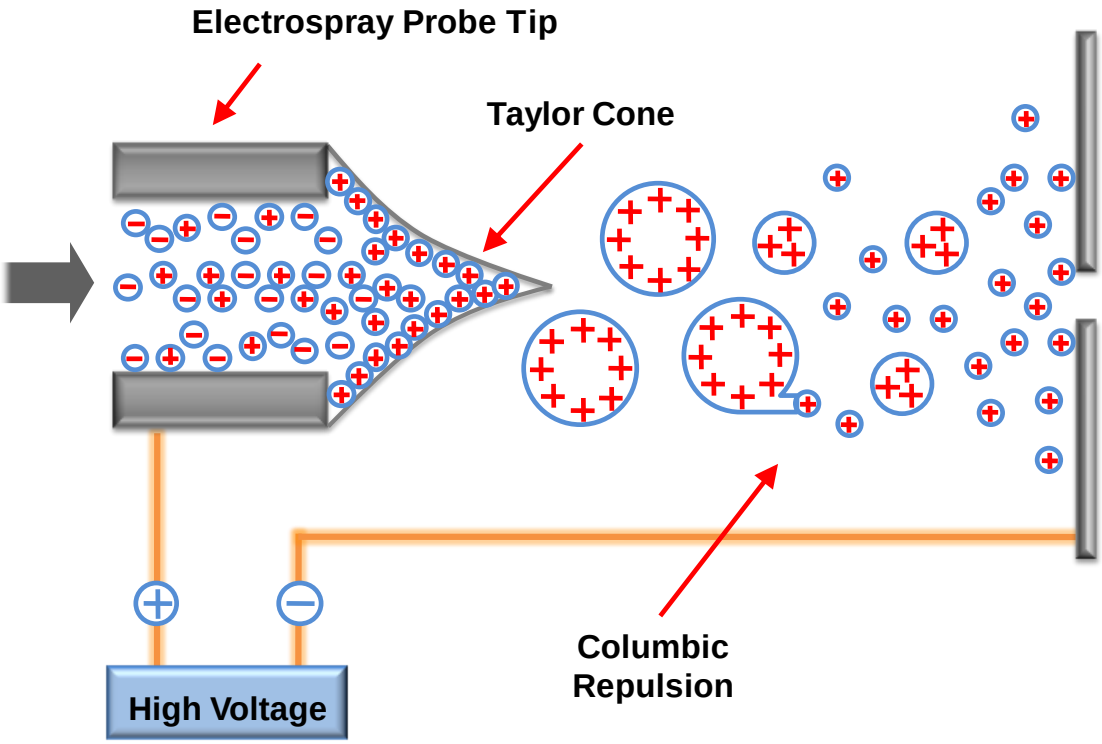


Figure 1.1. Process of electrospray ionization (ESI).

ions have unstable traveling trajectories and are lost. The transmitted ions enter the TOF tube.

In the TOF analyzer, it's a field free region; no electric or magnetic fields are applied across the length of the flight tube. The ions are pulsed in by an electric field applied on the top of the flight tube. The velocity of an accelerated ion is:

$$v = (\frac{2qU}{m})^{1/2}$$

Where U is the voltage, q is the charge of the ion and m is the mass of the ion. Its flight time in the TOF is:

$$t = \frac{d}{v}$$

Where d is the length of ion path. Since U and d are constant in a given flight tube with an electric field of known strength, the ion velocity or its flight time is determined by the ion's m/z ratio only.

Normally, in order to increase the mass analyzer's resolution, a V-mode reflectron TOF is used. The reflectron can partially compensate for initial energy dispersion of ions and focus ions having the same m/z value to the detector by using an ion reflector. The ion reflector consists of successive sets of plates, within which an electric field gradient is created. As ions with different kinetic energy enter this field, higher energy ions will penetrate deeper into the reflectron, increasing their flight path length and observed flight time. Compared with a linear TOF analyzer, the reflectron TOF increases mass resolution (m/Δm = ~10,000 to 20,000), with minimal losses in sensitivity.

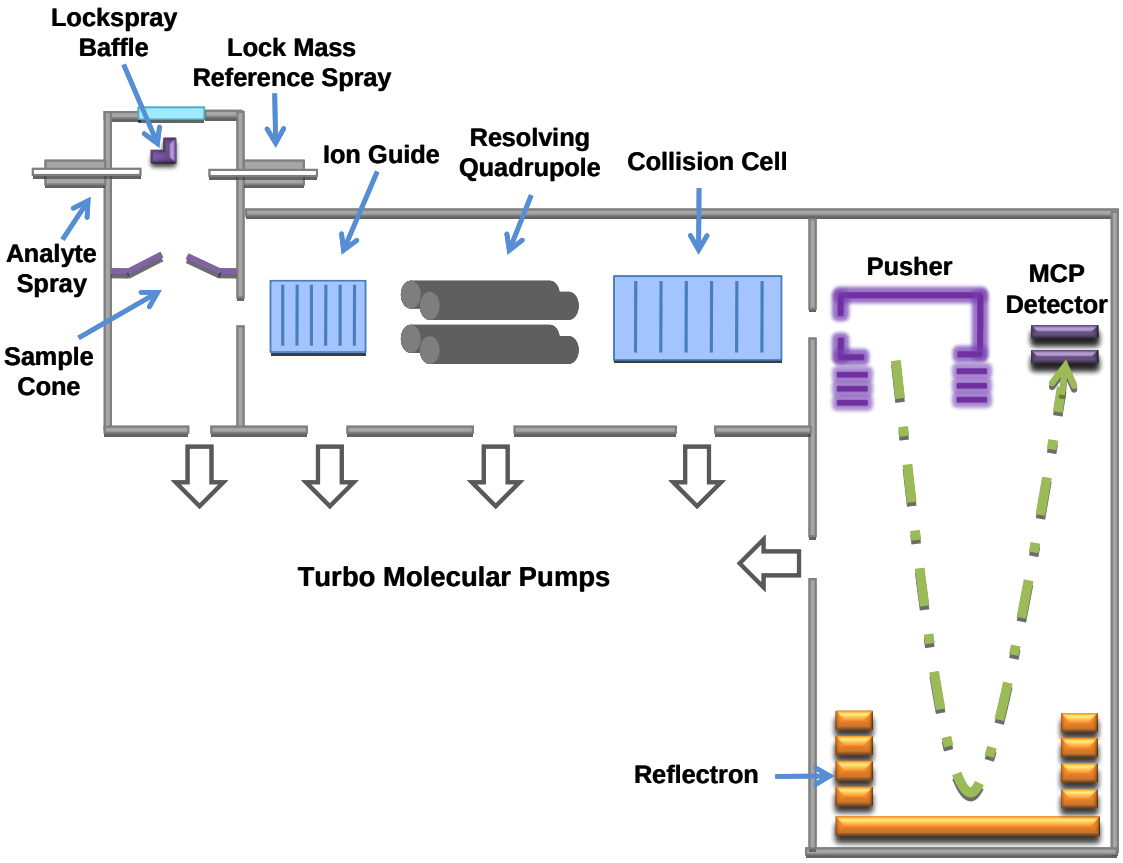


Figure 1.2. Schematic diagram of ESI QTOF MS from Waters.

1.2.3 Tandem MS

In a tandem mass spectrometer, collision-induced dissociation (CID) is one of the most popular methods for peptide ion fragmentation[23]. The collision cell is a hexapole or octopole within a set of acceleration plates. In CID, A small voltage is applied across the plates, and the molecular or precursor ions transmitted into the cell are accelerated by the applied electrical field to high kinetic energy in the vacuum. Then the ions collide with a neutral, inert bath gas, typically nitrogen or argon, to increase the internal energy within the ions. The collision leads to intramolecular energy distributions that result in bond fragmentation. Although there are many bonds within a peptide, the most common fragmentations occur along the amide backbone of the peptide. The most commonly observed ion types are b- and y- ions resulting from cleavage of the C-N bond in the amide backbone as shown in Figure 1.3. Other fragment ion series are less common and may be enhanced in higher energy fragmentation methods. Neutral losses of water (-18 Da) can be observed for serine, threonine, aspartic acid, and glutamic acid containing fragment ions and ammonia (-17 Da) can be observed for lysine, arginine, asparagine, and glutamine containing fragments, but these are generally weaker than the parent fragment.

A typical duty cycle during tandem mass spectrometric analysis of peptides is initiated with by an acquisition of an MS spectrum. The quadrupole is set as a broad bandpass filter, allowing ions over the entire *m/z* range (typically *m/z* 300 to 2000 for ESI-based instruments) through to be measured by the TOF mass spectrometer. Peaks are then quickly processed and the most intense peaks are selected for MS/MS fragmentation. To collect an MS/MS spectrum, the quadrupole is set to allow only ions with a particular *m/z* ratio through. Precursor ions are fragmented in the collision cell and the product ions are separated in the TOF analyzer and measured at the detector. The timescale for each MS and MS/MS spectrum is typically on the order of one second in our study. MS/MS spectral acquisition begins when ions selected by the quadrupole enter the collision cell to be fragmented.

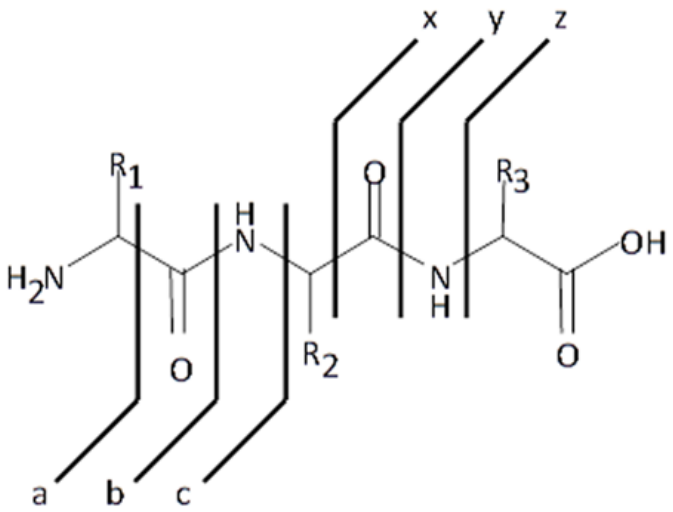


Figure 1.3. CID fragmentation pattern of a peptide ion.

1.3 Overview of MS-Based Quantification Proteomic Analysis

As proteins have great variations in their primary structures, each protein will produce peptides of variable length, amino acid composition, and amino acid sequence. Peptides with different structures will have different ionizing efficiency in MS. Furthermore, due to chemical interferences, absolute peak intensities measured by mass spectrometry do not necessarily indicate peptide concentrations. In order to address this issue, a variety of proteomic quantitative strategies for mass spectrometry have been developed. Based on data generated from these quantification strategies, they can be grouped as relative and absolute quantification.

In relative quantification, also called comparative quantification, usually two or more proteome samples were compared to get concentration ratios of individual proteins. It provides shifting patterns in protein expression between samples. However, in absolute quantification, a protein sample is usually compared with known-concentration internal standards. By comparing the peak intensity of analytes with internal standards in mass spectrometry, the absolute amount of analytes in a protein sample can be calculated. In the following discussion, strategies of these two types of quantification will be reviewed.

1.4 Relative Quantification

1.4.1 Label Free

Although signal response on mass spectrometry may not be directly related to peptide concentration, there is an increase in observed signal response as peptide concentration increases[24]. Based on this observation, a series of semi-quantification strategies, label free, are developed for quick sample evaluation.

A widely applied label free strategy is spectral counting. This strategy exploits the inherent MS and MS/MS duty cycle and uses the frequency of MS/MS

sequencing events as a surrogate for protein concentration[25-26]. In one duty cycle of peptide identification, an initial MS or survey scan identifies the precursor ions. For subsequent MS/MS sequencing, precursor ions with proper MS intensity are selected, transmitted into the collision cell, fragmented and the fragment ions are scanned for generating a MS/MS spectrum which contains peptide structure information. After a set of precursor ion peaks are sequentially analyzed by MS/MS, the mass spectrometer will begin the duty cycle again and acquire another MS spectrum to select more peaks for further analysis. In spectral counting strategy, the number of MS/MS scan for a given protein's peptides is considered to be correspondent with the overall concentration. With an increase of a protein's concentration, more of its constituent peptides will be selected from MS and redundantly analyzed by MS/MS.

Another method for label free analysis compares the ion current intensity in an extracted ion chromatogram of a peptide from different samples[27]. In this strategy, samples to be compared are prepared under identical conditions and run under the same LC-MS conditions. One of the samples is set as reference and its MS/MS peak lists are processed and searched against the appropriate database to identify the peptides. Utilizing a peptide's precursor m/z value and its chromatographic retention time, the elution profile of each identified peptide is reconstructed. Peptides are matched in comparison samples with their accurate m/z value and retention time. The reconstructed ion intensities are then integrated over the elution profiles for each peptide and the relative response is reported.

The direct comparison of peptide signals across multiple runs is deceptively simple: by comparing the extracted ion chromatograms of each component, the relative intensity of the same component in different samples can easily be gauged. However, feature alignment and correction are still an active area of research, aiming to continually refining current algorithms[28]. Because of the sample complexity and multiple charge states for each peptide, there can be many similar peaks within a reasonable m/z ratio tolerance and retention time range.

Another drawback of ion current based methods is that variations in LC or MS instrument performance could significantly affect the accuracy of peptide matching in comparison samples[29-30]. One corrective approach spikes peptides standards into samples in order to aid in chromatographic retention time alignment.[31] Although many algorithms and corrective approaches may be applied, data generated from label free quantification methods are still recognized as semi-quantitative, as they are less accurate than label based quantification.

1.4.2 Label Based

Label based quantification methods are characterized by incorporating stable isotopic standards to analytes. The stable isotopic standards are usually isotopomers of analytes. They have nearly identical chemical behaviour, but different m/z value. Depending on when isotope standards are introduced, analytes and standards go through the same sample preparation procedure, chromatographic separation, and MS analysis, that minimizes sample processing variations. Based on ways to introduce isotopes, label based quantification methods can be generally divided into two categories.

1.4.2.1 Metabolic Methods

Proteome in cells can be metabolically coded with isotopically labeled small molecules and inorganic salts by growing cells on isotopically enriched sources for quantitative proteomics experiments. Unlike other quantitation methods, metabolically labeled samples can be combined at the early sample preparation stage and subjected to protein level separation, such as SDS-PAGE and affinity chromatography, like phosphoprotein enrichment[32] or immunoprecipitation[33]. Early sample mixing ultimately reduces experimental variation. However, metabolic labeling methods have one major limitation: samples need to be cultured in a specific medium. Many clinically significant samples, such as human blood and plasma, cannot be metabolically labeled as they cannot be cultured in medium.

Besides that, subtle effects on cell proteome when cells are growing on isotopically enriched media may also need to be carefully evaluated. Additionally, metabolic labeling methods tend to be relatively expensive. Despite these shortcomings, metabolic labeling strategies are widely applied due to their many advantages. One example of metabolic labeling is stable isotope labeling by amino acids in cell culture (SILAC).

Stable isotope labeling by amino acids in cell culture (SILAC) was first reported in 2002, using d_{10}-leucine to label yeast proteome[34]. Modern versions of SILAC usually use a combination of ^{13}C and/or ^{15}N enriched lysine or arginine in the growth medium of bacteria, yeast, and human cells[35]. Lysine and arginine are used because they are trypsin recognition sites in protein digestion. The incorporation of a single isotopic label specifically at the C-terminus of peptides after tryptic digestion facilitates following data analysis.

1.4.2.2 Chemical and Enzymatic Labeling Methods

With the chemical and enzymatic labeling methods, intact proteins or peptides are labeled with isotopic reagents, ensuring equal reactivity and identical performance in downstream sample preparation.

The primary advantage of chemical labeling approaches is their general applicability to all types of samples, regardless of their origin. Chemical labeling approaches can be divided into two classes: non-isobaric and isobaric tags. Non-isobaric mass tags extract quantitation information from MS spectra, while isobaric tags obtain quantitation data from MS/MS scans. Peptides reacted with non-isobaric tags will have certain mass shift in an MS spectrum depending on the number of labels. One typical example of non-isobaric tags is dimethylation on peptides or proteins free amine groups with isotopic coded formaldehyde, ^{12}C-/^{13}C- or H/D. The labeling gives peptides 2 to 6 Da mass shift in MS spectra. Differently, isobaric tags labeled peptides have the same apparent mass in MS; quantitation

results are extracted from MS/MS after peptide ion fragmentation. Isobaric reagents have similar structures, but different patterns of isotope incorporation. After fragmentation, isobaric reagents yield fragment ions of different isotopic composition, which are named reporters or report ions. The observed ratio between reporters in MS/MS gives quantification information on comparison samples. Several isobaric labeling reagents are commercially available, like iTRAQ and TMT. The generalized structure of isobaric tags has three features: the reactive group, the mass balance group, and the reporter group. The reactive group usually reacts with amine groups in peptides, like N-terminal amine and amine on the side chain of lysine. The balance and reporter groups have isotopes (^{13}C, ^{15}N, and ^{18}O) incorporated at various atoms in order to maintain the same mass of the modifying group. However, the pattern of isotopic incorporation in the reporter group produces a reporter ion in CID with different m/z value. Current commercially available isobaric reagents are capable to simultaneously compare of six to eight samples. Both methods provide their own advantages, features, and drawbacks and both of them can be optimized based on the objective of a study.

Enzymatic methods can also been used to introduce stable isotopic codes into proteins or peptides. The most widely used method is tryptic digestion in ^{18}O water[36-37]. Samples are separately digested with trypsin in either $H_2{}^{16}O$ or $H_2{}^{18}O$. In addition to digestion, trypsin exchanges the ^{16}O atoms at the C-terminus of peptides with ^{18}O. The samples are then mixed and subjected to downstream processing. The main issue with enzymatic labeling is the potential back exchange. If trypsin remains active in solution[38] after samples are mixed, they could slowly converse the ^{18}O atoms back to ^{16}O under mildly acidic aqueous solutions[39].

1.5 Absolute Quantification

1.5.1 Target Protein Quantification (MRM)

Multiple reaction monitoring (MRM) is one strategy for absolute protein

quantification, as this method can be used to selectively measure ion intensities of known peptides from given proteins with improved quantification accuracy and precision. In MRM experiments, triple quadrupole (QQQ) mass spectrometers are widely used. Within the retention time window of a peptide, the first quadrupole is set to allow the precursor ions with a certain m/z value to pass through and the second quadrupole is set to monitor a certain fragment ion without scanning the entire mass range. This strategy focuses an instrument scan on a limited mass range and significantly increases a signal response, since only particular transitions of interest are measured. MRM timetables are designed to schedule selected transitions during the LC separation gradient in order to maximize the number of peptides measured in a single chromatographic run. By MRM, the sensitivity could be improved by at least 10-fold for monitoring known peptides. However the main drawback of MRM strategy is that it is only applicable to previously identified peptides.

1.5.2 Synthesized Standards

Nowadays, isotopic peptide standards can be synthesized using solid phase peptide synthesis methods and spiked into samples for quantitative analysis. This absolute quantification method is known as AQUA and was first published in 2003[40]. By comparing the signal responses of the analytes and synthetic peptides, absolute quantification information can be obtained if the amount of synthetic peptide added is known.

Another absolute quantification method, named QconCAT, was described in 2006[41-42]. In this approach, two or three peptides are selected from a given protein and monitored by MS. A chimeric protein standard that contains the selected peptides is expressed in cells by recombinant DNA techniques. By growing the expression vector transfected cell in isotopically labeled media, the produced chimeric protein standard is isotopically coded. After purification and quantification, the isotopically labeled chimeric protein standard can be spiked into protein samples

as an internal standard, and go through downstream sample preparation and analysis process. In a mass spectrometer, the isotopic peptides produced from the chimeric protein can be used as references for absolute quantification. The use of two or more peptide standards provides better accuracy and precision for quantifying a protein of interest, as the average value of the quantified signals can be used.

1.6 Scope of The Thesis Work

My thesis work focuses on analytical method development and applications of quantitative mass spectrometry for proteome and protein analysis. In Chapter 2, I will describe my work related to determining the total protein or peptide amount for assessing sample integrity in a proteome sample preparation workflow. In Chapter 3, I will describe a work on evaluating the effects on detergents on an isotope labeling technique, dimethylation after guanidination (2-MEGA), used for shotgun quantitative proteome analysis. In Chapters 4 and 5, I will focus on demonstrating the applications of the 2-MEGA labeling method in MS-based biomarker identification for breast cancer and deep tissue injury. In Chapter 6, I will discuss an absolute quantification method developed for analyzing a mixture of N-truncated protein and intact protein. The thesis ends with a conclusion chapter (Chapter 7) where I also briefly comment on the future work related to my research.

1.7 Literature Cited

1. Dworzanski, J. P.; Snyder, A. P., Classification and identification of bacteria using mass spectrometry-based proteomics. *Expert Rev Proteomics* **2005,** *2* (6), 863-78.

2. Leitner, A.; Lindner, W., Chemistry meets proteomics: the use of chemical tagging reactions for MS-based proteomics. *Proteomics* **2006,** *6* (20), 5418-34.

3. Sprenger, R. R.; Horrevoets, A. J., The ins and outs of lipid domain proteomics. *Proteomics* **2007,** *7* (16), 2895-903.

4. Wu, L.; Han, D. K., Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Rev Proteomics* **2006,** *3* (6), 611-9.

5. Isserlin, R.; Emili, A., Interpretation of large-scale quantitative shotgun proteomic profiles for biomarker discovery. *Curr Opin Mol Ther* **2008,** *10* (3), 231-42.

6. Kito, K.; Ito, T., Mass spectrometry-based approaches toward absolute quantitative proteomics. *Curr Genomics* **2008,** *9* (4), 263-74.

7. Smith, J. C.; Figeys, D., Recent developments in mass spectrometry-based quantitative phosphoproteomics. *Biochem Cell Biol* **2008,** *86* (2), 137-48.

8. Simpson, D. M.; Beynon, R. J., Acetone precipitation of proteins and the modification of peptides. *J Proteome Res* **2010,** *9* (1), 444-50.

9. McDonald, C. A.; Yang, J. Y.; Marathe, V.; Yen, T. Y.; Macher, B. A., Combining results from lectin affinity chromatography and glycocapture approaches substantially improves the coverage of the glycoproteome. *Mol Cell Proteomics* **2009,** *8* (2), 287-301.

10. Rodriguez, J.; Gupta, N.; Smith, R. D.; Pevzner, P. A., Does trypsin cut before proline? *Journal of Proteome Research* **2008,** *7* (1), 300-305.

11. Zhong, H.; Marcus, S. L.; Li, L., Microwave-assisted acid hydrolysis of proteins combined with liquid chromatography MALDI MS/MS for protein identification. *J. Am. Soc. Mass Spectrom.* **2005,** *16* (4), 471-481.

12. Slebos, R. J. C.; Brock, J. W. C.; Winters, N. F.; Stuart, S. R.; Martinez, M. A.; Li, M.; Chambers, M. C.; Zimmerman, L. J.; Ham, A. J.; Tabb, D. L.; Liebler, D. C., Evaluation of Strong Cation Exchange versus Isoelectric Focusing of Peptides for Multidimensional Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Proteome Research* **2008,** *7* (12), 5286-5294.

13. Helbig, A. O.; Gauci, S.; Raijmakers, R.; van Breukelen, B.; Slijper, M.; Mohammed, S.; Heck, A. J., Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome. *Mol Cell Proteomics* **2010,** *9* (5), 928-39.

14. Motoyama, A.; Xu, T.; Ruse, C. I.; Wohlschlegel, J. A.; Yates, J. R., 3rd, Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of

peptides and phosphopeptides. *Anal Chem* **2007,** *79* (10), 3623-34.

15. Mitchell, C. R.; Bao, Y.; Benz, N. J.; Zhang, S., Comparison of the sensitivity of evaporative universal detectors and LC/MS in the HILIC and the reversed-phase HPLC modes. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009,** *877* (32), 4133-9.

16. Takahashi, N.; Takahashi, Y.; Heiny, M. E.; Putnam, F. W., Purification of hemopexin and its domain fragments by affinity chromatography and high-performance liquid chromatography. *J Chromatogr* **1985,** *326*, 373-85.

17. Alpert, A. J., Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal Chem* **2008,** *80* (1), 62-76.

18. Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P., Multidimensional separations-based shotgun proteomics. *Chem Rev* **2007,** *107* (8), 3654-86.

19. Tomas, R.; Kleparnik, K.; Foret, F., Multidimensional liquid phase separations for mass spectrometry. *J Sep Sci* **2008,** *31* (11), 1964-79.

20. Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **2000,** *35* (7), 804-17.

21. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., Molecular beams of macroions. *J. Chem. Phys.* **1968,** *49* (5), 2240-9.

22. Iribarne, J. V.; Thomson, B. A., On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **1976,** *64* (6), 2287-94.

23. Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G., Mass spectrometry of peptides and proteins. *Methods* **2005,** *35* (3), 211-22.

24. Liu, H.; Sadygov, R. G.; Yates, J. R., III, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004,** *76* (14), 4193-4201.

25. Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M., Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **2005,** *4* (9), 1265-1272.

26. Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M., Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **2007,** *25* (1), 117-124.

27. Ryu, S.; Gallis, B.; Goo, Y. A.; Shaffer, S. A.; Radulovic, D.; Goodlett, D. R., Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method. *Cancer Inf.* **2008,** *4*, 243-255.

28. Savitski, M. M.; Fischer, F.; Mathieson, T.; Sweetman, G.; Lang, M.; Bantscheff, M., Targeted Data Acquisition for Improved Reproducibility and Robustness of Proteomic Mass Spectrometry Assays. *J. Am. Soc. Mass Spectrom.* **2010,** *21* (10), 1668-1679.

29. Paulovich, A. G.; Billheimer, D.; Ham, A.-J. L.; Vega-Montoto, L.; Rudnick, P. A.; Tabb, D. L.; Wang, P.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Clauser, K. R.; Kinsinger, C. R.; Schilling, B.; Tegeler, T. J.; Variyath, A. M.; Wang, M.; Whiteaker, J. R.; Zimmerman, L. J.; Fenyo, D.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Mesri, M.; Neubert, T. A.; Regnier, F. E.; Rodriguez, H.; Spiegelman, C.; Stein, S. E.; Tempst, P.; Liebler, D. C., Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. *Mol. Cell. Proteomics* **2010,** *9* (2), 242-254.

30. Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C., Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Proteome Research* **2010,** *9* (2), 761-776.

31. Christin, C.; Hoefsloot, H. C. J.; Smilde, A. K.; Suits, F.; Bischoff, R.; Horvatovich, P. L., Time Alignment Algorithms Based on Selected Mass Traces for Complex LC-MS Data. *Journal of Proteome Research* **2010,** *9* (3), 1483-1495.

32. Hilger, M.; Bonaldi, T.; Gnad, F.; Mann, M., Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Mol.*

Cell. Proteomics **2009,** *8* (8), 1908-1920.

33. Zhang, G.; Fenyo, D.; Neubert, T. A., Evaluation of the Variation in Sample Preparation for Comparative Proteomics Using Stable Isotope Labeling by Amino Acids in Cell Culture. *Journal of Proteome Research* **2009,** *8* (3), 1285-1292.

34. Jiang, H.; English Ann, M., Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *Journal of Proteome Research* **2002,** *1* (4), 345-50.

35. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen Dan, B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002,** *1* (5), 376-86.

36. Heller, M.; Mattou, H.; Menzel, C.; Yao, X., Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J. Am. Soc. Mass Spectrom.* **2003,** *14* (7), 704-718.

37. Fenselau, C.; Yao, X., 18O2-Labeling in Quantitative Proteomic Strategies: A Status Report. *Journal of Proteome Research* **2009,** *8* (5), 2140-2143.

38. Petritis, B. O.; Qian, W.-J.; Camp, D. G., II; Smith, R. D., A Simple Procedure for Effective Quenching of Trypsin Activity and Prevention of 18O-Labeling Back-Exchange. *journal of proteome research* **2009,** *8* (5), 2157-2163.

39. Niles, R.; Witkowska, H. E.; Allen, S.; Hall, S. C.; Fisher, S. J.; Hardt, M., Acid-Catalyzed Oxygen-18 Labeling of Peptides. *Anal. Chem. (Washington, DC, U. S.)* **2009,** *81* (7), 2804-2809.

40. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (12), 6940-5.

41. Pratt, J. M.; Simpson, D. M.; Doherty, M. K.; Rivers, J.; Gaskell, S. J.; Beynon, R. J., Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* **2006,** *1* (2), 1029-43.

42. Rivers, J.; Simpson Deborah, M.; Robertson Duncan, H. L.; Gaskell Simon, J.;

Beynon Robert, J., Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol Cell Proteomics* **2007,** *6* (8), 1416-27.

Chapter 2 - Quantification of Total Peptide Amount By an Optimized LC-UV Method for Assessing Sample Integrity During Proteome Sample Preparation

2.1 Introduction

Mass spectrometry (MS) is a powerful tool for mapping the proteome from a proteome sample[1-2]. A large scale proteome profiling work is generally carried out using hundreds of micrograms or milligrams of starting materials. To produce this quantity of sample, millions or billions of cells are used. However, in some clinical or biological studies, it may be difficult to collect a large amount of samples, like circulating cancerous cells captured from a blood sample of a patient with early sign of tumor in a specific organ[3], and stem cells isolated from a large population of other cells, which usually only count for a rare portion of the total cell population[4]. These few-cell samples do not produce a large amount of proteins. In order to ensure maximum proteome coverage for these limited cell samples, both the sample preparation procedure and MS analysis method need to be optimized according to the sample amount.

Shotgun proteomics has emerged as a powerful approach for the analysis of complex protein mixtures, including biofluids, tissues, cells, organelles or protein complexes. It's relatively sensitive compared to other proteome analysis methods such as gel-based proteome analysis[5]. In a shotgun proteomic workflow (Figure 2.1), proteins are firstly extracted from a biological source, for example, cells or tissue. The extracted proteome is chemically or enzymatically digested, followed by peptide separation with multidimensional chromatographic techniques. The separated peptides are subsequently introduced into a mass spectrometer for MS and MS/MS analysis. Any one of these steps can potentially involve the loss of some proteins. In working with a large quantity of samples, this sample loss may not be very significant so long as the sample loss is not biased towards a particular group of proteins.

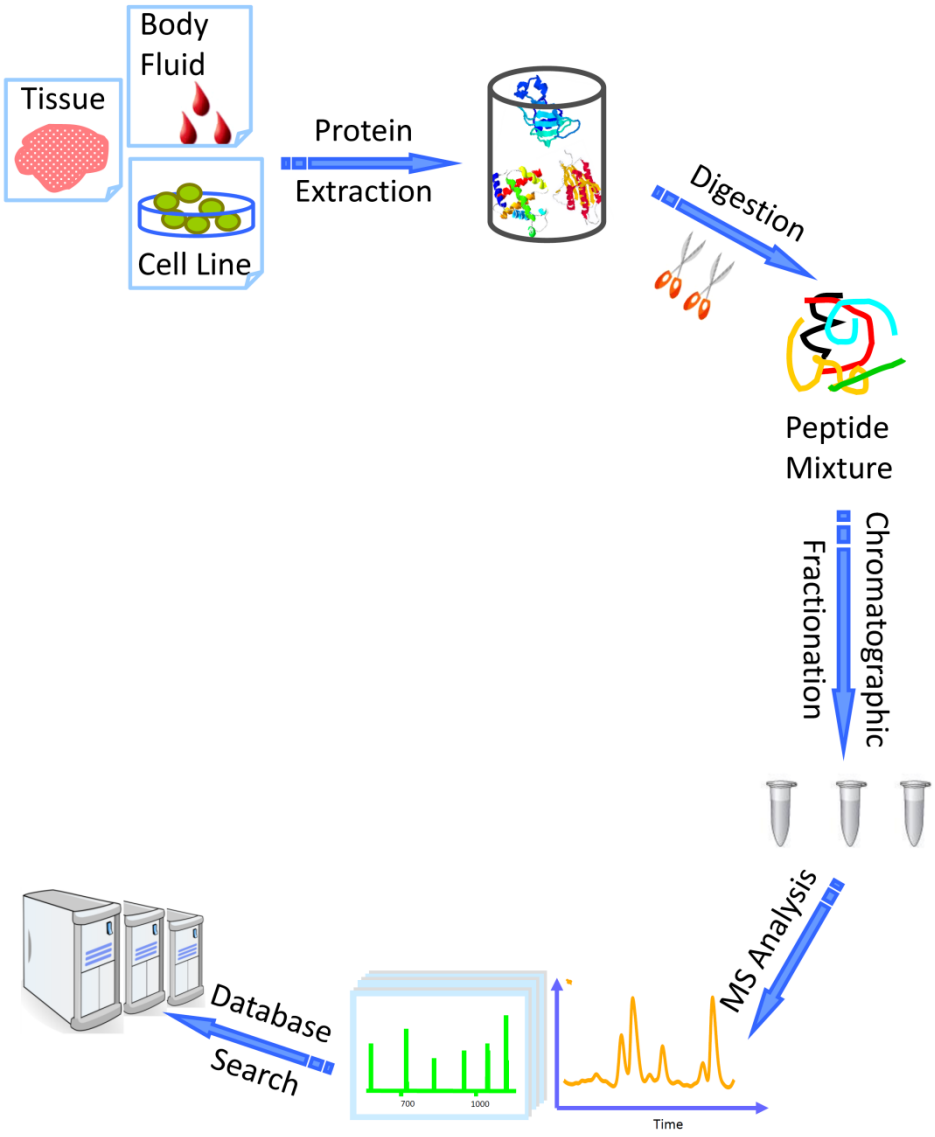


Figure 2.1. Workflow of a shotgun proteomic experiment.

However, in handling small numbers of cells, sample loss of any type can be detrimental to the proteome coverage. The reason is that the amount of sample generated from a small number of cells will be limited and it will often not meet the optimal sample amount required for peptide sequencing in LC-MS/MS (e.g., < 1 μg). Thus it's very important to evaluate any sample loss in preparation and optimize the LC-MS/MS conditions in order to detect a maximum number of peptides and proteins from a limited amount of sample.

In this work, we have developed a micro-LC/UV method with a step gradient elution to desalt and, at the same time, rapidly quantify the total peptide amount in proteomic digest samples. By coupling with LC-ESI MS/MS, the proteome of less than 1000 cells (250, 500 and 1000 MCF-7 cells) was analyzed in our shotgun proteomic experiment. Adequate coverage of the proteome from this number of cells may lead to several important applications. For example, 1000 cells may be collected from a patient blood containing rare circulating cancerous cells from an early stage of metastasis of a solid tumor[6-10]. The proteome of these cells may be used as a fingerprint for diagnosis or prognosis of a cancer. Another example is that about 1000 cells may be procured from a tissue section using laser capture microdissection (LCM) within a couple of hours. Analyzing these cells may assist in identifying specific protein markers for disease diagnosis. The ultimate goal of this research is to analyze single cell proteome[11]. Unfortunately, this is a huge challenge at this moment for mass spectrometry based technologies due to limited sensitivity. Developing and applying techniques for analyzing the proteome of a few hundred cells is a more realistic goal. However, very few studies of proteome analysis from a few hundreds of cells have been reported[12-16].

2.2 Experimental Section

2.2.1 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).

2.2.2 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 dithiothreitol (DTT) and alkylation with the same volume of 40 mM iodoacetamide (IAA), the protein mixture was then digested by trypsin at a final enzyme concentration of 8 ng/µl at 37 ^{o}C for 8 hours.

2.2.3 RPLC Calibration

RPLC columns (C18, particle size: 30 µm, pore size: 300 Å, length: 50 mm) with different inner diameters: 4.6-mm, 2-mm and 1-mm were calibrated with the four-protein mixture digest on an Agilent 1100 LC-UV system. The flow rates of RPLC columns with 4.6-mm, 2-mm and 1-mm inner diameters were 1 mL/min, 200 µL/min and 100 µL/min, respectively. Peptides were eluded from columns by 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. The UV absorbance of eluted peptides was detected at 214nm.

2.2.4 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 ºC in DMEM Gibco medium 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 ºC. The harvested cells were then fluorescently stained by incubating with a FITC–conjugated mouse anti–human HEA 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.

2.2.5 Cell Lysate and In-solution Digestion

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 DTT and alkylated with iodoacetamide. Acetone (precooled 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 uL) at 37 ^{o}C for 4 hr.

2.2.6 Peptide Desalting and Quantification by RPLC

The tryptic peptides were desalted and quantified on an Agilent 1100 HPLC system, with a 1 mm × 50 mm Polaris C18 A column (3 µm particle and 300 Å pore). The desalting and quantification step of tryptic peptides obtained from MCF-7 cells was following the above description of LC calibration with four-protein standard mixtures.

2.2.7 Mass Spectrometry Analysis of Small Number Cells

After desalting and quantification, the digests were analyzed using a QTOF Premier mass spectrometer equipped with a nanoACQUITY Ultra Performance LC system. Briefly, the desalted digests were concentrated on Speedvac and

reconstituted with 0.1% formic acid to 0.2 μg/μL. Then a specific amount of 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 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.

2.2.8 Protein Database Search

Raw LC-ESI data were lock-mass corrected, de-isotoped, and converted to peak list files by 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.

2.3 Results and Discussions

2.3.1 Calibration of RPLC Columns with Different Inner Diameters

Four-proteins mixture digests were linearly diluted and injected into the RPLC system as calibration standards. A step gradient, as described in experimental section, was used to rapidly elute peptides from a LC column to generate one integral peak.

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 at 214 nm. Fortunately, the system peak area was constant. Thus the peak area from the peptides could be determined by integrating the overall peak area from the UV chromatogram and subtracting out the system peak contribution. System peaks were reproducible when 4.6- and 2-mm columns were used, while using a 1-mm column required careful optimization of the flow rates and solvent conditions to achieve reproducible results. This is probably because the column with smaller inner diameter is more sensitive to factors like flow rate and solvent concentration. Figure 2.2 demonstrates the signal responses as a function of sample loading amounts on columns. In Figure 2.2-a, besides the sample signal peak eluted at 8.25 min, there is another small peak appeared at 9 min, which is also increased with increasing sample loading amount. This peak might be from the heme molecules released in the protein digests.

Columns with different inner diameters gave different linear ranges, as well as different slopes and intersections in their calibration curves (Table 2.1). The slope of calibration curve for each column was inversely proportional to its flow rate. The linear range for all columns was about two orders of magnitude. The relative standard deviation in signal response at different sample loading amounts for each column on the same day were <5%, while the day-to-day relative standard deviations for each column were <10%.

Table 2.1. Flow Rate, Calibration Equations and Linear Range of Columns with Different Inner Diameters.

Column ID	Flow Rate (mL/min)	Calibration Equation	R^2	Linear Range
4.6×50 mm	1	y = 416.28x - 91.625	0.9999	0.5 – 20 μg
2.0×50 mm	0.2	y = 1865x – 142.25	0.9983	0.3 – 7 μg
1.0×50 mm	0.1	y = 4519.6x – 48.35	0.9991	0.039 – 0.3 μg
		y = 8196.6x – 550.87	0.9958	0.6 – 5 μg

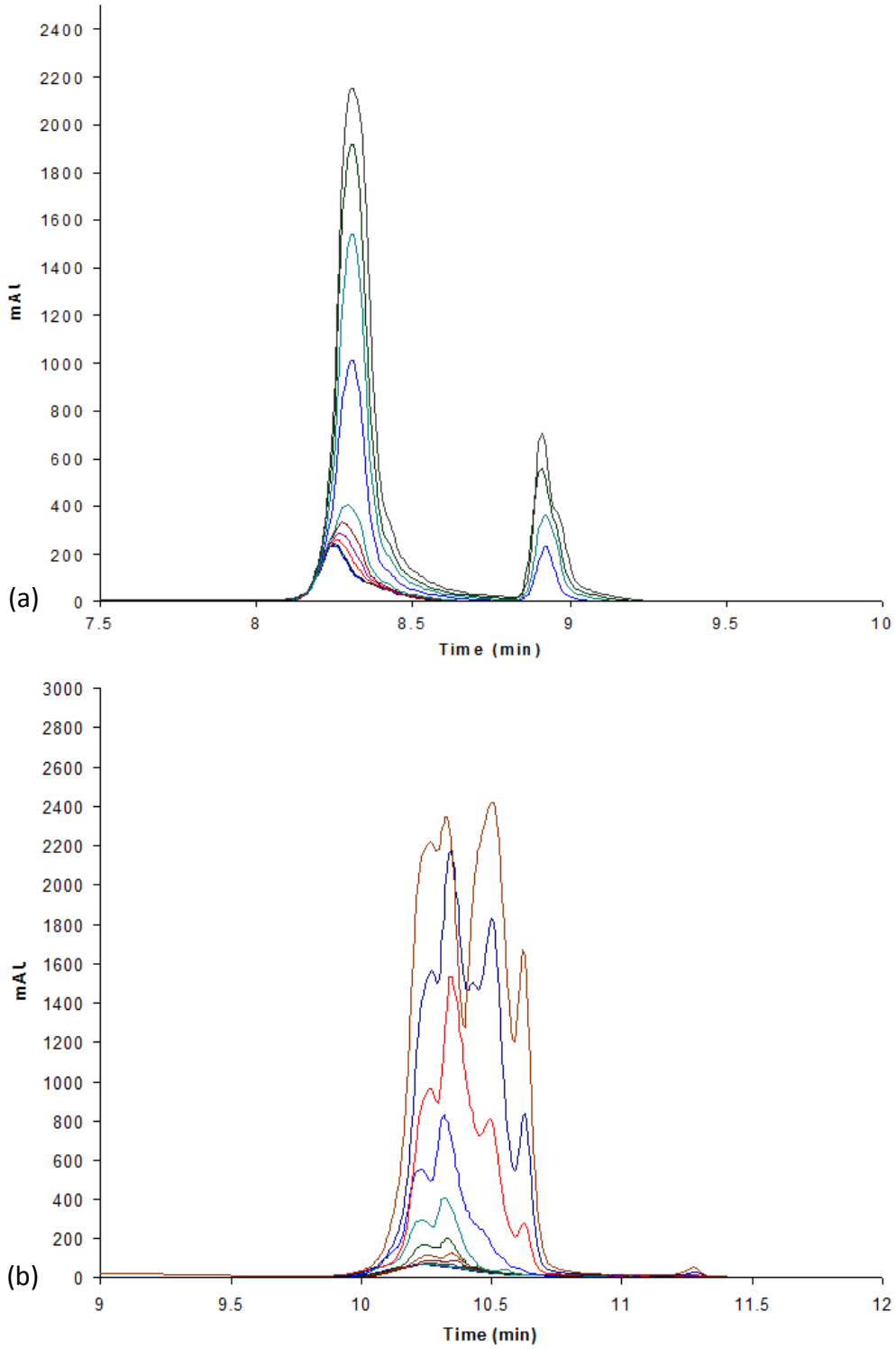


Figure 2.2. (a) LC-UV signals for different sample loading amounts on 2.0×50 mm C18 column, normal HPLC system. Detecting wavelength: 214nm. Sample loading amount: 0.01-10 µg. (b) LC-UV signals for different sample loading amounts on 1.0×50 mm C18 column, micro-bore HPLC system. Detecting wavelength: 214nm. Sample loading amount: 0.01-10 µg.

With a 1-mm column, the limit of quantification was <0.04 μg. From 1-mm column's calibration curve built by integrating peak area, we found its linear range was 39 ng – 0.3 μg and 0.6μg – 5 μg , because peptides' absorbance was not linear to the peptides amount in the range between 0.3 μg to 0.6 μg (Figure 2.3). By defining the linear ranges with different calibration equation, it's possible to calculate peptides amount in the ranges. This is not quite convenient in practical application, since many times, there might be no clue about the sample amount. We have examined if a derivative integration method could be used to build a better linear calibration curve.

Micro-bore LC was sensitive to system conditions as mentioned above. Derivative integration, which calculates and displays chromatograms by d(absorbance)/d(time) vs. time, could emphasize 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[17]. There are several methods to do integration on derivative chromatogram, for instance, peak height, maximum to minimum, etc. We used the height of the peak to do integration, as we found this peak is increasing proportionally to the sample amount (Figure 2.4). The derivative integration shows very good linearity from 39 ng to 5 μg. In the low sample loading amount range, from 39 ng to 1.25 μg, the calibration curve still shows good linearity and minor slope change, comparing with the whole range calibration equation (Figure 2.5).

However, we found that derivative integration was more sensitive to the effect of the peak shapes than normal integration. While it gives better linear range in calibration, when it comes to a real sample, it may have some problems as derivative integration requires constant derivative chromatogram, which is, however, very sensitive to changes in sample complexity and contaminants. If the sample components were not co-eluting, and showed a spectrum of splitting peaks, its derivative chromatogram would change. And the peak used for calibration may not be able to be used for integration. To give an example, complex protein samples, like

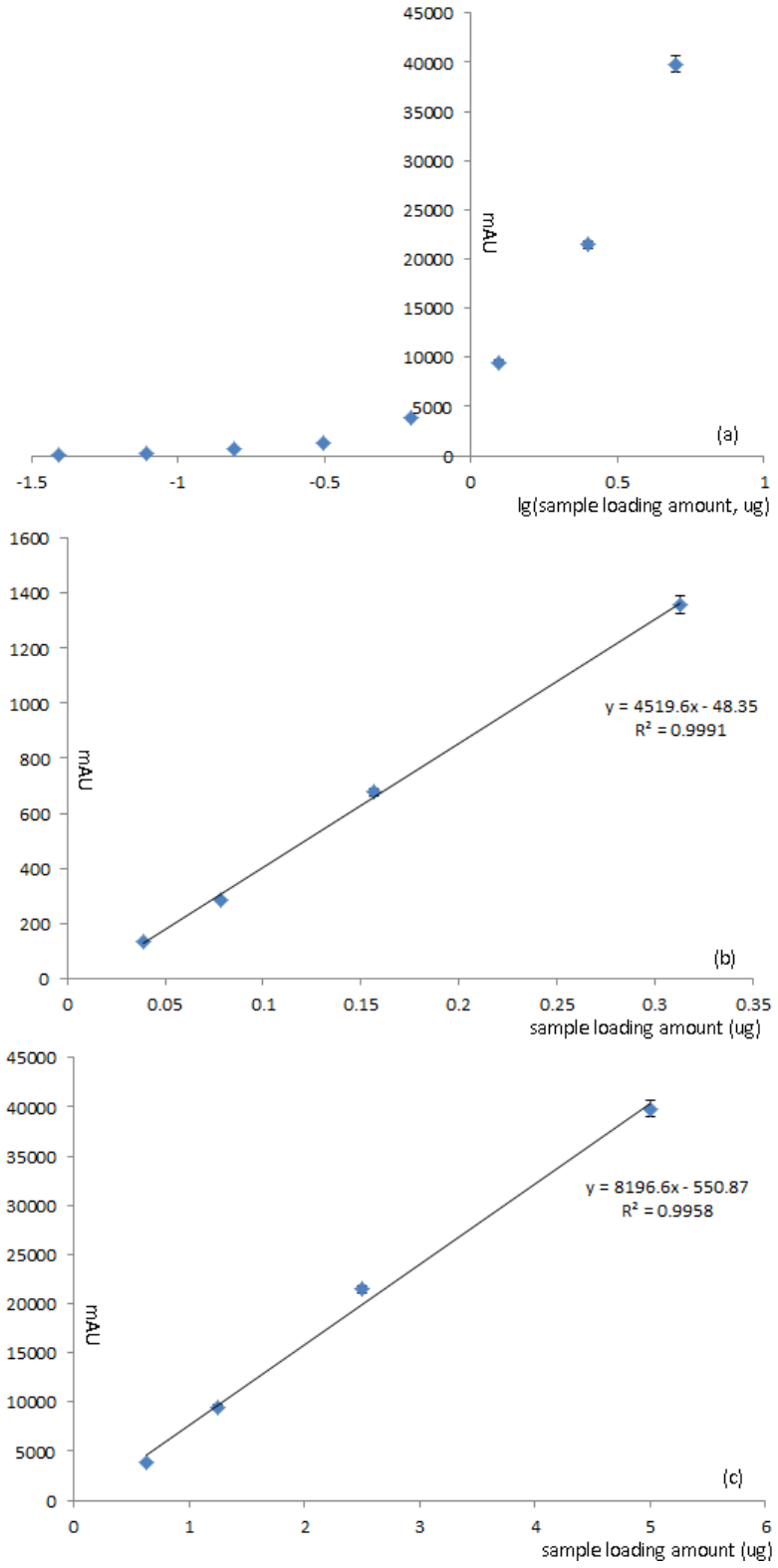


Figure 2.3. Dynamic range of 4-protein standards on 1.0×50 mm C18 column. (a) Calibration curve between signal responses vs. log(sample loading amount), 39 ng to 5 µg. (b) Linear dynamic range between 39 ng and 0.3 µg. (c) Linear dynamic range between 0.6 µg and 5 µg.

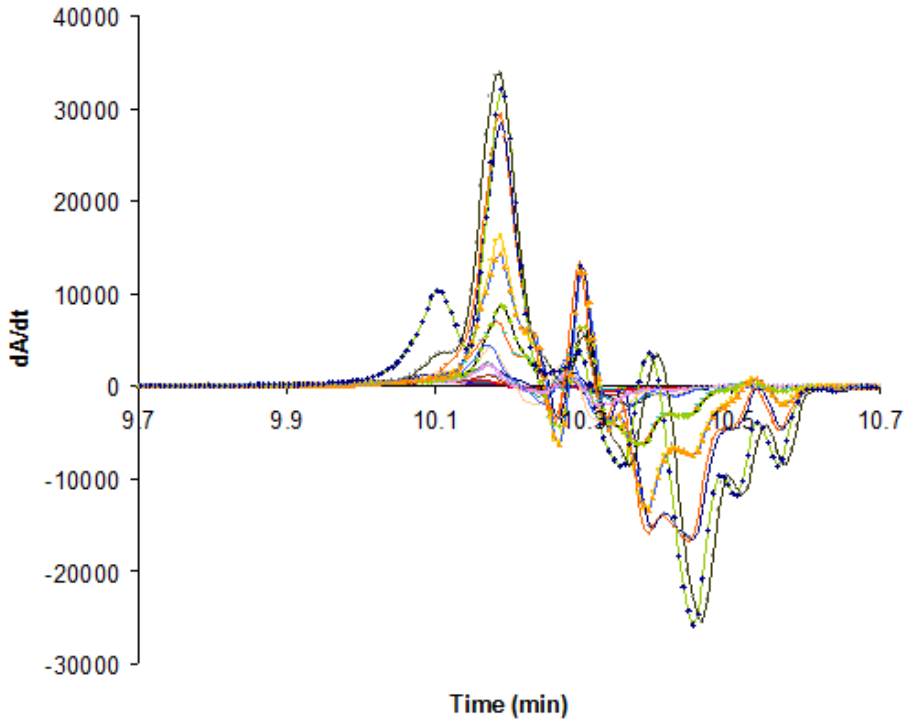


Figure 2.4. Derivative chromatograms of 4-protein standards on 1.0×50 mm C18 column. Sample loading amount: 0.01-10 µg. The peak intensities of the analyte peaks were proportionally related to the sample loading amount.

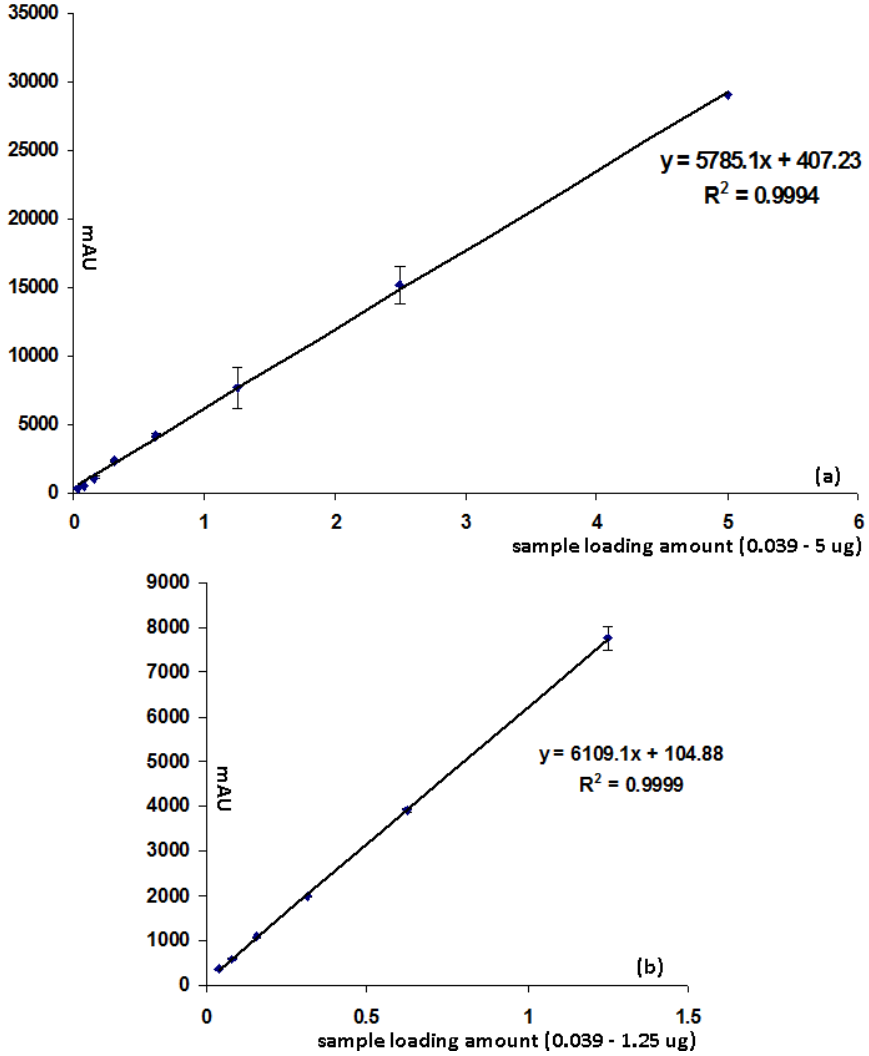


Figure 2.5. Calibration curve of 4-protein standards on 1.0×50 mm C18 column. Integrated by derivative chromatograms. (a) Linear calibration curve from 39 ng to 5 µg. (b) Linear calibration curve from 39 ng to 1.25 µg.

whole cell lysates, their derivative chromatograms were not similar with the calibration derivative chromatograms produced by the four-protein mixture standard (data not shown). Giving this consideration, it's difficult to apply the derivative integration method to a real sample. Therefore, in our later experiment of MCF-7 cells, we used peak area for integration on 1-mm column.

2.3.2 Cell Linear Dilution

In order to evaluate the influence of possible contaminants in MCF-7 cell lysates sample to UV absorbance and also to evaluate whether the calibration curves generated from the four-protein mixture standard could be applied for complex protein sample quantification, linear dilution experiment was performed. 5000 MCF-7 cells were lysed on ice, DTT reduced, IAA alkylated and trypsin digested. The digest was diluted at ½ rates to 2500, 1250, 625, 312.5, 156.25 and 78.125 cells. The linearly diluted cell digests were quantified on 1.0×50 mm C18 RPLC column (3 μm particles and 300 Å pore). The total peptide amounts of each sample were calculated based on the calibration curve built by the four-protein mixture digest (Table 2.2). Based on the calculation, the peptide amounts of 78 and 156 cells were lower than the limit of quantification (LOQ). So their amounts cannot be trusted, which was also proven by their RSD values obtained. For the samples supposed to have more than 312.5 cells, their RSDs were less than 10%, which was smaller than the instrument's inter- and intra-day deviation. These results may suggest that the deviations on the calculated peptide amounts from the UV absorbance measurements were from the LC system, not contaminants.

Table 2.2. Linear dilution experiment of MCF-7 cells. Peptide amounts were quantified on 1.0×50 mm C18 column, measured at 214 nm.

Calibration Equation: y = 4519.6x - 48.35

Diluted Cell Number	Absorbance	Peptide Amount (μg)	SD	RSD
78.125	46.82	0.021	0.0097	46.22%
156.25	50.54	0.022	0.0036	16.55%
312.5	258.12	0.068	0.0011	1.56%
625	763.69	0.18	0.0093	5.17%
1250	1729.42	0.39	0.0056	1.42%
2500	3641.60	0.82	0.064	7.78%
5000	8131.69	1.81	0.065	3.57%

To further prove this, we started with the peptide amount calculated from 5000 cells, and worked out the theoretical peptide amounts of other series of diluted cell samples (Table 2.3). The theoretical peptide amounts and the absorbance actually measured in the experiment were fit into a curve (Figure 2.6). Since the theoretical amounts did not count in any contaminants in samples, if the curve's linearity slope between the theoretical peptide amounts and the experimental absorbance greatly differed from the experimental calibration curve, it would indicate that contaminants in the complex protein sample had influence on the sample's UV absorbance and the calibration curve generated from the four-protein mixture standard cannot be applied to the real sample. However, Table 2.4 shows us that the RSD between slopes of these two curves was less than 1 %. To further evaluate the difference between the calibration curve constructed from 4-proteim standards and the fitting curve built with MCF-7 cells, both of these two quantification equations were applied to quantify the same MCF-7 cell dilution samples by their UV absorbance. The quantification results generated from both equations indicated few differences between these two methods (Figure 2.7). These results indicate that the influence of contaminants in complex cell lysate sample on UV absorbance were not significant, and the calibration curves built by the four-protein mixture standard could be applied to quantifying real proteome samples on RPLC.

Table 2.3. Theoretical peptide amounts of a series of diluted cell samples, calculated from 5000 MCF-7 cells.

Diluted Cell Number	Theoretic Peptide Amount (Calc. from 5000 cells, μg)	Absorbance
5000	1.81	8131.69
2500	0.90	3641.60
1250	0.45	1729.42
625	0.23	763.69
312.5	0.11	258.12
156.25	0.057	50.54
78.125	0.028	46.82

Table 2.4. Slopes of 4-protein standards calibration curve and 5000 MCF-7 cells dilution fitting curve.

	4-Protein Calibration Curve	5000 Cells Dilution Curve	SD	RSD
Slope	4519.6	4554.9	24.75	0.55%

2.3.3 Protein Identification of Few Cell Samples on MS

The amounts of peptides produced from cell lysates were determined using the LC-UV system as described in experimental section. The average amount (n=3) of peptides from the 1000-cell sample was found to be 0.610±0.004 μg. And the average amount of the 500-cell sample was 0.290±0.007 μg, which was very close to half of the amount of peptides produced from the 1000-cell sample. The lower limit of regular UV-LC system using 4.6-mm column to measure peptide amount is about 0.25 μg. By downsizing the inner diameter of RPLC column to 1-mm, and switching to a micro-bore UV-LC system, the LOQ of peptides was lowered to tens of nanograms. This lower LOQ facilitates the measurement of peptide amounts during the sample preparation of small amount samples.

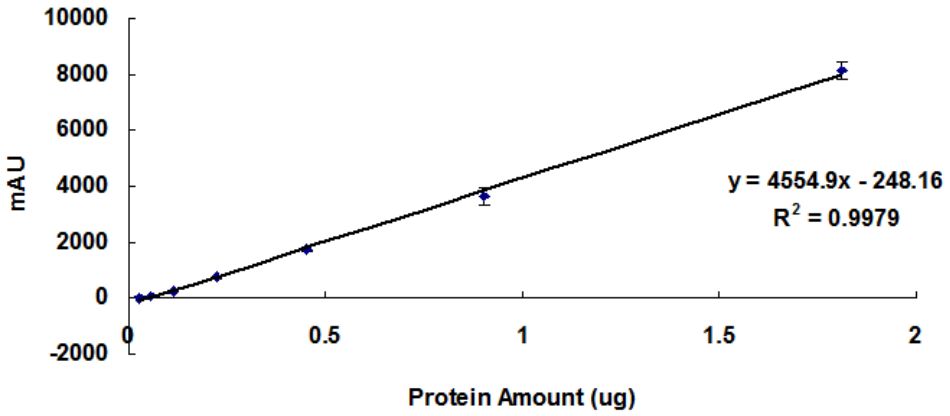


Figure 2.6. Fitting curve constructed by theoretically calculated peptide amounts from 5000 MCF-7 cells dilution experiment and observed LC-UV absorbance.

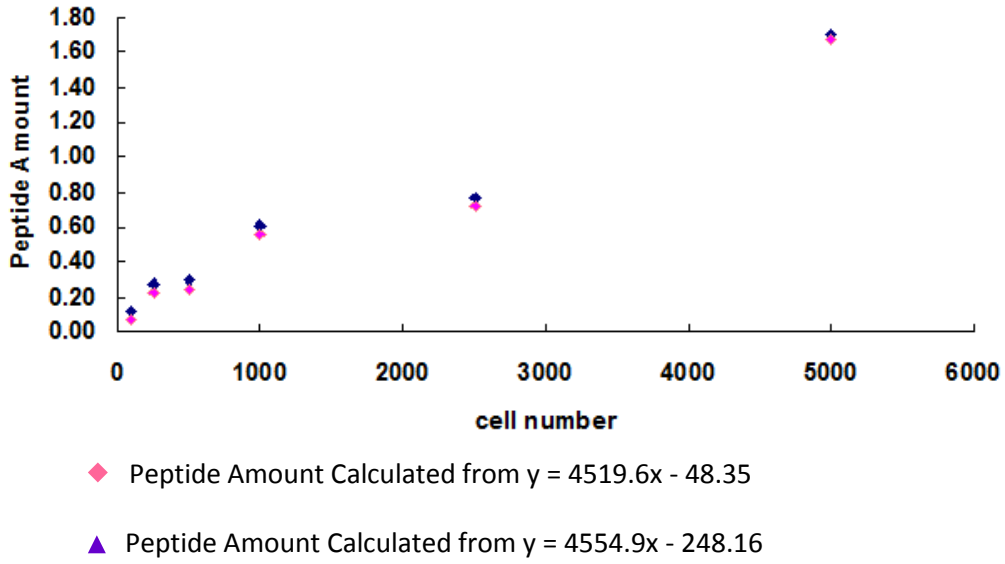


Figure 2.7. Peptide amounts of a series of diluted MCF-7 cell lysates calculated from two quantification equations.

Based on the quantification of peptide amount of 1000, 500 and 250 MCF-7 cell lysates, their separation and analysis on LC-ESI MS/MS were optimized. The gradient speed can significantly affect the detectability of peptides in LC-MS/MS. If a fast gradient is used, a peptide elutes quickly to form a fast rising peak in an ion chromatogram, resulting in intense signals in both MS and MS/MS spectra. But, in this case, only a few MS and MS/MS spectra can be acquired within the peak elution time. If a slow gradient is used, the same peptide would elute out more slowly to form a broader peak and the mass spectral signal of the peptide would be less intense. If an injected sample amount is limited, the gradient should be carefully optimized according to the sample amount to generate database-searchable MS/MS spectra, and at the same time, ensure enough separation power. The gradient speed's effects on the number of peptides identified by LC-ESI MS/MS had been investigated. Generally speaking, the optimum gradient time increased as the number of cells in a sample increased. 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 2.5. Identified peptide and protein numbers for 250, 500 and 1000 MCF-7 cells on LC-ESI MS/MS.

	250 MCF-7 cells		500 MCF-7 Cells		1000 MCF-7 Cells	
Protein I.D. Number	126	81	154	122	282	256
Unique Peptide I.D. Number	368	258	410	387	836	784

Table 2.5 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 numbers of peptides and proteins identified increased as the cell number increased and the number change was not quite in linear proportion to cell numbers, especially for fewer cell samples. For example, an average of 398±12 peptides or 138±16 proteins (n=2) were identified from the 500-cell sample, while 313±55 or 103±23 proteins were identified from the 250-cell sample. Although the cell number decreased by 2-fold, the number of peptides and proteins identified decreased by about 1.2- and 1.3-fold, respectively. However, the peptide/protein ratio kept around 3 for the 1000, 500 and 250 cell sample.

In all cases, the run-to-run reproducibility was good, indicating that the experimental protocol used in this study can be used to generate reproducible results from as few as 250 cells.

2.4 Conclusions

A micro-bore LC/UV method for quantifying peptides from a few-cell samples was developed. By using a micro-bore LC system and a small (1-mm) column, the LOQ of peptides was in the range of tens of nanograms. At the meantime, solvent consumption was significantly reduced, compared to the regular LC system using 2- or 4.6-mm column. Based on the results obtained from the quantification of sample amount, LC ESI-MS/MS gradient could be optimized. In analyzing MCF-7 cells, by coupling with an optimized LC ESI-MS/MS setup, we could identify an average of 103±23, 138±16, and 269±13 proteins from 250, 500, and 1000 cells, respectively. We envisage that this method will be useful in proteome profiling of small numbers of cells for disease diagnosis and prognosis.

2.5 Literature Cited

1. Haynes, P. A.; Yates, J. R., 3rd, Proteome profiling-pitfalls and progress. *Yeast* **2000,** *17* (2), 81-7.

2. Lee, K. H., Proteomics: a technology-driven and technology-limited discovery science. *Trends Biotechnol* **2001,** *19* (6), 217-22.
3. de Roos, B.; Duthie, S. J.; Polley, A. C.; Mulholland, F.; Bouwman, F. G.; Heim, C.; Rucklidge, G. J.; Johnson, I. T.; Mariman, E. C.; Daniel, H.; Elliott, R. M., Proteomic methodological recommendations for studies involving human plasma, platelets, and peripheral blood mononuclear cells. *J Proteome Res* **2008,** *7* (6), 2280-90.
4. Pardal, R.; Clarke, M. F.; Morrison, S. J., Applying the principles of stem-cell biology to cancer. *Nat Rev Cancer* **2003,** *3* (12), 895-902.
5. McDonald, W. H.; Yates, J. R., 3rd, Shotgun proteomics and biomarker discovery. *Dis Markers* **2002,** *18* (2), 99-105.
6. Chosy, E. J.; Nakamura, M.; Melnik, K.; Comella, K.; Lasky, L. C.; Zborowski, M.; Chalmers, J. J., Characterization of antibody binding to three cancer-related antigens using flow cytometry and cell tracking velocimetry. *Biotechnol Bioeng* **2003,** *82* (3), 340-51.
7. Weir, I. E.; Maddumage, R.; Allan, A. C.; Ferguson, I. B., Flow cytometric analysis of tracheary element differentiation in Zinnia elegans cells. *Cytometry A* **2005,** *68* (2), 81-91.
8. Utz, P. J., Protein arrays for studying blood cells and their secreted products. *Immunol Rev* **2005,** *204*, 264-82.
9. Schneider, T.; Moore, L. R.; Jing, Y.; Haam, S.; Williams, P. S.; Fleischman, A. J.; Roy, S.; Chalmers, J. J.; Zborowski, M., Continuous flow magnetic cell fractionation based on antigen expression level. *J Biochem Biophys Methods* **2006,** *68* (1), 1-21.
10. Swerts, K.; Ambros, P. F.; Brouzes, C.; Navarro, J. M.; Gross, N.; Rampling, D.; Schumacher-Kuckelkorn, R.; Sementa, A. R.; Ladenstein, R.; Beiske, K., Standardization of the immunocytochemical detection of neuroblastoma cells in bone marrow. *J Histochem Cytochem* **2005,** *53* (12), 1433-40.
11. Harwood, M. M.; Christians, E. S.; Fazal, M. A.; Dovichi, N. J., Single-cell protein analysis of a single mouse embryo by two-dimensional capillary electrophoresis.

J Chromatogr A **2006,** *1130* (2), 190-4.

12. Marko-Varga, G.; Berglund, M.; Malmstrom, J.; Lindberg, H.; Fehniger, T. E., Targeting hepatocytes from liver tissue by laser capture microdissection and proteomics expression profiling. *Electrophoresis* **2003,** *24* (21), 3800-5.
13. Seshi, B., Proteomics strategy based on liquid-phase IEF and 2-D DIGE: application to bone marrow mesenchymal progenitor cells. *Proteomics* **2007,** *7* (12), 1984-99.
14. Sitek, B.; Potthoff, S.; Schulenborg, T.; Stegbauer, J.; Vinke, T.; Rump, L. C.; Meyer, H. E.; Vonend, O.; Stuhler, K., Novel approaches to analyse glomerular proteins from smallest scale murine and human samples using DIGE saturation labelling. *Proteomics* **2006,** *6* (15), 4337-45.
15. Umar, A.; Luider, T. M.; Foekens, J. A.; Pasa-Tolic, L., NanoLC-FT-ICR MS improves proteome coverage attainable for approximately 3000 laser-microdissected breast carcinoma cells. *Proteomics* **2007,** *7* (2), 323-9.
16. Wang, H.; Qian, W. J.; Mottaz, H. M.; Clauss, T. R.; Anderson, D. J.; Moore, R. J.; Camp, D. G., 2nd; Khan, A. H.; Sforza, D. M.; Pallavicini, M.; Smith, D. J.; Smith, R. D., Development and evaluation of a micro- and nanoscale proteomic sample preparation method. *J Proteome Res* **2005,** *4* (6), 2397-403.
17. Xu, F.; Zou, L.; Liu, Y.; Zhang, Z.; Ong, C. N., Enhancement of the capabilities of liquid chromatography-mass spectrometry with derivatization: general principles and applications. *Mass Spectrom Rev* **2011,** *30* (6), 1143-72.

Chapter 3 - Evaluation of Different Cell Lysis Buffers and Protein Solubilizing Detergents on 2-MEGA Labeling Chemistry

3.1 Introduction

In addition to profile proteins expressed in a bio-system, the identification and quantification of protein profile changes in cells, tissues or biofluids of different origins or states of diseases is increasingly being recognized as a key objective in proteomics research[1-3]. The measurement of proteome changes in health and disease condition provides more direct and more accurate information in cellular dynamic changes, comparing with genomics study[4]. Therefore, quantitative proteomics is an important way to discover biomarkers for disease diagnosis and prognosis, and also to understand biological processes and mechanisms.

Mass spectrometry (MS) is one of the most efficient methods to identify and quantify large scale proteins due to its high sensitivity, selectivity, and accuracy in molecular masses. The ability of MS to be coupled with separation techniques, e.g., liquid chromatography (LC) or capillary electrophoresis (CE), further expands the quantification ability of MS to complex protein samples[5]. Based on MS, a range of proteomic quantification approaches have been developed. They can be broadly categorized as either label-based or label-free, depending on whether or not isotopes are introduced for quantitation[6-7]. Regardless of the strategy employed, both general methods offer their own advantages. Label-free methods typically use additional information from identified peptides across multiple runs, such as ion current intensity or frequency of MS/MS sequencing, to determine relative changes between various samples[8]. Label-based methods utilize relative signal intensities from isotopically-encoded references. 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 labelled peptides[9]. Because the usage of internal standards, label-based methods are more accurate than label-free.

Our group has recently demonstrated that differential 2-MEGA (dimethylation after guanidation) labeling of N-termini of peptides with $^{12}CH_2O$- and $^{13}CD_2O$-formaldehyde, after blocking the amino groups on the side chains of lysines by guanidination with O-methylisourea, is a promising labeling strategy for MS-based global quantitative proteome analysis[10-11], mainly due to the following reasons: (1) the uniform 6.032 Da mass difference between each derivatized peptide pair eliminates the significant overlapping of isotope envelopes, even for a peptide pair of around 3000 Da, and simplifies the quantification data analysis process; (2) the reaction itself is simple, fast, and complete and also can be done with commercially available and inexpensive reagents. The 2-MEGA labeling chemistry coupled with LC-MS analysis has been successfully applied to understand signal pathways of Bax-regulated apoptosis in HCT116 human colon carcinoma cells[12]. In our group, this stable-isotope labeling method is continuously applied in a number of proteomic quantification analyses, including comparing healthy and diseased biologic systems to identify potential biomarkers.

In these proteomic quantification studies, before proteins are digested and peptides are labeled by 2-MEGA chemistry, biological samples, like cells and tissues, need to be lysed to release their proteome. Some bacteria cells, like *E. coli*, can be lysed by simple physical methods, like French Press[13]. However, for most animal or plant cells, simple physical methods may not be enough to lyse cells effectively. Usually, a combination of lysis buffers and physical cell lysis methods is applied to ensure high efficient cell lysis and protein extraction[14-16]. After cell lysis, before many analyses can be done, extracted proteins need to be dissolved in solution. In this step, detergent molecules are normally used to help protein dissolution[17]. Because of the wide application of lysis buffers and detergents in biological sample preparations, in order to perform label-based quantitative analysis, the compatibility of cell lysis buffers or detergents, used in early stages of biological sample preparation, with the 2-MEGA chemical reactions need to be evaluated.

In this study, the compatibility of two commonly applied lysis buffers and seven

detergents with the 2-MEGA labeling chemistry is evaluated using LC-ESI quadrupole time-of-flight (QTOF) MS. By comparison between the proteomic analyses of cell lysis buffers and detergents from the same *E. coli* K-12 proteome, it is demonstrated that the efficiency of 2-MEGA labeling chemistry is not affected by the presence of the two tested cell lysis buffers and most detergents that used to assist protein dissolving. We show that more than 95% of the peptides can be correctly labeled.

3.2 Experimental Section

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 Lysis

Escherichia coli K-12 (*E. coli*, ATCC 47076) was from the American Type Culture Collection. A single *E. coli* K12 colony was used to inoculate 10 mL of LB broth (BBL, Becton Dickinson). The culture was incubated overnight with shaking at 37 °C. This saturated culture (1.5 mL) was added to 90 mL of growth medium in a 500-mL baffled Erlenmeyer flask. Cells were harvested in the mid-log phase by centrifugation at 3200 g for 10 min at 4 °C, resuspended, washed in 50 mM MOPS buffer, pH 7.3, and collected by centrifugation at 3200 g for 10 min at 4 °C.

Cell pellets were lysated using CelLytic^{TM} M buffer (Sigma-Aldrich) or TM buffer (Sigma-Aldrich), supplemented with 1 mM phenylmethylsulfonyl fluoride and protease and phosphatase inhibitor cocktails (Sigma-Aldrich). Cell debris was removed by centrifuging at 100×g for 15 min and the supernatant was collected. Then, the protein supernatant was divided in 7 vials, and precipitated by adding acetone to a final concentration of 80% and incubated in -80℃ for 4 hr, in order to remove salts and other impurities. The mixture was centrifuged at 20,000 g for 15

min.

3.2.3 Protein Solubilization and In-solution Digestion

Proteins pellets were respectively redissolved by 7 different salts and detergents: NH_4HCO_3, MeOH, Urea, SDS, Rapigest (Waters, Mississauga, ON, Canada), ProteaseMax (Promega, Madison, WI, U.S.) and Anionic Acid Labile Surfactant (AALS, VPQ Scientific, Toronto, ON, Canada). The protein solutions were reduced with dithiothreitol (DTT) and alkylated with iodoacetaminde (IAA), followed by trypsin digestion with a ratio of 1:50 at 37℃ overnight.

3.2.4 2-MEGA Labeling

Peptides dissolving in different detergents were isotopic labeled with 2-MEGA method on a liquid handler (Gilson, MA) as described previously[12, 18] with some modifications. Trypsin in the 500 µL tryptic digest solution (about 1 µg/µL) was irreversibly inactivated by adding 24 µL 2 M sodium hydroxide to adjust pH equal to 11. The ε-amino groups of lysines were blocked by reacting with 100 µL 6 M O-methylisourea at 37℃ for 1 hr and the reaction was stopped by adjusting pH to 5 with 48 µL 3 M hydrogen chloride. Dimethylation with 10 µL d(0), ^{12}C-formaldehyde (4%, w/w) was carried out at 37℃ for 30min, using 40 µL 2 M 2-picoline borane as reductive reagent. 32 µL 1 M ammonium bicarbonate was added to the reaction mixture followed by incubation at 37℃ for 15 min to consume excess formaldehyde. Then, the reaction was stopped by adjust pH to 2 with 10% trifluoroacetic acid (TFA). This whole reaction process was carried out on a liquid handler automatically.

3.2.5 Peptide Desalting and Quantification by RPLC

The 2-MEGA labeled peptides were desalted and quantified on an Agilent 1100 HPLC system, with a 4.6 mm × 50 mm Polaris C18 A column (3 µm particle and 300 Å pore). Solvent A consisted of 0.1% TFA in water, and Solvent B consisted of 0.1% TFA

in Acetonitrile (ACN). The labeled peptides were desalted by a step gradient: 0-10 min, 5% Solvent B; 10.01-15 min, 85%; 15.01-25 min, 5%.

3.2.6 Protein Identification with LC-MS

After desalting and quantification, the digests were analyzed using a QTOF Premier mass spectrometer equipped with a nanoACQUITY Ultra Performance LC system. Briefly, the desalted digests were concentrated on Speedvac and reconstituted with 0.1% formic acid to 0.2 µg/µL. Then a specific amount of digest solution was injected onto a 75 µm × 100 mm Atlantis dC18. 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 solvent gradient of 130 min: 0-2 min, 2-7% Solvent B; 2-85 min, 7-20%; 85-105 min, 20-30%; 105-110 min, 30-45%; 110-120 min, 45-90%; 120-125 min, 90%; and 125-130 min, 90-2%. The separated peptides were electrosprayed into the mass spectrometer fitted with a nanoLockSpray source at a flow rate of 350 nL/min. One MS scan was acquired from m/z 350-1600 for 0.8 s, followed by 8 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 by using ProteinLynx Global Server 2.2.5 (Waters). 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 Escherichia coli 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: Dimethylationd0 (N-term), Dimethylationd0 (K), Guanidination

(N-term), Guanidination (K), Oxidation (M). 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.

3.3 Results and Discussion

3.3.1 Two Cell Lysis Buffers' Influences on 2-MEGA Labeling Chemistry

In our group, CelLytic™ M and TM buffer are two most commonly used cell lysis buffers when breaking cells. In order to answer the question of whether the presence of these two buffers will cause any problem in 2-MEGA labeling, we labeled the *E. coli* digests, lysed with CelLytic™ M or TM buffer, with or without acetone precipitation and compared their labeling efficiencies (Figure 3.1).

E. coli cells pellets were respectively lysed as described in their protocol with CelLytic™ M buffer or TM buffer. After cell lysis, each of them was divided into two vials; one was added pre-cooled acetone (4×sample volume) and incubated in -80℃ for 4 hr, and another one stayed in 4℃. Acetone precipitation was used to remove salts and detergents. Then, protein precipitates were collected by centrifuging at 14000 rpm, 4℃, for 10 min, and redissolved by 100 mM NH_4HCO_3. After protein reduction, alkylation and trypsin digestion, 2-MEGA labeling on trypic peptides was performed. The labeled peptides were desalted and quantified on RPLC before injected onto LC- ESI-QTOF.

Table 3.1 summarizes the correctly labeled peptide number, total detected peptide number and correct rate of 2-MEGA reaction under each condition. With the use of one dimension (1D) separation before MS analysis, we identified more than 1000 peptides for each sample. *E. coli* sample lysed with TM buffer gave arise the identification of 100 more peptides than *E. coli* lysed with CelLytic™ M buffer.

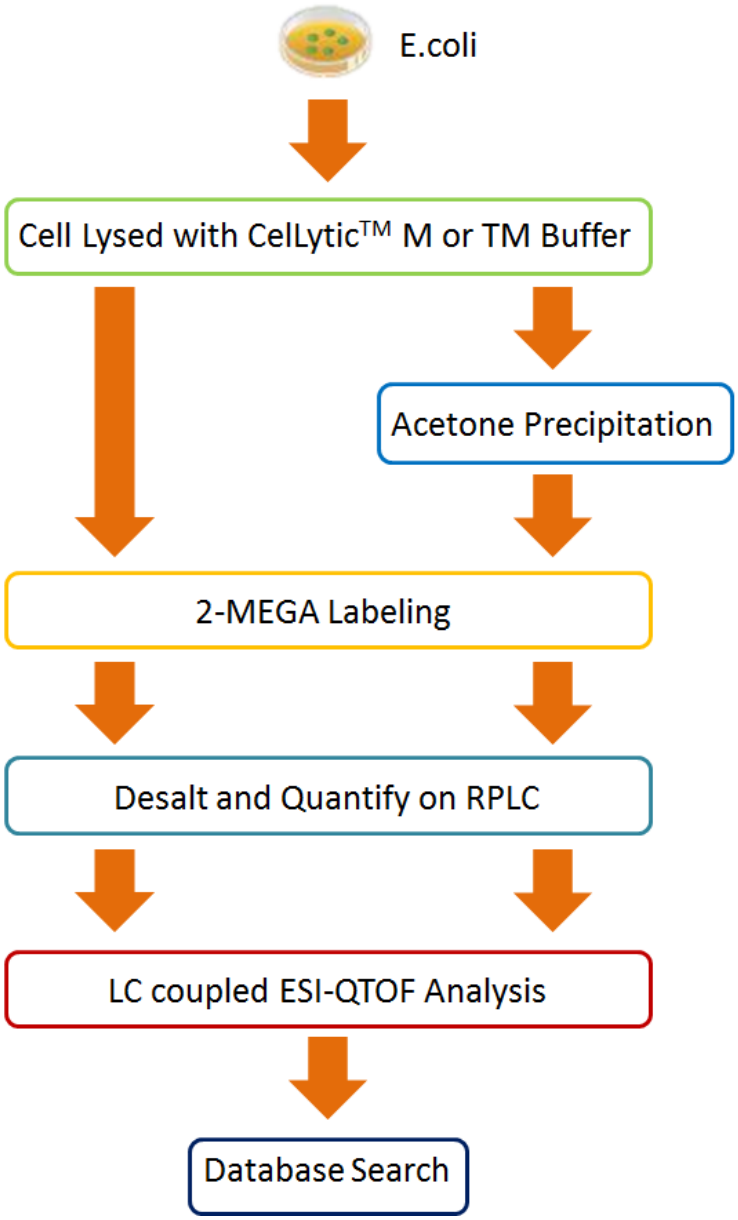


Figure 3.1. Workflow of cell lysis buffers experiment.

Table 3.1. Correctly and total labeled peptide number, correct rates of 2-MEGA labeling chemistry for *E. coli* proteome with or without CelLytic M™, TM buffer.

	CelLytic M™		CelLytic M™ with Acetone Precipitate		TM		TM with Acetone Precipitate	
Correct Labeled Peptide #	975	966	1029	1010	1144	1143	1109	1107
Total Labeled Peptide #	1037	1024	1094	1067	1204	1204	1180	1177
Correct Labeling Rate	94.02%	94.34%	94.06%	94.66%	95.02%	94.93%	93.98%	94.05%

In each case, with or without cell lysis buffer, the correct labeling efficiency of 2-MEGA on peptides was over 94%. If we looked at the peptides identified from each lysis buffer or sample before or after acetone precipitation, they have larger than 60% overlaps (Figure 3.2). Considering the instrument variations, which were usually counted for ~30% in MS-based peptide identification[19], the 60% overlaps may indicate that almost the same sample components were analyzed. Figure 3.3 shows the percentage of frequently happened side reactions under each reaction condition. We found the most important side reaction for all four reaction conditions was guanidination on peptides' N-terminal, which was about 1.5 to 2% in all the identified peptides. The second frequently happened side reaction was missing quanidination on lysine's side chain (K), this was happened in around 1.5% of total identified peptides. Other side reactions, like missing dimethylation on N-terminal, guanidination on N-terminal but missing guanidination on K, etc., their occurrence rate were less than 1% of the total identities. These results indicate that in 2-MEGA labeling, the tricky part was the guanidination reaction; most percentage of side reactions were happened in this step, while the dimethylation was relatively robust. In the four labeling conditions, they all had similar correct labeling efficiency and side reaction profiles. These data indicate that 2-MEGA labeling reaction is compatible with CelLytic™ M and TM buffer.

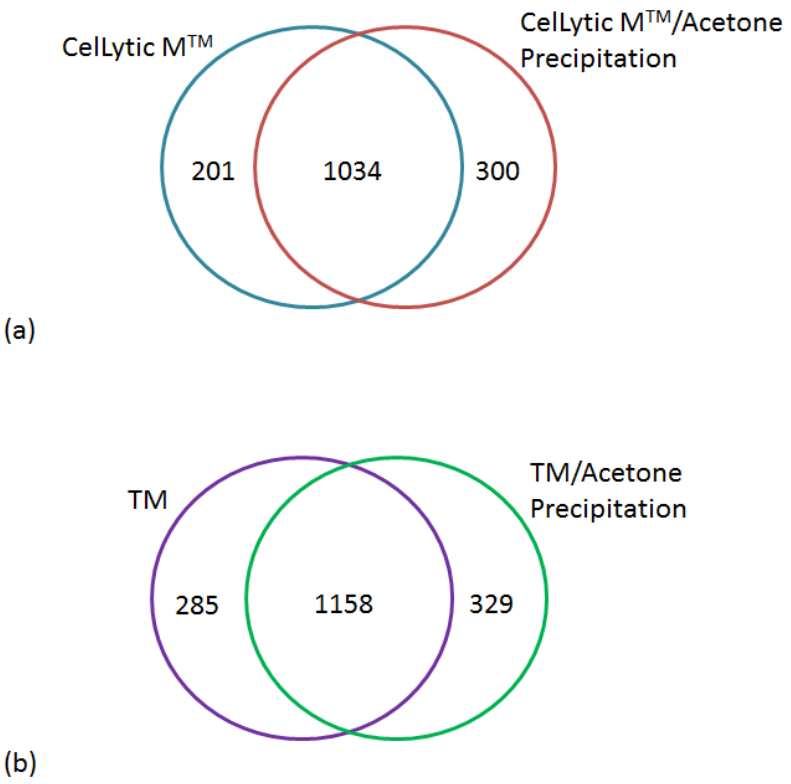


Figure 3.2. Overlap of identified *E. coli* peptides from 2-MEGA labeling with or without cell lysis buffers. (a) Overlaps between *E. coli* proteome samples with CelLytic M™ and CelLytic M™/acetone precipitation. (b) Overlaps between *E. coli* proteome samples with TM and TM/acetone precipitation.

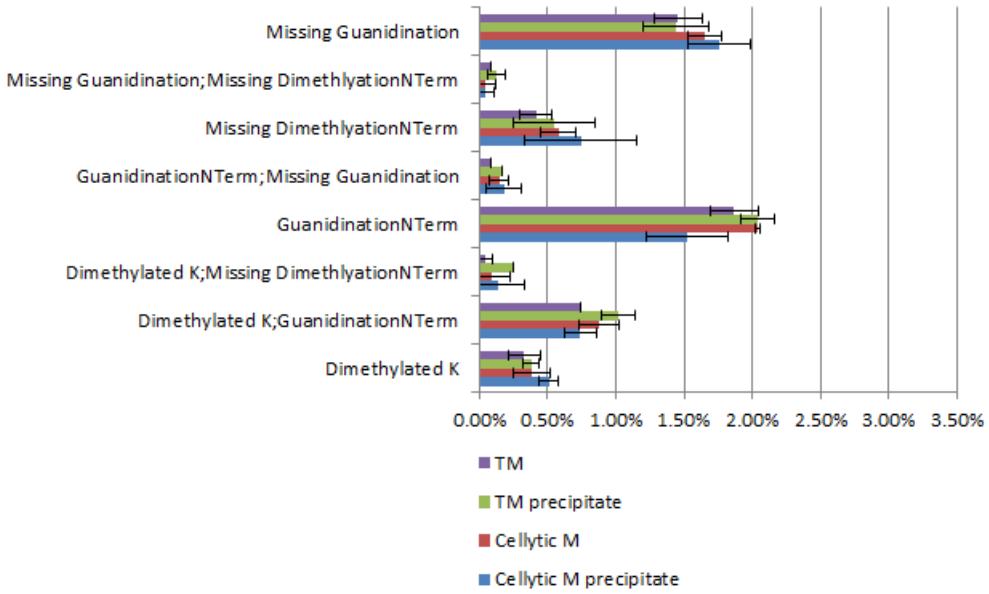


Figure 3.3. Mis-reactions in 2-MEGA labeling with or without cell lysis buffers.

3.3.2 Seven Salts and Detergents' Influences on 2-MEGA Labeling Chemistry

In proteomic research, one most important step in sample preparation is protein solubilization, as proteins in solid phase are not suitable for many analyses. The situation in MS is the same. In shotgun proteome analysis, proteins have to be dissolved in a proper solution before they can be processed, separated and analyzed. This is why researchers try to use various kinds of salts, organic solutions and detergents to help protein solubilization[20-23]. In this project, we studied several widely applied molecules helping protein solubilization, and evaluated their influences on the efficiency of 2-MEGA reaction (Figure 3.4). As all reported study of 2-MEGA labeling chemistry was made in peptide ammonium bicarbonate solution, *E. coli* digests in ammonium bicarbonate solution would become the control in our experiment.

E. coli cells were lysed by sonication with TM buffer. After removing cell debris, protein pellets were acetone precipitated and collected as described in the experimental section. The protein pellets were then dissolved, respectively, in 100 mM NH_4HCO_3, 60% MeOH, 6 M Urea, 0.6% SDS, 0.6% Rapigest, 0.15% ProteaseMax and 0.6% AALS. The first four are most commonly used molecules in proteomics to help protein solubilization, and the last three are acid cleavable surfactants, which are compatible with MS. The seven protein solutions were processed as described through reduction, alkylation, trypsin digestion and 2-MEGA labeling. After desalting and quantification, they were analyzed by LC-MS. Peptide identification number and 2-MEGA labeling efficiency for each solution are shown in Table 3.2. From this table, we can see that AALS gave the highest peptide identification number, ~1500, and SDS gave lower identities, ~950. This is understandable if we consider the extra SDS-removal step that was necessary for SDS dissolved protein samples before loading into LC-MS; SDS is not compatible with MS analysis. The extra purification step caused extra sample loss in sample preparation. Other solutions, MeOH, Urea, Rapigest and ProteaseMax, gave similar peptide identification numbers, which were only slightly higher than the control group, NH_4HCO_3.

Table 3.2. Correct and total labeled peptide number, correct rates of 2-MEGA labeling chemistry for *E. coli* proteome dissolved in seven commonly used reagents.

	Correct Labeled Peptide #	Total Labeled Peptide #	Correct Labeling Rate
NH_4HCO_3	960	1008	95.24%
	903	965	93.58%
MeOH	1003	1052	95.34%
	1234	1297	95.14%
Urea	1089	1183	92.05%
	1167	1263	92.40%
SDS	915	965	94.82%
	893	944	94.60%
Rapigest	1090	1140	95.61%
	1250	1322	94.55%
ProteaMax	1232	1290	95.50%
	1103	1162	94.92%
AALS	1492	1561	95.58%
	1363	1434	95.05%

Comparing their correct labeling rates, six salt and detergent molecules, NH_4HCO_3, MeOH, Urea, Rapigest, ProteaseMax and AALS, had similar 2-MEGA labeling efficiency, which may indicate they will not have effects on the reaction. However, from Table 3.2, we can see that the correct labeling rate for urea was only around 92%. What caused the 3% decrease for 2-MEGA labeling efficiency in the urea peptide solution? In order to answer this question, we analyzed the category of side reactions in the 7 solutions (Figure 3.5). From this figure, we can see that all other side reactions had similar distributions in all 7 solutions: the most frequently happened side reaction in 2-MEGA labeling was still guanidination on peptide N-terminal, in most solutions it's over 2.5% in total peptide identities. However, in urea solution, it had one side reaction, which was greatly increased compared to other 6 solutions: missing guanidination on lysine (K). In urea solution, it went up to 1.85%, while in other solutions, it's less than 0.2%. There are reports about the possible peptide modifications in sample preparation, and many of them had mentioned that urea may cause carbamylation on peptide or protein's free amine

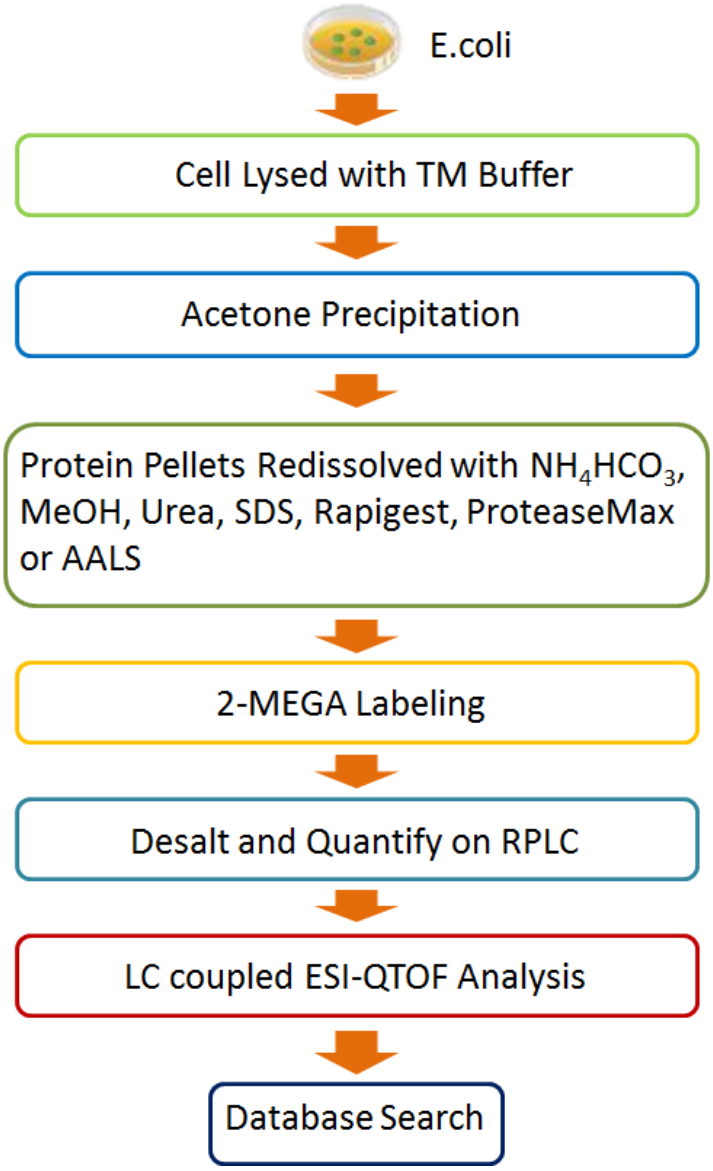


Figure 3.4. Workflow of protein dissolving detergents experiment.

group, like N-terminal amine or side chain of lysine or arginine[24-26]. Based on this information, we searched all identification data again in our database, with carbamyl (N-term) and carbamyl (K) as variable modification, to evaluate the frequency of carbamylation in our samples (Figure 3.6). In urea solution, carbamylation was detected in over 2% of the total identified peptides; in other 6 solutions, it only happened on less than 0.5% of identified peptides. In this sense, carbamylation could be one reason of the lower correct 2-MEGA labeling efficiency in urea peptide solutions. As the urea molecules may react with the free amine group on the side chain of K and N-terminus, carbamylation blocked the reaction spots for 2-MEGA labeling, thus caused missing labeling or side reactions.

3.4 Conclusion

We have examined the correct labeling efficiency of 2-MEGA on peptides with or without two cell lysis buffers: CelLytic™ M and TM buffer. Our results showed the two commonly used lysis buffers would not cause any influence on 2-MEGA labeling chemistry. We have also evaluated seven reagents commonly used to help protein solubilization, namely NH_4HCO_3, MeOH, Urea, SDS, Rapigest, ProteaseMax and AALS. From our experiment results, we demonstrate that NH_4HCO_3, MeOH, SDS, Rapigest, ProteaseMax and AALS are compatible with 2-MEGA labeling, without causing any extra mis-labeling or side reaction. However, urea may cause a decrease in the correct labeling efficiency of 2-MEGA reaction, because of carbamylation reaction occured in the sample preparation process, causing interference with the 2-MEGA labeling step.

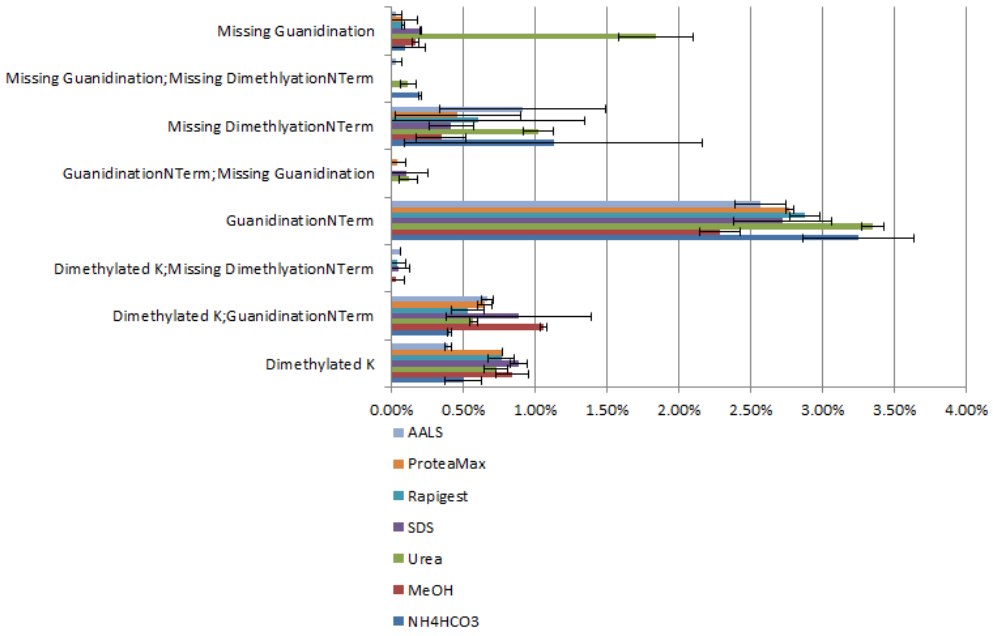


Figure 3.5. Mis-reactions in 2-MEGA labeling with seven commonly used reagents.

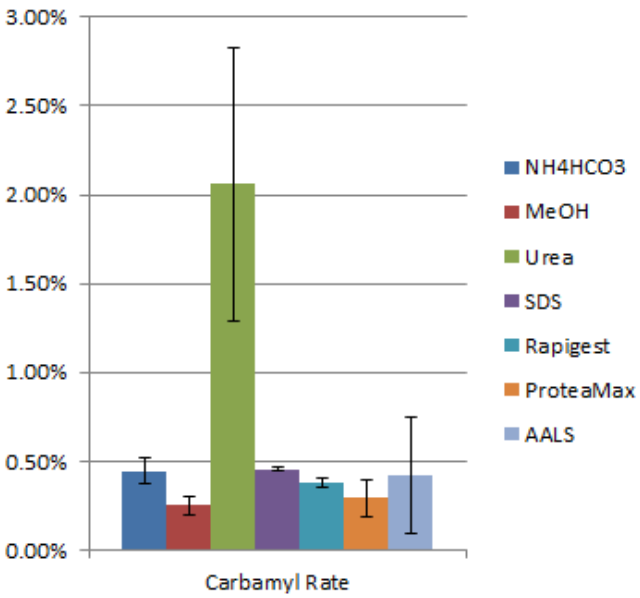


Figure 3.6. Carbamyl percentage of *E. coli* proteome in 2-MEGA labeling with seven commonly used reagents.

3.5 Literature Cited

1. Brasier, A. R.; Garcia, J.; Wiktorowicz, J. E.; Spratt, H. M.; Comach, G.; Ju, H.; Recinos, A., 3rd; Soman, K.; Forshey, B. M.; Halsey, E. S.; Blair, P. J.; Rocha, C.; Bazan, I.; Victor, S. S.; Wu, Z.; Stafford, S.; Watts, D.; Morrison, A. C.; Scott, T. W.; Kochel, T. J., Discovery proteomics and nonparametric modeling pipeline in the development of a candidate biomarker panel for dengue hemorrhagic fever. *Clin Transl Sci* **2012,** *5* (1), 8-20.
2. van de Guchte, M.; Chaze, T.; Jan, G.; Mistou, M. Y., Properties of probiotic bacteria explored by proteomic approaches. *Curr Opin Microbiol* **2012,** *15* (3), 381-9.
3. van Gool, A. J.; Hendrickson, R. C., The proteomic toolbox for studying cerebrospinal fluid. *Expert Rev Proteomics* **2012,** *9* (2), 165-79.
4. Vogel, C.; Marcotte, E. M., Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **2012,** *13* (4), 227-32.
5. Thelen, J. J.; Miernyk, J. A., The proteomic future: where mass spectrometry should be taking us. *Biochem J* **2012,** *444* (2), 169-81.
6. Tao, W. A.; Aebersold, R., Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr Opin Biotechnol* **2003,** *14* (1), 110-8.
7. Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **2012,** *404* (4), 939-65.
8. Zhu, W.; Smith, J. W.; Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010,** *2010*, 840518.
9. Lill, J., Proteomic tools for quantitation by mass spectrometry. *Mass Spectrom Rev* **2003,** *22* (3), 182-94.
10. Ji, C.; Li, L., Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS. *J Proteome Res* **2005,** *4* (3), 734-42.

11. Ji, C.; Zhang, N.; Damaraju, S.; Damaraju, V. L.; Carpenter, P.; Cass, C. E.; Li, L., A study of reproducibility of guanidination-dimethylation labeling and liquid chromatography matrix-assisted laser desorption ionization mass spectrometry for relative proteome quantification. *Anal Chim Acta* **2007,** *585* (2), 219-26.
12. Wang, P.; Lo, A.; Young, J. B.; Song, J. H.; Lai, R.; Kneteman, N. M.; Hao, C.; Li, L., Targeted quantitative mass spectrometric identification of differentially expressed proteins between Bax-expressing and deficient colorectal carcinoma cells. *J Proteome Res* **2009,** *8* (7), 3403-14.
13. Amari, J. V.; Mazsaroff, I., Analysis of recombinant human interleukin-11 fusion protein derived from Escherichia coli lysate by combined size-exclusion and reversed-phase liquid chromatography. *J Chromatogr A* **1996,** *729* (1-2), 113-24.
14. Thornhill, A. R.; McGrath, J. A.; Eady, R. A.; Braude, P. R.; Handyside, A. H., A comparison of different lysis buffers to assess allele dropout from single cells for preimplantation genetic diagnosis. *Prenat Diagn* **2001,** *21* (6), 490-7.
15. Connor, P.; Talavera, F.; Kang, J. S.; Burke, J.; Roberts, J.; Menon, K. M., Epidermal growth factor activates protein kinase C in the human endometrial cancer cell line HEC-1-A. *Gynecol Oncol* **1997,** *67* (1), 46-50.
16. Gilbert, C.; Rollet-Labelle, E.; Caon, A. C.; Naccache, P. H., Immunoblotting and sequential lysis protocols for the analysis of tyrosine phosphorylation-dependent signaling. *J Immunol Methods* **2002,** *271* (1-2), 185-201.
17. Majumdar, D.; Rosser, R.; Havard, S.; Lobo, A. J.; Wright, P. C.; Evans, C. A.; Corfe, B. M., An integrated workflow for extraction and solubilization of intermediate filaments from colorectal biopsies for proteomic analysis. *Electrophoresis* **2012,** *33* (13), 1967-74.
18. Ji, C.; Guo, N.; Li, L., Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. *J Proteome Res* **2005,** *4* (6), 2099-108.
19. Barnidge, D. R.; Tschumper, R. C.; Jelinek, D. F.; Muddiman, D. C.; Kay, N. E., Protein expression profiling of CLL B cells using replicate off-line strong cation exchange chromatography and LC-MS/MS. *J Chromatogr B Analyt Technol*

Biomed Life Sci **2005,** *819* (1), 33-9.

20. Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D., Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. *J Proteome Res* **2002,** *1* (4), 351-60.
21. Goshe, M. B.; Blonder, J.; Smith, R. D., Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J Proteome Res* **2003,** *2* (2), 153-61.
22. Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R., Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* **2001,** *19* (10), 946-51.
23. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001,** *19* (3), 242-7.
24. Angel, P. M.; Orlando, R., Quantitative carbamylation as a stable isotopic labeling method for comparative proteomics. *Rapid Commun Mass Spectrom* **2007,** *21* (10), 1623-34.
25. McCarthy, J.; Hopwood, F.; Oxley, D.; Laver, M.; Castagna, A.; Righetti, P. G.; Williams, K.; Herbert, B., Carbamylation of proteins in 2-D electrophoresis--myth or reality? *J Proteome Res* **2003,** *2* (3), 239-42.
26. Ye, X.; Li, L., Microwave-assisted protein solubilization for mass spectrometry-based shotgun proteome analysis. *Anal Chem* **2012,** *84* (14), 6181-91.

Chapter 4 - Quantitative Proteomic Analysis of HER2 Normal and Overexpressing MCF-7 Breast Cancer Cells Revealed Proteomic Changes Accompanied With HER2 Gene Amplification

4.1 Introduction

Breast cancer is among the most common malignancies, and remains a major cause of mortality in women worldwide[1]. In recent decades, age-adjusted mortality from breast cancer has improved in the developed world. Clinical outcomes are heavily dependent on stage of the disease at diagnosis and the 5-year survival rate decreases from 98% for localized disease to 26% for late stage disease[2]. Treatment for breast cancer is increasingly informed by the molecular heterogeneity of the disease, where biomarkers drive breast cancer prognostication and informed clinical treatment decisions. One critical breast cancer biomarker is the human epidermal growth factor receptor 2 (HER2, Neu or ErbB-2). HER2 is encoded by *ERBB2* gene and about 15~20% of breast cancer cases have *ERBB2* DNA amplification with consequent overexpression of HER2 protein[3], which is associated with aggressive disease, as evidenced by poor prognosis and increased likelihood of recurrence. The aberrant overexpression and activation of HER2 is thought to trigger multiple cellular signaling pathways that drive abnormal cell proliferation, survival loss of normal growth inhibitory constraints and metastasis, but the mediators of these downstream effects are incompletely understood. FDA approved HER2 directed therapies include trastuzumab and lapatinib, but fewer than 50% of breast cancer patients respond to anti-HER2 monotherapy, and even initially sensitive tumours typically develop clinical resistance[4]. The mechanisms of resistance to anti-HER2 therapies are also poorly understood.

The HER2 phenotype is triggered by gene amplification event(s) that lead to an increase in the number of *ERBB2* gene copies in the malignant cell nucleus. This abnormality is accompanied by increases in *ERBB2* mRNA and HER2 protein. However, genomic changes do not necessarily have a direct correlation with

proteomic changes due to translational regulation, post-translational modification and alternative splicing. Based on this understanding, diverse proteomic studies have been done in this area and several proteins have been reported as protential HER2-associated biomarkers for breast cancer[5-7]. But, to date, these proteins have not been evaluated in the context of clinical relevance or clinical prognosis.

We sought to identify proteomic changes accompanying *ERBB2* gene amplification in order to i) identify new biomarkers associated with the HER2 phenotype, ii) gauge the magnitude of the proteomic changes triggered by amplification of this single gene, and iii) derive a better understanding of the downstream biological changes triggered by HER2 overexpression.

To address these questions, we first performed quantitative proteomic analysis using isogenetic MCF-7 cell lines differing only in *ERBB2* gene copy number. We performed quantitative proteomic analysis of wild-type (parental) and overexpressing (stably transfected) Michigan Cancer Foundation-7 (MCF-7) human breast cancer cell lines using offline two-dimensional liquid chromatography (2D-LC) coupled mass spectrometry (MS) based isotopic labeling quantitative proteomic approach. For stable isotopic labeling, we applied the technique of N-terminal dimethylation after lysine guanidination (2-MEGA), which had been developed by our group[8-9]. Based on the differential protein list obtained from these two cell lines, biomarker candidates for HER2-positive breast cancer were selected, validated by western blotting and evaluated by immunohistochemical analysis of clinical breast cancer samples.

4.2 Experimental Section

4.2.1 Chemicals and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich (Oakville, ON) unless stated otherwise. The isotope reagent, d(6), ^{13}CD-formaldehyde (20%, w/w in deuterated H_2O), was purchased from Cambridge Isotope Laboratories, Inc. (Andover,

MA). Progenta™ Anionic Acid Labile Surfactant I was obtained from Protea Bioscience, Inc. (Morgantown, WV). Acetonitrile and water were purchased from Fisher Scientific Canada (Edmonton, AB). AIF rabbit monoclonal antibody was purchased from Epiomics. CPTC-GSTMu3-1-s mouse monoclonal antibody was purchased from Developmental Studies Hybridoma Bank, University of Iowa, (Iowa City, IA). FAM129A rabbit polyclonal antibody and RanGAP1 rabbit monoclonal [EPF3295-ab92360] antibody were obtained from Abcam (Cambridge, MA). SQSTM1 (D-3) mouse monoclonal antibody was obtained from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA).

4.2.2 Cell Culture

MCF-7 and MCF-7 stable transfectants expressing neomycin (vector) or HER2+ (kindly provided by Dr. Dennis J. Slamon[10]) were cultured in MEM (Gibco BRL, Life Technologies, Burlington, ON) supplemented with 10% bovine serum (FBS) and antibiotics. Cells were harvested at 70-80% confluence. Three replicates of wild type and HER2 overexpressing MCF-7 cells were cultured under the same conditions in three different days to obtain bio-triplicate samples.

4.2.3 Cell Lysate and In-solution Digestion

Cells were lysed using CelLytic™ M buffer (Sigma-Aldrich, Oakville, ON), supplemented with 1 mM phenylmethylsulfonyl fluoride and protease and phosphatase inhibitor cocktails (Sigma-Aldrich, Oakville, ON). Cell debris was removed by centrifugation at 100×g for 15 min and the supernatant was collected. Protein concentration of the supernatant was determined by the BCA assay (Bio-Rad, Mississauga, ON). Proteins were reduced with dithiothreitol (DTT) and alkylated with iodoacetaminde. The reduced and alkylated proteins were then precipitated by adding acetone to a final concentration of 80% and incubated in -80°C for 4 hours, in order to remove salts and other impurities. The mixture was centrifuged at 20,000×g for 15 min. The supernatant was decanted and protein pellets were redissolved with

0.1% Progenta™ Anionic Acid Labile Surfactant I (AALS I) and digested by trypsin at a ratio of 1:50 at 37℃ overnight.

4.2.4 2-MEGA Isotopic Labeling

Peptides were isotopic labeled with 2-MEGA method on liquid handler as described previously[8-9] with some modifications. Trypsin in the 500 µL tryptic digest solution (about 1 µg/µL) was irreversibly inactivated by adding 24 µL 2 M sodium hydroxide to adjust pH equal to 11. The ε- amino groups of lysines were blocked by reacting with 100 µL 6M O-methylisourea at 37℃ for 1 hour and the reaction was stopped by adjusting pH to 5 with 48 µL 3 M hydrogen chloride. Dimethylation with 10 µL d(0), ^{12}C-formaldehyde, or d(6), ^{13}CD-formaldehyde (4%, w/w) was carried out at 37℃ for 30min, using 40 µL 2 M 2-picoline borane as reductive reagent. Then, 32 µL 1 M ammonium bicarbonate was added to the reaction mixture followed by incubation at 37℃ for 15 min to consume excess formaldehyde. Then, the reaction was stopped by adjust pH to 2 with 10% TFA. This whole reaction process was carried out on liquid handler automatically.

4.2.5 Strong Cation Exchange (SCX) Liquid Chromatography

After isotopic labeling, peptide mixtures were fractionated on SCX. A gradient for elution was established with solvent A (50 mM KH_2PO_4, pH 2.7) and solvent B (1 M KCl in 50 mM K_2PO_4, pH 2.7): 0-1 min, 0-4% solvent B; 1-12 min, 4-20%; 12-45 min, 20-60%; 45-50 min, 60-100%; 50-53 min, 100%; 53-55 min, 100-0%; and kept flushing the column with 0% solvent B until 62 min. Peptide fractions were collected in 1 min fractions from 2 min to 60 min, followed by desalting and quantification on RPLC. Less abundant neighbor fractions were pooled together to obtain fractions of ~1 µg peptides.

4.2.6 Liquid Chromatography Coupled Mass Spectrometric Analysis

An electrospray ionization (ESI) quadrupole time-of-flight (Q-TOF) Premier mass

spectrometer (Waters, Mississauga, ON) equipped with a nanoACQUITY Ultra Performance LC system (Waters, Mississauga, ON) was used to analyze the peptide fractions. Solvent A consisted of 0.1% formic acid in deionized water, and solvent B contained 0.1% formic acid in acetonitile. The following 130 min LC gradient was used to separate peptides: 0-2 min, 2-7% solvent B; 2-85 min, 7-20%; 85-105 min, 20-30%; 105-110 min, 30-45%; 110-120 min, 45-90%; 120-125 min, 90%; and 125-130 min, 90-2%. The separated peptides were electrosprayed into the mass spectrometer at a flow rate of 350 nl/min. After acquiring condition optimization, Mass spectra were acquired from m/z 350 to 1600 for 1 sec followed by four data dependent MS/MS analyses form m/z 50 to 1990 for 0.8 sec each, with a precursor ion exclusion strategy to eliminate redundant identification from two adjacent SCX fractions.

4.2.7 Database Search and Bioinformatics

MS and MS/MS spectral data were processed and searched using MASCOT DISTILLER. The software picked out paired peaks from individual MS spectra, calculated peak areas and worked out peptide ion relative intensity ratios in two comparative samples. MS/MS spectra taken for the paired peptides peaks were searched with following parameters: database: SWISSPROT; taxonomy: Homo Sapiens (Human); enzyme: trypsin; missed cleavage: 2; fixed modification: Carbamidomethylation (C); variable modification: Guanidinyl (K), Dimethylationd(0) (N-term, + C_2H_4, + 28.0313 Da); Dimethylationd(6) (N-term, + $^{13}C_2D_4$, + 34.0631 Da); MS tolerance: 0.2 Da; MS/MS tolerance: 0.1 Da.

Metacore was employed to predict the cellular signaling pathways of differentially expressed proteins and the Human Protein Atlas (http://www.proteinatlas.org/) was used to help select breast cancer biomarker candidates.

4.2.8 Western Blotting

Cell lysate were prepared as described above. Equal amount of total cellular proteins from the two cell lines were separated on SDS-PAGE gels, and transferred onto nitrocellulose membranes (Bio-Rad, Mississauga, ON) in an electronic field. Membranes were blocked in 5% (w/v) skim milk blocking solution at room temperature for 1 hour, followed by incubation with various primary antibodies at 4℃ overnight. Membranes were then washed, incubated with the horseradish perioxidase-conjugated secondary antibodies at room temperature for 1 hour and processed with standard enhanced chemiluminescence (ECL) reagents (Perkin Elmer, Waltham, MA).

4.2.9 Tumor Tissue Microarrays and Immunohistochemistry

Tissue microarrays were constructed from 371 specimens of primary breast cancers from women who consented to participate in the Canadian Breast Cancer Foundation (CBCF) Tumor Bank. The study populations (n=184; and n =187) have previously been described[11-12]. The study is approved by the institutional research ethics board, Alberta Cancer Research Ethics Committee (protocol #23134 and # 23140, respectively). Tissue micoarrays were made from formalin fixed breast tissues in triplicate 0.6 mm cores using the TMArrayer™ or 1.0 mm cores using the Beecher ATA-27. Tissues on slides were deparaffinized in xylene and rehydrated in decreasing concentrations of ethanol to water. Endogenous peroxidase was quenched in 3% H_2O_2 for 10 min. For antigen retrieval slides were placed in boiling citrate buffer pH 6.0 for 10 min or in Tris Target Retrieval Solution, pH 9 (Dako, Burlington, ON) followed by rinsing in water for 10 min. Tissues were incubated with the primary antibodies at room temperature in a humidified container for 1 hr. Slides were washed 2 times in PBS for 5 min. Slides were incubated in the secondary antibodies, anti-Rabbit or anti-Mouse EnVision+System-HRP (Dako, Burlington, ON), at room temperature for 30 min. They were washed 2 times in PBS and and incubated with 3,3'-dichlorobenzidine (DAB) (Dako, Burlington, ON). Slides were then rinsed in water

for 10 min and incubated in 1% copper sulfate for 5 min. They were counterstained with haematoxylin, dipped 3 times in saturated lithium carbonate, rinsed in water, dehydrated in increasing concentrations of ethanol and xylene and coverslipped. Specifics for each antibody are summarized in Table 4.1. Slides were scored by a single pathologist blinded to clinical outcomes. Cytoplasmic and nuclear immunoreactivity were scored separately, based on the average staining signal intensity throughout the tumor tissue on a scale of 0 (negative), 1 (weak), 2 (intermediate), and 3 (strong). Each tumor was represented in triplicate, and the staining was analyzed as a mean.

Table 4.1. List of the antibodies for tissue microarrays and immunohistochemistry.

Antigen	Antibody and Source	Antigen Retrieval Method	Antibody Dilution	Secondary Antibody	DAB (min)
AIFM1	AIF antibody RabMab Epitomics	Tris	1/50	anti-rabbit	10
GSTM3	CPTC-GSTMu3-1-s Developmental Studies Hybridoma Bank (DSHB), The University of Iowa	Tris	1/10	anti-mouse	2
NIBAN-cytoplasmic staining	FAM129A Antibody Abcam	Citrate	1/200	anti-rabbit	2
RAGP1	RanGAP1 Antibody [EPR3295] Abcam	Tris	1/400	anti-rabbit	10
SQSTM1	SQSTM1 (D-3) Santa Cruz Biotechnology	Citrate	1/50	anti-mouse	10

4.2.10 Prognostic Evaluation of Biomarkers

Overall survival was defined as the time from diagnosis to the time of death. Patients who were alive at the time of last follow up were considered censored. Recurrence free survival was calculated from the time of diagnosis to the time of recurrence, and patients who did not experience any kind of recurrence were censored for this analysis. The selected biomarkers for the study were dichotomized using the receiver operating curve (ROC) characteristics. The cut point chosen to dichotomize continuous data into categories is based on the point that provides maximum sensitivity and specificity. Kaplan Meier curves were plotted for overall survival as well as for the recurrence free survival and log rank tests were used to compare the curves based on the dichotomized biomarkers. Cox's proportional hazard model was used to obtain the hazard ratio and the corresponding 95% confidence interval. For our analysis, we had assumed the two tailed significance and a p-value < 0.05 was considered statistically significant. SAS (SAS Institute Inc., Cary, NC) software version 9.2 was used for statistical analysis.

4.3 Results and Discussion

4.3.1 Proteomic Quantification

Figure 4.1 shows the workflow using the forward (A_HB_L) and reverse (A_LB_H) labeling strategy for MS based quantitative proteomic analysis. Wild type HER2 expressing and HER2 overexpressing cells were lysed to extract proteins. The extracted proteins from cells were reduced, alkylated, precipitated and then digested with trypsin. Both MCF-7 wild type and HER2 overexpressing proteome samples were divided into two equal parts and labeled by the 2-MEGA method, with quantification and desalting on RPLC. Based on the quantification results of the total peptide amount in each cell sample, the labeled peptides were accurately mixed with 1:1 ratio. The heavy isotopic labeled wild type MCF-7 sample (A_H) was mixed with the light isotopic labeled HER2 overexpressing MCF-7 sample (B_L) to produce an

A_HB_L mixture, while the mixing of light isotopic labeled wild type sample and heavy isotopic labeled HER2 overexpressing sample giving an A_LB_H mixture. The two mixtures were fractionated on SCX respectively and were injected onto LC coupled ESI-QTOF for separation and MS analysis.

To extract identification and comparative quantification information, the software MASCOT Distiller was used. If the same peptide was present in both wild type and HER2 overexpressing samples, it would be identified as a peak pair in a MS spectrum with a mass difference of 6.032 Da for single charged ion. The program would calculate the intensity ratio of the peptide peak pair and identify it from its MS/MS spectra. Running through the SCX fractions, a quantitative peptides list would be generated by the program. For the forward (A_HB_L) and reverse (A_LB_H) labeled peptides samples, two different SCX fractions were analyzed and two quantified peptides lists were produced. The individual quantified peptides lists were matched against each other to find common peptides. Based on the identified and quantified common peptides list, a list of corresponding proteins was generated.

In order to increase the confidence of quantitative data, biological triplicates of wild type and HER2 overexpressing MCF-7 cells were analyzed on ESI-QTOF after 2-MEGA labeling. Three batches of identified and quantified protein lists were generated and matched with each other. A total of 2455 unique proteins were identified and quantified for the bio-triplicate cell samples. We applied 1.50/0.67 as the threshold for differential expression to filter out proteins having expression level changes from the protein lists, and 1278 proteins were found to be differentially expressed between wild type and HER2 overexpressing MCF-7 cells. Among these proteins, 231 were found to be present in all three bio-triplicates and had the same expression level changes, upregulated or downregulated. Out of the 231 differentially expressed proteins, 146 were downregulated, and 85 were upregulated in the HER2 overexpressing MCF-7 cells. Table 4.2 lists the proteins based on their relative abundance. The expression level changes of these 231 proteins were considered of

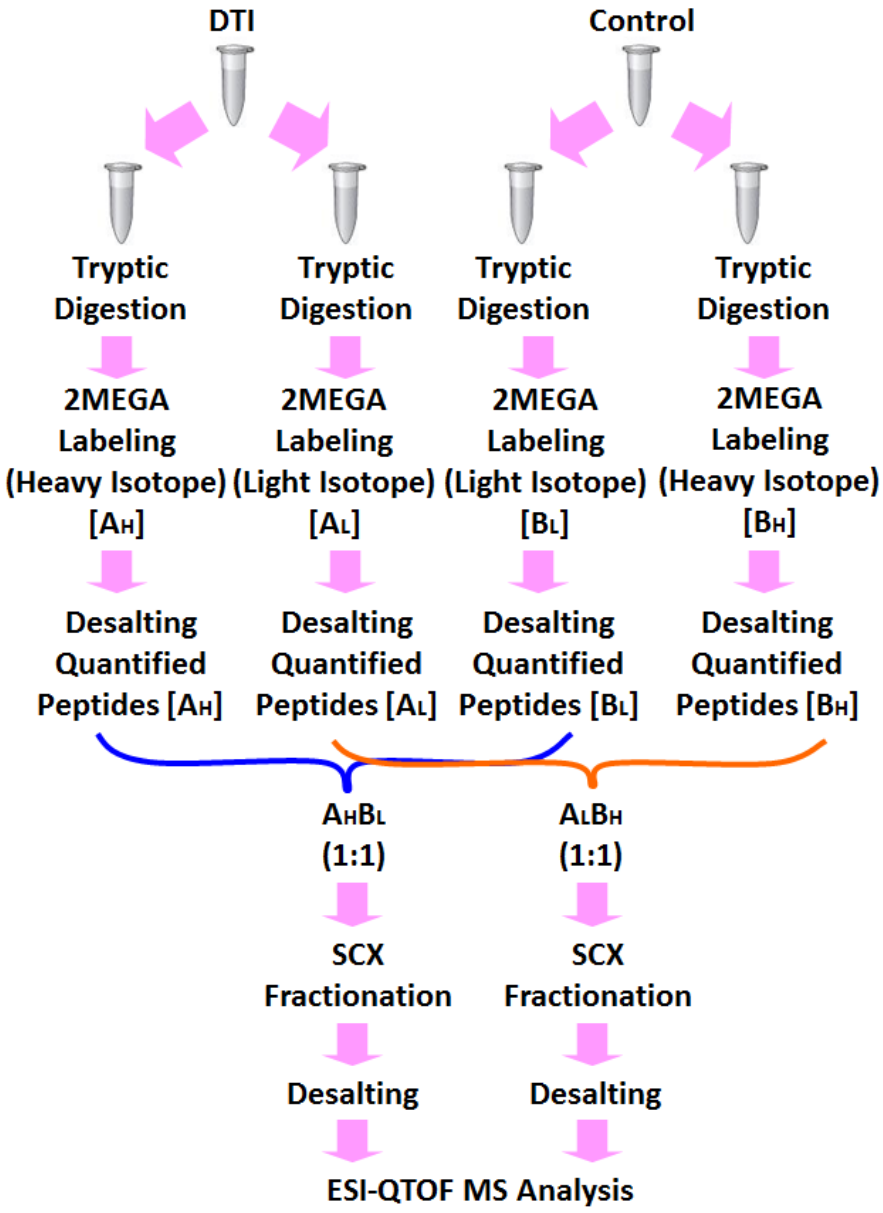


Figure 4.1. Workflow using the forward (A_HB_L) and reverse (A_LB_H) labeling strategy for MS based quantitative proteomic analysis

high confidence, and the protein list was applied for later biomarker candidates' selection and the following biological studies.

4.3.2 Bioinformatics and Selection of Biomarker Candidates

Based on their molecular functions, the 231 proteins were classified into different groups as shown in Figure 4.2A. 56.3% (130/231) of differentially expressed proteins have protein binding functions, while 37.2% (86/231) have catalytic activities. Proteins that have catalytic activities generally exhibit substrate specificity, catalyze certain biochemical reactions and participate in some biological signal transduction pathways. For example, MK03, which is better known by its gene name *MAPK3*, is a mitogen-activated protein kinase of the ERKs family. This protein which is known to be upregulated in HER2 positive breast cancers and functions downstream of HER2 intracellular signaling pathway[13-14] provided a positive internal control of these quantitative methods.

To predict the cellular signaling pathways, the 231 differentially expressed proteins with high confidence were analyzed with IPA[15-17] and MetaCore[18-19]. According to the prediction, 21.6% (50/231) of proteins were involved in pathways of cellular movement; 17.3% (40/231) were found to play roles in cell to cell signaling and interactions, 21.2% (49/231) participated in in cell death pathways, 16% (37/231) were involved in cell morphology, and 22.9% (53/231) were found in signaling pathways that regulate cell growth and proliferation (Figure 4.2B). Many known cancer-related cellular signaling pathways were found to be modulated by several of the identified proteins including p53, ERK/MAPK and vascular endothelial growth factor [(VEGF); see Figure 4.3].

Table 4.2. List of proteins identified to be differentially expressed between HER2+/Neo MCF-7 cells in biological triplicates.

Swiss-Prot ID	PROTEIN	RATIO
	Up-Regulated Proteins	
P04626	Receptor tyrosine-protein kinase erbB-2 (ERBB2_HUMAN)	4.76
O14561	Acyl carrier protein (ACPM_HUMAN)	3.73
O95994	Anterior gradient protein 2 homolog (AGR2_HUMAN)	20.18
O95831	Apoptosis-inducing factor 1 (AIFM1_HUMAN)	2.83
Q9HDC9	Adipocyte plasma membrane-associated protein (APMAP_HUMAN)	2.50
P53004	BIEA_HUMAN	2.40
Q8WY22	BRI3-binding protein (BRI3B_HUMAN)	3.77
Q9BRJ6	CG050_HUMAN	5.56
O75390	Citrate synthase (CISY_HUMAN)	3.40
Q96DG6	Carboxymethylenebutenolidase homolog (CMBL_HUMAN)	7.66
Q9Y2B0	Protein canopy homolog (CNPY2_HUMAN)	2.38
Q9GZN8	UPF0687 protein C20orf27 (CT027_HUMAN)	2.81
P00167	Cytochrome b5 (CYB5_HUMAN)	5.87
Q9H773	dCTP pyrophosphatase 1 (DCTP1_HUMAN)	3.26
Q96HY6	DDRGK domain-containing protein 1 (DDRGK_HUMAN)	2.51
Q3LXA3	FAD-AMP lyase (DHAK_HUMAN)	6.33
Q9UBM7	7-dehydrocholesterol reductase (DHCR7_HUMAN)	5.67
Q13268	Dehydrogenase/reductase SDR family member 2 (DHRS2_HUMAN)	34.28
Q9BTZ2	Dehydrogenase/reductase SDR family member 4 (DHRS4_HUMAN)	5.97
P13804	Electron transfer flavoprotein subunit alpha (ETFA_HUMAN)	3.94
P09467	Fructose-1,6-bisphosphatase 1 (F16P1_HUMAN)	56.31
P49327	Fatty acid synthase (FAS_HUMAN)	4.09
Q9NYL4	FK506-binding protein 11 (FKB11_HUMAN)	32.96
Q02790	FK506-binding protein 4 (FKBP4_HUMAN)	2.50
P11413	Glucose-6-phosphate 1-dehydrogenase (G6PD_HUMAN)	3.58
P80404	4-aminobutyrate aminotransferase (GABT_HUMAN)	129.84
Q8NCL4	Polypeptide N-acetylgalactosaminyltransferase 6 (GALT6_HUMAN)	9.79
O76003	Glutaredoxin-3 (GLRX3_HUMAN)	5.02
P43304	Glycerol-3-phosphate dehydrogenase	3.53

	(GPDM_HUMAN)	
P62993	Growth factor receptor-bound protein 2 (GRB2_HUMAN)	2.51
P21266	Glutathione S-transferase Mu 3 (GSTM3_HUMAN)	13.23
P16401	Histone H1.5 (H15_HUMAN)	24.98
Q9Y5Z4	Heme-binding protein 2 (HEBP2_HUMAN)	2.65
Q01581	Hydroxymethylglutaryl-CoA synthase (HMCS1_HUMAN)	4.77
P04792	Heat shock protein beta-1 (HSPB1_HUMAN)	7.90
P48735	Isocitrate dehydrogenase [NADP] (IDHP_HUMAN)	5.48
Q8TEX9	Importin-4 (IPO4_HUMAN)	2.04
Q96CN7	Isochorismatase domain-containing protein 1 (ISOC1_HUMAN)	5.86
Q6UXG2	UPF0577 protein KIAA1324 (K1324_HUMAN)	13.13
P12532	Creatine kinase (KCRU_HUMAN)	18.85
Q16719	Kynureninase (KYNU_HUMAN)	2.92
Q6P1M3	Lethal(2) giant larvae protein homolog 2 (L2GL2_HUMAN)	14.63
P17931	Galectin-3 (LEG3_HUMAN)	3.45
P50851	Lipopolysaccharide-responsive and beige-like anchor protein (LRBA_HUMAN)	3.03
Q9BS40	Latexin (LXN_HUMAN)	36.71
Q9HCC0	Methylcrotonoyl-CoA carboxylase beta chain (MCCB_HUMAN)	5.05
P27361	Mitogen-activated protein kinase 3 (MK03_HUMAN)	2,64
P49006	MARCKS-related protein (MRP_HUMAN)	3.52
Q9UM54	Myosin-VI (MYO6_HUMAN)	2.87
O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1 (NHRF1_HUMAN)	14.40
P15559	NAD(P)H dehydrogenase [quinone] 1 (NQO1_HUMAN)	17.34
P04181	Ornithine aminotransferase (OAT_HUMAN)	2.13
P32322	Pyrroline-5-carboxylate reductase 1 (P5CR1_HUMAN)	3.10
Q53H96	Pyrroline-5-carboxylate reductase 3 (P5CR3_HUMAN)	1.94
O95340	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthetase 2 (PAPS2_HUMAN)	2.36
Q16822	Phosphoenolpyruvate carboxykinase [GTP] (PCKGM_HUMAN)	4.29
Q9Y365	PCTP-like protein (PCTL_HUMAN)	12.24
Q15084	Protein disulfide-isomerase A6 (PDIA6_HUMAN)	2.66
P30086	Phosphatidylethanolamine-binding protein 1 (PEBP1_HUMAN)	3.33
Q8TBX8	Phosphatidylinositol-5-phosphate 4-kinase type-2 gamma (PI42C_HUMAN)	3.93
P14923	Junction plakoglobin, catenin gamma (PLAK_HUMAN)	6.38

Q9NVS9	Pyridoxine-5'-phosphate oxidase (PNPO_HUMAN)	3.97
P50897	Palmitoyl-protein thioesterase 1 (PPT1_HUMAN)	3.52
Q06830	Peroxiredoxin-1 (PRDX1_HUMAN)	2.37
P32119	Peroxiredoxin-2 (PRDX2_HUMAN)	2.87
P29762	Cellular retinoic acid-binding protein 1 (RABP1_HUMAN)	5.41
P29373	Cellular retinoic acid-binding protein 2 (RABP2_HUMAN)	21.09
Q00765	Receptor expression-enhancing protein 5 (REEP5_HUMAN)	2.66
P05109	Protein S100-A8 (S10A8_HUMAN)	15.20
P06702	Protein S100-A9 (S10A9_HUMAN)	23.37
Q9HCY8	Protein S100-A14 (S10AE_HUMAN)	35.58
P55809	Succinyl-CoA:3-ketoacid-coenzyme A transferase 1 (SCOT1_HUMAN)	5.23
Q8NBX0	Probable saccharopine dehydrogenase (SCPDH_HUMAN)	3.27
O43175	D-3-phosphoglycerate dehydrogenase (SERA_HUMAN)	5.64
O75368	SH3 domain-binding glutamic acid-rich-like protein (SH3L1_HUMAN)	3.67
Q9UNE7	STIP1 homology and U box-containing protein 1 (STUB1_HUMAN)	5.80
Q9NSD9	Phenylalanyl-tRNA synthetase beta chain (SYFB_HUMAN)	2.39
O76070	Gamma-synuclein (SYUG_HUMAN)	10.46
Q16762	Thiosulfate sulfurtransferase (THTR_HUMAN)	2.39
Q6UW68	Transmembrane protein 205 (TM205_HUMAN)	6.52
Q9Y3Q3	Transmembrane emp24 domain-containing protein 3 (TMED3_HUMAN)	11.35
Q12931	Tumour necrosis factor type 1 receptor-associated protein (TRAP1_HUMAN)	4.17
O43396	Thioredoxin-like protein 1 (TXNL1_HUMAN)	2.37
O60701	UDP-glucose 6-dehydrogenase (UGDH_HUMAN)	5.83
O95292	Vesicle-associated membrane protein-associated protein B/C (VAPB_HUMAN)	2.68
	Down-Regulated Proteins	
P01892	HLA class I histocompatibility antigen, A-2 alpha chain (1A02_HUMAN)	0.075
P21589	5'-nucleotidase (5NTD_HUMAN)	0.067
O95336	6-phosphogluconolactonase (6PGL_HUMAN)	0.36
P21399	Cytoplasmic aconitate hydratase (ACOC_HUMAN)	0.23
P12814	Alpha-actinin-1 (ACTN1_HUMAN)	0.24
P35609	Alpha-actinin-2 (ACTN2_HUMAN)	0.25
O43707	Alpha-actinin-4 (ACTN4_HUMAN)	0.27

Q09666	Neuroblast differentiation-associated protein AHNAK (AHNK_HUMAN)	0.49
Q16352	Alpha-internexin (AINX_HUMAN)	0.050
P15121	Aldose reductase (ALDR_HUMAN)	0.054
P04083	Annexin A1 (ANXA1_HUMAN)	0.026
P07355	Annexin A2 (ANXA2_HUMAN)	0.12
P12429	Annexin A3 (ANXA3_HUMAN)	0.19
P09525	Annexin A4 (ANXA4_HUMAN)	0.28
P08758	Annexin A5 (ANXA5_HUMAN)	0.28
P08133	Annexin A6 (ANXA6_HUMAN)	0.13
Q96CW1	AP-2 complex subunit mu-1 (AP2M1_HUMAN)	0.45
Q06481	Amyloid-like protein 2 (APLP2_HUMAN)	0.20
P61160	Actin-related protein 2 (ARP2_HUMAN)	0.38
O15144	Actin-related protein 2/3 complex subunit 2 (ARPC2_HUMAN)	0.49
P35613	Tumour cell-derived collagenase stimulatory factor (BASI_HUMAN)	0.22
Q9Y376	Calcium-binding protein 39 (CAB39_HUMAN)	0.053
Q05682	Caldesmon (CALD1_HUMAN)	0.038
O43852	Calumenin (CALU_HUMAN)	0.23
P17655	Calpain-2 catalytic subunit (CAN2_HUMAN)	0.14
Q01518	Adenylyl cyclase-associated protein 1 (CAP1_HUMAN)	0.48
P40121	Macrophage-capping protein (CAPG_HUMAN)	0.051
P48509	CD151 antigen (CD151_HUMAN)	0.18
Q9Y5K6	CD2-associated protein (CD2AP_HUMAN)	0.34
P16070	CD44 antigen (CD44_HUMAN)	0.068
P60033	CD81 antigen (CD81_HUMAN)	0.50
P48960	CD97 antigen (CD97_HUMAN)	0.049
O00299	Chloride intracellular channel protein 1 (CLIC1_HUMAN)	0.49
Q9Y696	Chloride intracellular channel protein 4 (CLIC4_HUMAN)	0.24
P09543	2',3'-cyclic-nucleotide 3'-phosphodiesterase (CN37_HUMAN)	0.21
Q15021	Condensin complex subunit 1 (CND1_HUMAN)	0.34
Q15417	Calponin-3 (CNN3_HUMAN)	0.099
Q9ULV4	Coronin-1C (COR1C_HUMAN)	0.21
Q14019	Coactosin-like protein (COTL1_HUMAN)	0.11
Q99829	Copine-1 (CPNE1_HUMAN)	0.16
P04632	Calpain small subunit 1 (CPNS1_HUMAN)	0.23
O14578	Citron Rho-interacting kinase (CTRO_HUMAN)	0.045
P17661	Desmin (DESM_HUMAN)	0.054
Q9Y2H0	Disks large-associated protein 4 (DLGP4_HUMAN)	0.25

Q14204	Cytoplasmic dynein 1 heavy chain 1 (DYHC1_HUMAN)	0.53
Q9BUP0	EF-hand domain-containing protein D1 (EFHD1_HUMAN)	0.029
Q96C19	EF-hand domain-containing protein D2 (EFHD2_HUMAN)	0.10
Q9H4M9	EH domain-containing protein 1 (EHD1_HUMAN)	0.21
Q9NZN4	EH domain-containing protein 2 (EHD2_HUMAN)	0.022
Q7L2H7	Eukaryotic translation initiation factor 3 subunit M (EIF3M_HUMAN)	0.44
P29317	Ephrin type-A receptor 2 (EPHA2_HUMAN)	0.035
O75477	Erlin-1 (ERLN1_HUMAN)	0.21
A0FGR8	Extended synaptotagmin-2 (ESYT2_HUMAN)	0.33
O75369	Filamin-B (FLNB_HUMAN)	0.31
Q14315	Filamin-C (FLNC_HUMAN)	0.13
P51114	Fragile X mental retardation syndrome-related protein 1 (FXR1_HUMAN)	0.26
Q13283	Ras GTPase-activating protein-binding protein 1 (G3BP1_HUMAN)	0.30
Q14376	UDP-glucose 4-epimerase (GALE_HUMAN)	0.31
Q9UBI6	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12 (GBG12_HUMAN)	0.17
Q9H0R5	Guanylate-binding protein 3 (GBP3_HUMAN)	0.023
P52566	Rho GDP-dissociation inhibitor 2 (GDIR2_HUMAN)	0.018
Q04446	1,4-alpha-glucan-branching enzyme (GLGB_HUMAN)	0.46
P04899	Guanine nucleotide-binding protein G(i), alpha-2 subunit (GNAI2_HUMAN)	0.22
Q9NZ01	Synaptic glycoprotein SC2 (GPSN2_HUMAN)	0.33
Q8NBJ5	Glycosyltransferase 25 family member 1 (GT251_HUMAN)	0.23
P11166	Solute carrier family 2, facilitated glucose transporter member 1 (GTR1_HUMAN)	0.14
P17096	High mobility group protein HMG-I/HMG-Y (HMGA1_HUMAN)	0.19
O14979	Heterogeneous nuclear ribonucleoprotein D-like (HNRDL_HUMAN)	0.45
Q6YN16	Hydroxysteroid dehydrogenase-like protein 2 (HSDL2_HUMAN)	0.29
P56537	Eukaryotic translation initiation factor 6 (IF6_HUMAN)	0.36
O00458	Interferon-related developmental regulator 1 (IFRD1_HUMAN)	0.23
Q13418	Integrin-linked protein kinase (ILK_HUMAN)	0.20
O00410	Importin-5 (IPO5_HUMAN)	0.27

P17301	Integrin alpha-2 (ITA2_HUMAN)	0.068
P05556	Integrin beta-1 (ITB1_HUMAN)	0.078
P08729	Keratin, type II cytoskeletal 7 (K2C7_HUMAN)	0.094
P33176	Kinesin-1 heavy chain (KINH_HUMAN)	0.43
Q14847	LIM and SH3 domain protein 1 (LASP1_HUMAN)	0.31
P07195	L-lactate dehydrogenase B chain (LDHB_HUMAN)	0.079
P09382	Galectin-1 (LEG1_HUMAN)	0.15
P02545	Lamin-A/C (LMNA_HUMAN)	0.47
P29966	Myristoylated alanine-rich C-kinase substrate (MARCS_HUMAN)	0.46
Q15691	Microtubule-associated protein RP/EB family member 1 (MARE1_HUMAN)	0.28
P19105	Myosin regulatory light chain 12A (ML12A_HUMAN)	0.086
P26038	Moesin (MOES_HUMAN)	0.13
O15427	Monocarboxylate transporter 4 (MOT4_HUMAN)	0.096
Q6WCQ1	Myosin phosphatase Rho-interacting protein (MPRIP_HUMAN)	0.087
P35580	Myosin-10 (MYH10_HUMAN)	0.11
P35749	Myosin-11 (MYH11_HUMAN)	0.11
P35579	Myosin-9 (MYH9_HUMAN)	0.039
P08590	Myosin light chain 3 (MYL3_HUMAN)	0.24
P60660	Myosin light polypeptide 6 (MYL6_HUMAN)	0.13
P14649	Myosin light chain 6B (MYL6B_HUMAN)	0.23
P24844	Myosin regulatory light polypeptide 9 (MYL9_HUMAN)	0.042
Q9NZM1	Myoferlin (MYOF_HUMAN)	0.38
P43490	Nicotinamide phosphoribosyltransferase (NAMPT_HUMAN)	0.33
P07196	Neurofilament light polypeptide (NFL_HUMAN)	0.048
Q9UMS0	NFU1 iron-sulfur cluster scaffold homolog (NFU1_HUMAN)	0.17
Q9BZQ8	Protein Niban (NIBAN_HUMAN)	0.16
P30419	Glycylpeptide N-tetradecanoyltransferase 1 (NMT1_HUMAN)	0.49
O95747	Serine/threonine-protein kinase OSR1 (OXSR1_HUMAN)	0.33
Q9NVD7	Alpha-parvin (PARVA_HUMAN)	0.16
Q96HC4	PDZ and LIM domain protein 5 (PDLI5_HUMAN)	0.20
Q9NR12	PDZ and LIM domain protein 7 (PDLI7_HUMAN)	0.060
P41219	Peripherin (PERI_HUMAN)	0.049
Q15149	Plectin-1 (PLEC1_HUMAN)	0.15
P13796	Plastin-2 (PLSL_HUMAN)	0.039
P13797	Plastin-3 (PLST_HUMAN)	0.045
P07737	Profilin-1 (PROF1_HUMAN)	0.47

Q6NZI2	Polymerase I and transcript release factor (PTRF_HUMAN)	0.029
P06737	Glycogen phosphorylase, liver form (PYGL_HUMAN)	0.13
P46060	Ran GTPase-activating protein 1 (RAGP1_HUMAN)	0.40
P13489	Ribonuclease inhibitor (RINI_HUMAN)	0.43
Q9NQC3	Reticulon-4 (RTN4_HUMAN)	0.16
P06703	Protein S100-A6 (S10A6_HUMAN)	0.13
P60903	Protein S100-A10 (S10AA_HUMAN)	0.11
Q99808	Equilibrative nucleoside transporter 1 (S29A1_HUMAN)	0.27
Q9Y617	Phosphoserine aminotransferase (SERC_HUMAN)	0.024
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3 (SH3L3_HUMAN)	0.27
Q9H2G2	STE20-like serine/threonine-protein kinase,CTCL tumour antigen se20-9 (SLK_HUMAN)	0.27
Q9NS25	Sperm protein associated with the nucleus on the X chromosome B/F (SPNXB_HUMAN)	0.0088
Q13813	Spectrin alpha chain (SPTA2_HUMAN)	0.27
Q01082	Spectrin beta chain, brain 1 (SPTB2_HUMAN)	0.19
Q13501	Sequestosome-1 (SQSTM_HUMAN)	0.33
Q9UJZ1	Stomatin-like protein 2 (STML2_HUMAN)	0.46
P16949	Stathmin (STMN1_HUMAN)	0.48
Q01995	Transgelin (TAGL_HUMAN)	0.063
P68366	Tubulin alpha-4A chain (TBA4A_HUMAN)	0.19
P13726	Tissue factor (TF_HUMAN)	0.082
P02786	Transferrin receptor protein 1 (TFR1_HUMAN)	0.37
P10599	Thioredoxin (THIO_HUMAN)	0.30
Q9Y490	Talin-1 (TLN1_HUMAN)	0.20
Q8NFQ8	Torsin-1A-interacting protein 2 (TOIP2_HUMAN)	0.41
P06753	Tropomyosin alpha-3 chain (TPM3_HUMAN)	0.33
P67936	Tropomyosin alpha-4 chain (TPM4_HUMAN)	0.34
Q6IBS0	Twinfilin-2 (TWF2_HUMAN)	0.29
P04818	Thymidylate synthase (TYSY_HUMAN)	0.16
Q8WVY7	Ubiquitin-like domain-containing CTD phosphatase 1 (UBCP1_HUMAN)	0.36
O00762	Ubiquitin-conjugating enzyme E2 C (UBE2C_HUMAN)	0.21
Q9NPD8	Ubiquitin-conjugating enzyme E2 T (UBE2T_HUMAN)	0.19
P05161	Interferon-induced 17 kDa protein (UCRP_HUMAN)	0.15
Q99536	Synaptic vesicle membrane protein VAT-1 homolog (VAT1_HUMAN)	0.32
P08670	Vimentin (VIME_HUMAN)	0.019
P18206	Vinculin (VINC_HUMAN)	0.30
O75083	WD repeat-containing protein 1 (WDR1_HUMAN)	0.47
Q15942	Zyxin (ZYX_HUMAN)	0.19

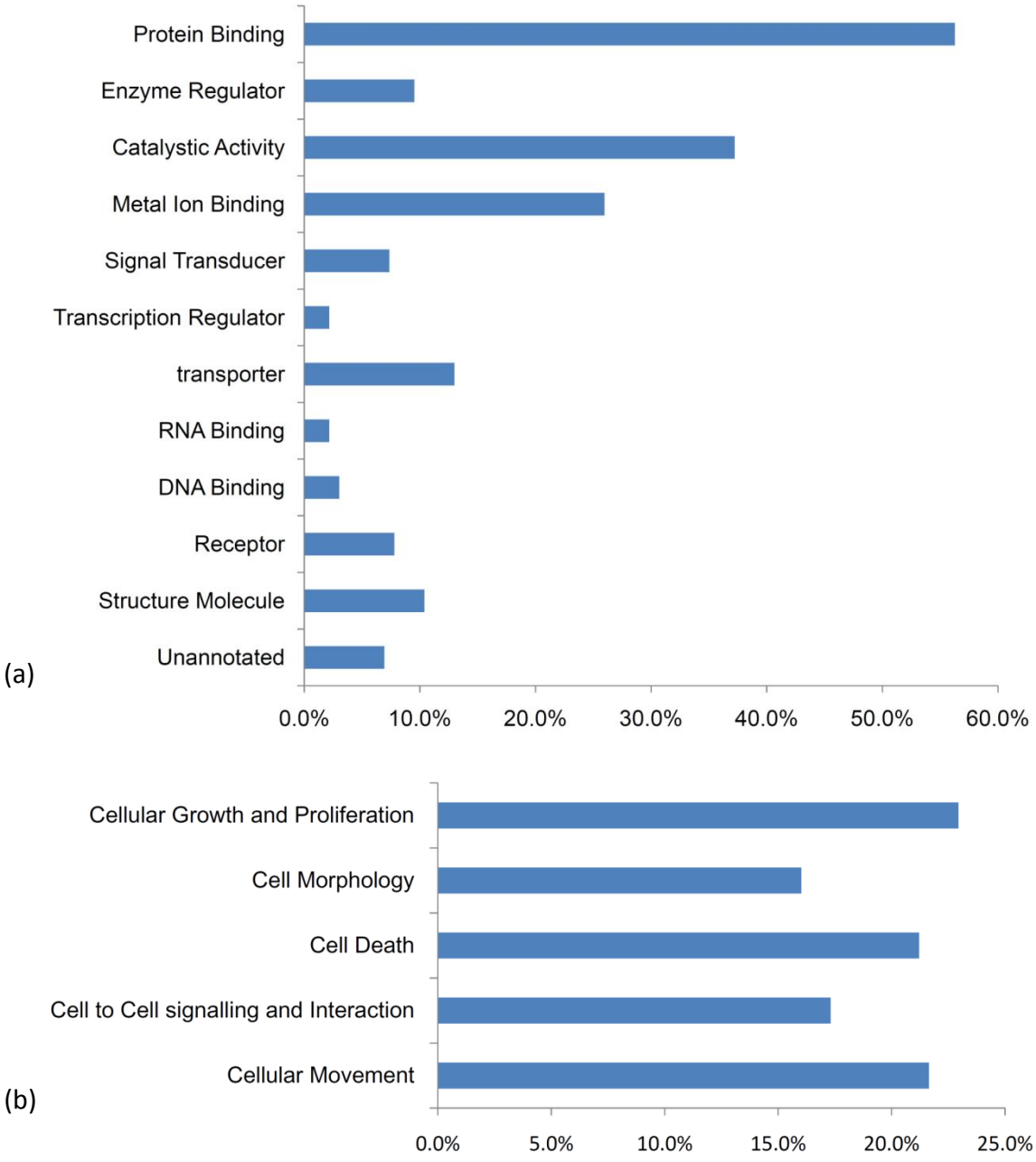


Figure 4.2. (a) Classification of 231 proteins into different groups based on their molecular functions. (b) Classification of 231 proteins based on the intracellular signaling pathways involved.

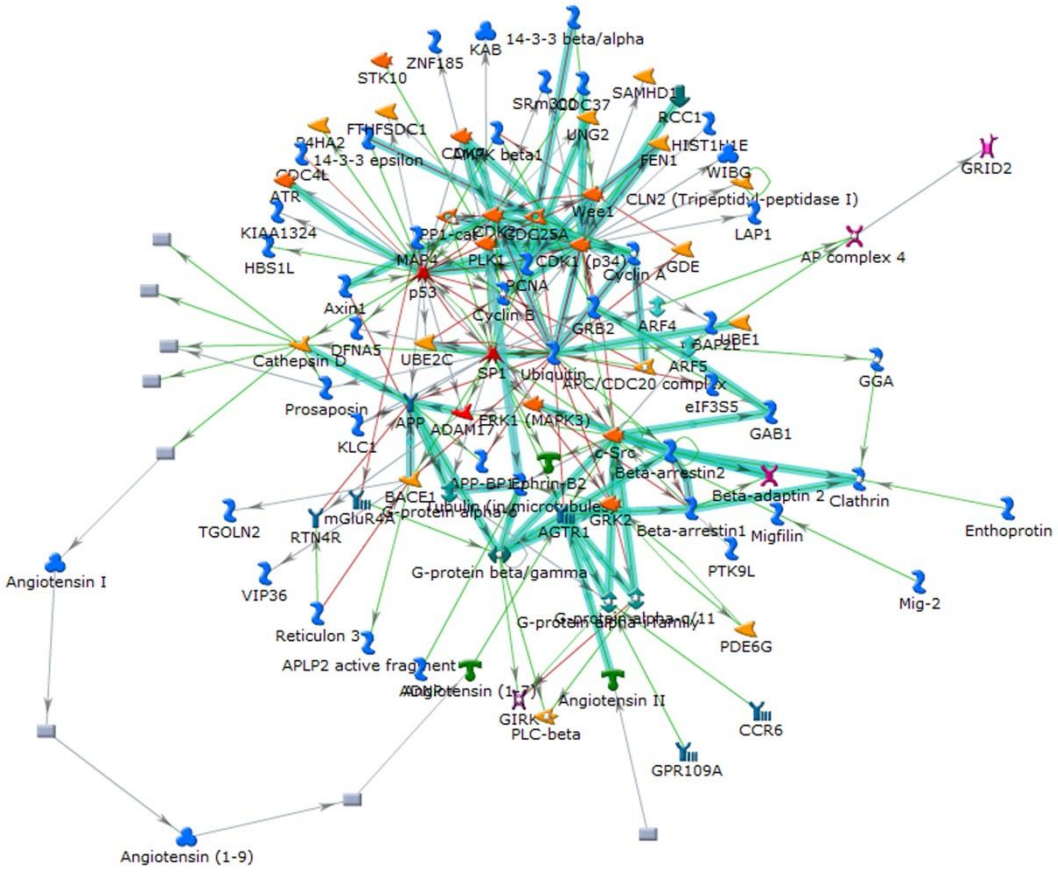


Figure 4.3. A partial representative of the intracellular signaling pathways in which the 231 identified proteins are involved.

Twenty four proteins identified by the above bioinformatic means were selected based on their involvement in cancer-related cellular signaling pathway and their potential clinical relevance. Next, proteins were searched in the Human Protein Atlas (HPA) Database[20-22]; the proteins that had different expression levels in normal and malignant breast tissue and also had been detected in the MCF-7 cell lines were filtered out. These analyses generated a candidate biomarker list of AIFM1, GSTM3, NIBAN, RAGP1 and SQSTM, which were uploaded into MetaCore to build biological networks and to find their potential interacting proteins.

4.3.3 Evaluation of Selected Protein Biomarkers

The differential protein abundance of the 5 biomarker candidates, AIFM1, GSTM3, NIBAN, RAGP1 and SQSTM, in wild type and HER2 overexpressing MCF-7 cells were validated by western blot (Figure 4.4). In HER2 overexpressing MCF-7 cells, AIFM1 and GSTM3 were found to be significantly upregulated and NIBAN, RAGP1 and SQSTM downregulated, consistent with the MS quantification data. The confirmation of these five biomarker candidates by western blot analysis increases the confidence in MS results. After western blot validation, the five proteins' expression levels in primary breast cancer tissue biopsies (treatment naïve) were examined by immunohistochemistry (IHC) on tissue microarray to further evaluate their relationship to clinically determined HER2 status.

The mean and maximal protein abundance were assessed by immunohistochemical analysis of clinical samples, in which HER2 normal and HER2 positive tumors were compared. AIFM1 and GSTM3 expression were positively correlated with HER2 status, while NIBAN and RAPG1 displayed an inverse trend with HER2 positive. These results from breast cancer clinical samples were consistent with the protein abundance detected in wild type and HER2 overexpressing MCF-7 cells. In contrast, SQSTM which showed downregulation in HER2 expressing cells was increased in the tumor biopsies (Figure 4.5). The p-value of the five biomarker candidates in HER2 normal and HER2 positive tumor biopsies are presented in Table

4.3.

Overall, the results obtained from biopsies were generally consistent with the findings from cell line experiments, i.e., the differential expression of proteins and their status to HER2 expression levels were in the same direction and/or magnitude between in vitro (cell lines) and the primary breast cancer tissues. This validates the feasibility of the proteomics based approach described here for identification of potential novel markers for therapeutics development.

Table 4.3. P-values of the five biomarker candidates' TMA scores.

Gene Cut_Point	Her2 Negative	Her2 Positive	Chi_square P value	Fisher's exact P value
AIFM1_Mean<1.25	25	8	0.105	0.133
AIFM1_Mean>=1.25	49	33	0.105	0.133
AIFM1_Max<2.00	20	8	0.368	0.497
AIFM1_Max>=2.00	54	33	0.368	0.497
GSTM3_Mean<2.25	50	25	0.306	0.404
GSTM3_Mean>=2.25	21	16	0.306	0.404
GSTM3_Max<3.00	43	24	0.833	0.844
GSTM3_Max>=3.00	28	17	0.833	0.844
NIBAN_CP_Mean<2.25	58	39	0.057	0.066
NIBAN_CP_Mean>=2.25	15	3	0.057	0.066
NIBAN_CP_Max<3.00	54	38	0.033	0.051
NIBAN_CP_Max>=3.00	19	4	0.033	0.051
RagP1_Mean<1.67	65	31	0.182	0.211
RagP1_Mean>=1.67	11	10	0.182	0.211
RagP1_Max<1.00	24	5	0.021	0.025
RagP1_Max>=1.00	52	36	0.021	0.025
SQSTM1_Mean<1.67	54	23	0.059	0.070
SQSTM1_Mean>=1.67	21	19	0.059	0.070
SQSTM1_Max<1.00	24	6	0.035	0.047
SQSTM1_Max>=1.00	51	36	0.035	0.047

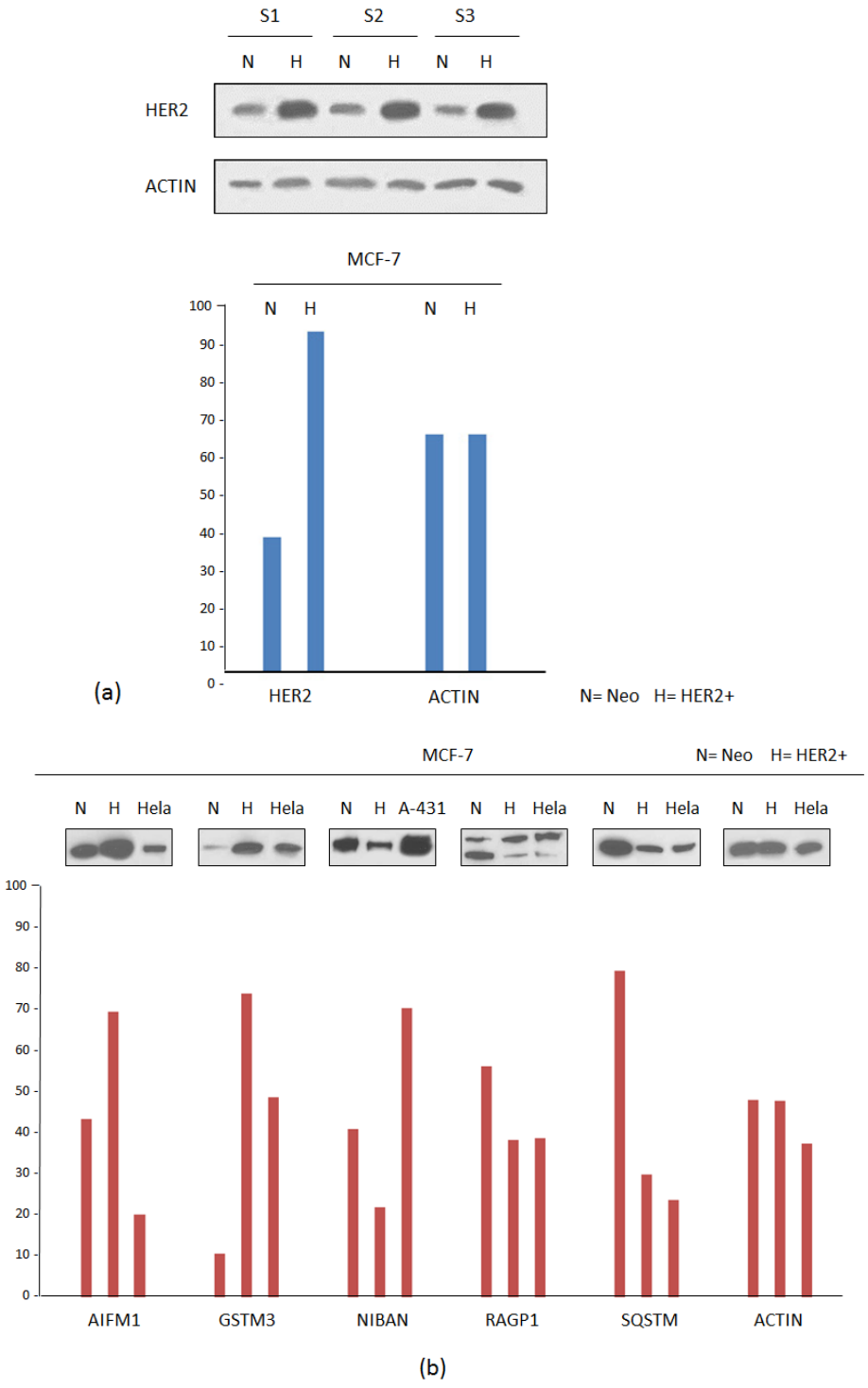


Figure 4.4: (a) HER2 protein levels in vector transfected (Neo) and HER2 overexpressing (HER2+) MCF-7 human breast cancer cells. Total cellular proteome from bio-triplicate Neo- or HER2^{+}- MCF-7 cell extracts were separated by SDS-PAGE and processed for immunoblotting using HER2 antibody. Western blots using actin antibody were performed to confirm equal loadings. (b) Western blot of the differentially expressed 5 biomarker candidates, AIFM1, GSTM3, NIBAN, RAGP1, SQSTM in Neo- and HER2^{+}- MCF-7 cells. HeLa and A-431 cell extracts were used as positive controls for various antibodies and actin was used as the loading control.

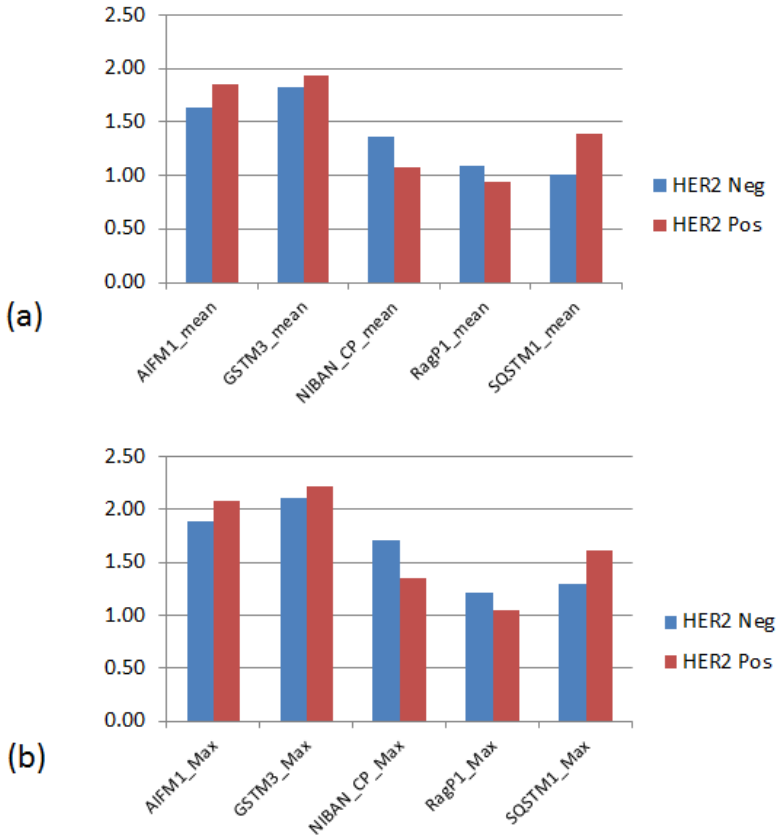


Figure 4.5. (a) Mean values of 5 protein biomarker candidates' TMA mean score with 75 HER2 negative and 42 HER2 positive human breast cancer biopsies. (b) Mean values of 5 protein biomarker candidates' TMA max score with 75 HER2 negative and 42 HER2 positive human breast cancer biopsies.

After biological validations, further bioinformatic studies were done with the five proteins to understand the biomarkers candidates' biofunctions and to explore potential links between their differential expression and HER2's overexpression in human breast cancer cells.

RAGP1 is the GTPase-activating protein (GAP), which specifically induces the activation of RAN, a GTPase involved nucleocytoplamic transportation[23]. RAGP1 is also the first documented substrate for the ubiquitin-like protein SUMO-1[24-25], that could explain the double band of RAGP1 observed in western blotting: presumably, the upper band represents SUMO-RAGP1 while the lower band is the unmodified RAGP1[26]. It's reported that both RAGP1 and SUMO-RAGP1 are active in modulating RAN[25] activity, thereby regulating the bidirectional transport of proteins and ribonucleoproteins across the nuclear pore complex (NPC)[27] and ultimately regulating transcription and cell cycle progression[28]. From MS based quantification, RAGP1 was downregulated in the HER2 overexpressing MCF-7 cells. This overall decrease in RAGP1 expression was also supported by western blot and immunohistochemical assays of HER2 positive breast tumor biopsies. This observation suggests a mechanical linkage between the downregulation of RAGP1 and the poor chemotherapeutic prognosis of HER2 positive breast cancers[29].

GSTM3, glutathione-S-transferase M3 is an enzyme that catalyzes the conjugation of reduced glutathione to electrophilic centers on a wide variety of substrates via a sulfhydryl group. This activity detoxifies endogenous compounds, such as environmental toxins, carcinogens and therapeutic drugs. Consistent with our results, this enzyme was reported to be upregulated in breast cancer cells upon stimulation of the MEK/ERK pathway[30]. Due to its role in inactivation of cytotoxic chemotherapeutics, its increased protein levels may contribute to the anti-apoptotic effects of HER2 overexpression in breast cancer and their poor outcomes following chemotherapy [31].

SQSTM, sequestosome or p62 is an adapter protein that binds ubiquitin

noncovalently. Studies employing knockout, transgenic, and knockin mice have shown that p62 plays critical roles in a number of cellular functions, including bone remodeling, obesity, and cancer[32-35]. This protein functions as a nodal point in regulating NF-kappaB signaling pathway and cell differentiation[36-37], its suppression would induce apoptosis in cells[38]. In our MS quantification experiment and western blot studies, the expression of SQSTM in HER2 overexpressing cells was downregulated; however, immunohistochemical analysis of tumor biopsies indicated upregulation of SQSTM in HER2 positive cases, consistent with their anti-apoptotic phenotype. This discrepancy may be due to the microenvironment differences between cells lines (in vitro) and tissue (in vivo biopsies).

AIFM1, apoptosis-inducing factor, mitochondrion associated 1 (also abbreviated as AIF), is a flavoprotein, which is involved in initiating caspase-independent pathway of apoptosis[39-40]. AIF negative tumors have poorer chemotherapeutic prognosis and survival ratio[41-43]. Increased expression of AIF in HER2 overexpressing human breast cancer cells is suggestive of apoptosis activation. However, the overall behavior of HER2 overexpressing MCF-7 cells is the opposite. This seemingly contradictory situation was also observed for NIBAN.

NIBAN also known as FAM 129A is thought to play a role in cell apoptosis regulation by binding to nucleophosmin (NPM)[44]. Interactions between NIBAN and NPM lead to the disassociation of NPM from the MDM2 complex, promoting MDM2-p53 interactions and subsequent p53 degradation, therefore providing an anti-apoptotic effect[44]. Like the upregulation of AIFM1, the downregulation of NIBAN in our HER2 overexpressing MCF-7 cells and breast cancer tissue should trigger cell apoptosis. However, HER2 positive MCF-7 cells are reportedly anti-apoptotic[45-46]. The possible explanation for the dysfunction of AIFM1 and NIBAN in HER2 overexpressing MCF-7 cells and tumor biopsies may be the expression of an anti-apoptotic factor(s) that override the proapoptotic effects of these proteins. To address this possibility, we further explored our proteomic data, and identified NQO1 (NAD(P)H dehydrogenase, quinone 1), a well-known apoptosis regulating protein. NQO1 is an

important factor in nitric oxide (NO) production and NO is involved in many cellular pathways including apoptosis[47-48]. Based on MS quantification, NQO1 was significantly amplified in HER2 overexpressing MCF-7 cells and this upregulation inhibits NO production[49] and potentially apoptosis in HER2 overexpressing MCF-7 cells[50]. AIFM1 induced apoptosis is caspase-independent and NO also regulates cell apoptosis via both caspase-dependent and independent pathways[51]. It is possible that the concentration changes of NO in cell would counteract AIFM1 induced apoptosis. In the case of NIBAN, which regulates apoptosis through p53 pathway, NO could overcome the proapoptotic effects of NIBAN downregulation by directly regulating the downstream MDM2 in the p53 pathway[52].

After the bioinformatics studies, we also tried to evaluate the potential prognostic value of the four protein biomarkers in the 371 breast cancer biopsies with available clinicopathological data. Overall survival analysis showed no statistically significant correlations between the five differentially expressed proteins and breast cancer development (Supplemental Figure S4.1). However, it is noteworthy that most of the patients in the cohorts described received adjuvant chemotherapy. Therefore, these results are not surprising and even reassuring given the context of the aggressive adjuvant treatment provided to the clinical population; HER2 positive patients received adjuvant trastuzumab in combination with anthracycline and taxane chemotherapy, which have been shown to confer a prognosis equivalent to HER2 negative disease[53]. Another limitation in our analysis was the sample size. Biomarkers for prognostication and tumor subtype classifications in published reports are typically from thousands of samples. Here, we only had access to 371 cases. Considering the different subtypes within this group and the anti-cancer treatments they received, the sample size for each subtype were even less, making it less likely to give statistically significant results. Currently, we are planning to assess the prognostic values of the five proteins in appropriate cohorts with reasonable sample size.

4.4 Conclusions

We used 2-MEGA labeling technique to isotopically label bio-triplicate proteome samples from wild type and HER2 overexpressing MCF-7 cell lines. Following offline 2D-LC separation, the labeled samples were quantified them on ESI-QTOF. In order to get better quantification results, the MS and MS/MS conditions were optimized. From these experiments, 2455 unique proteins were quantified for the bio-triplicate samples, 1278 proteins were identified to be differentially expressed, with 231 proteins exhibiting the same expression level changes in the bio-triplicates of the two clones. The multiplicity of proteins differentially expressed in these isogenic cells lines as a result of transfection of additional copies of the HER2 gene suggests that the biological effects of HER2 are mediated by multiple pathways and various effector molecules. Of the 231 proteins, five proteins were selected and validated by western blotting and immunohistochemical analysis. The direction of change in protein levels observed in MCF-7 cell lines was recapitulated in clinical samples for four of the five proteins. Although these proteins' prognostic value haven't been fully validated in this study because of limited sample size and aggressive adjuvant anti-HER2 therapy treated cohorts, identification of the proteins which potentially mediate the biological effects of HER2-positive phenotype may provide a better understanding of the mediators of clinically aggressive disease and potential mechanisms for resistance to anti-HER2 therapies. These mediators may have the potential to be used as biomarkers for diagnosis, prognosis and development of effective therapeutic strategies for treatment of HER2 mediated metastatic breast carcinomas.

4.5 Acknowledgment

This work was funded by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs program (LL and YT), Alberta Cancer Foundation and Alberta Breast Cancer Research Initiative (JM and SD) and Canadian

Breast Cancer Foundation-Prairies/NWT Region (JM, SD and MP). We thank Dr. Dennis Slamon, University of California Los Angeles, for kindly sharing the cell lines used in this study. Funding support for CBCF-Tumor Bank is from Canadian Breast Cancer Foundation-Prairies/NWT Region, Alberta Cancer Foundation, Alberta Cancer Prevention and Legacy Fund managed by Alberta Innovates-Health Solutions.

4.6 Literature Cited

1. Hicks, D. G.; Kulkarni, S., HER2+ breast cancer: review of biologic relevance and optimal use of diagnostic tools. *Am J Clin Pathol* **2008,** *129* (2), 263-73.
2. Gast, M. C.; Schellens, J. H.; Beijnen, J. H., Clinical proteomics in breast cancer: a review. *Breast Cancer Res Treat* **2009,** *116* (1), 17-29.
3. Baselga, J., Treatment of HER2-overexpressing breast cancer. *Ann Oncol 21 Suppl 7*, vii36-40.
4. Nahta, R.; Esteva, F. J., HER2 therapy: molecular mechanisms of trastuzumab resistance. *Breast Cancer Res* **2006,** *8* (6), 215.
5. Schaub, N. P.; Jones, K. J.; Nyalwidhe, J. O.; Cazares, L. H.; Karbassi, I. D.; Semmes, O. J.; Feliberti, E. C.; Perry, R. R.; Drake, R. R., Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. *J Am Coll Surg* **2009,** *208* (5), 970-8; discussion 978-80.
6. Tuszynski, G. P.; Rothman, V. L.; Zheng, X.; Gutu, M.; Zhang, X.; Chang, F., G-protein coupled receptor-associated sorting protein 1 (GASP-1), a potential biomarker in breast cancer. *Exp Mol Pathol 91* (2), 608-13.
7. Kang, U. B.; Ahn, Y.; Lee, J. W.; Kim, Y. H.; Kim, J.; Yu, M. H.; Noh, D. Y.; Lee, C., Differential profiling of breast cancer plasma proteome by isotope-coded affinity tagging method reveals biotinidase as a breast cancer biomarker. *BMC Cancer 10*, 114.
8. Ji, C.; Guo, N.; Li, L., Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. *J Proteome Res* **2005,** *4* (6), 2099-108.
9. Wang, P.; Lo, A.; Young, J. B.; Song, J. H.; Lai, R.; Kneteman, N. M.; Hao, C.; Li, L.,

Targeted quantitative mass spectrometric identification of differentially expressed proteins between Bax-expressing and deficient colorectal carcinoma cells. *J Proteome Res* **2009,** *8* (7), 3403-14.

10. Pegram, M. D.; Finn, R. S.; Arzoo, K.; Beryt, M.; Pietras, R. J.; Slamon, D. J., The effect of HER-2/neu overexpression on chemotherapeutic drug sensitivity in human breast and ovarian cancer cells. *Oncogene* **1997,** *15* (5), 537-47.
11. Craik, A. C.; Veldhoen, R. A.; Czernick, M.; Buckland, T. W.; Kyselytzia, K.; Ghosh, S.; Lai, R.; Damaraju, S.; Underhill, D. A.; Mackey, J. R.; Goping, I. S., The BH3-only protein Bad confers breast cancer taxane sensitivity through a nonapoptotic mechanism. *Oncogene* **2010,** *29* (39), 5381-91.
12. Germain, D. R.; Graham, K.; Glubrecht, D. D.; Hugh, J. C.; Mackey, J. R.; Godbout, R., DEAD box 1: a novel and independent prognostic marker for early recurrence in breast cancer. *Breast Cancer Research and Treatment* **2011,** *127* (1), 53-63.
13. Pancholi, S.; Lykkesfeldt, A. E.; Hilmi, C.; Banerjee, S.; Leary, A.; Drury, S.; Johnston, S.; Dowsett, M.; Martin, L. A., ERBB2 influences the subcellular localization of the estrogen receptor in tamoxifen-resistant MCF-7 cells leading to the activation of AKT and RPS6KA2. *Endocr Relat Cancer* **2008,** *15* (4), 985-1002.
14. Vlajnic, T.; Andreozzi, M. C.; Schneider, S.; Tornillo, L.; Karamitopoulou, E.; Lugli, A.; Ruiz, C.; Zlobec, I.; Terracciano, L., VEGFA gene locus (6p12) amplification identifies a small but highly aggressive subgroup of colorectal patients. *Mod Pathol 24* (10), 1404-12.
15. Balacescu, L.; Balacescu, O.; Crisan, N.; Fetica, B.; Petrut, B.; Bungardean, C.; Rus, M.; Tudoran, O.; Meurice, G.; Irimie, A.; Dragos, N.; Berindan-Neagoe, I., Identifying molecular features for prostate cancer with Gleason 7 based on microarray gene expression profiles. *Rom J Morphol Embryol 52* (4), 1195-202.
16. Karve, T. M.; Li, X.; Saha, T., BRCA1-mediated signaling pathways in ovarian carcinogenesis. *Funct Integr Genomics*.
17. Varnum, S. M.; Webb-Robertson, B. J.; Hessol, N. A.; Smith, R. D.; Zangar, R. C., Plasma Biomarkers for Detecting Hodgkin's Lymphoma in HIV Patients. *Plos One*

6 (12), e29263.

18. Ekins, S.; Nikolsky, Y.; Bugrim, A.; Kirillov, E.; Nikolskaya, T., Pathway mapping tools for analysis of high content data. *Methods Mol Biol* **2007,** *356*, 319-50.
19. Rhodes, D. R.; Tomlins, S. A.; Varambally, S.; Mahavisno, V.; Barrette, T.; Kalyana-Sundaram, S.; Ghosh, D.; Pandey, A.; Chinnaiyan, A. M., Probabilistic model of the human protein-protein interaction network. *Nat Biotechnol* **2005,** *23* (8), 951-9.
20. Berglund, L.; Bjoerling, E.; Oksvold, P.; Fagerberg, L.; Asplund, A.; Szigyarto, C. A. K.; Persson, A.; Ottosson, J.; Wernerus, H.; Nilsson, P.; Lundberg, E.; Sivertsson, A.; Navani, S.; Wester, K.; Kampf, C.; Hober, S.; Ponten, F.; Uhlen, M., A Genecentric Human Protein Atlas for Expression Profiles Based on Antibodies. *Molecular & Cellular Proteomics* **2008,** *7* (10), 2019-2027.
21. Persson, A.; Hober, S.; Uhlen, M., A human protein atlas based on antibody proteomics. *Current Opinion in Molecular Therapeutics* **2006,** *8* (3), 185-190.
22. Uhlens, M.; Bjorling, E.; Agaton, C.; Szigyarto, C. A.; Amini, B.; Andersen, E.; Andersson, A. C.; Angelidou, P.; Asplund, A.; Asplund, C.; Berglund, L.; Bergstrom, K.; Brumer, H.; Cerjan, D.; Ekstrom, M.; Elobeid, A.; Eriksson, C.; Fagerberg, L.; Falk, R.; Fall, J.; Forsberg, M.; Bjorklund, M. G.; Gumbel, K.; Halimi, A.; Hallin, I.; Hamsten, C.; Hansson, M.; Hedhammar, M.; Hercules, G.; Kampf, C.; Larsson, K.; Linskog, M.; Lodewyckx, W.; Lund, J.; Lundeberg, J.; Magnusson, K.; Malm, E.; Nilsson, P.; Odling, J.; Oksvold, P.; Olsson, I.; Oster, E.; Ottosson, J.; Paavilainen, L.; Persson, A.; Rimini, R.; Rockberg, J.; Runeson, M.; Sivertsson, A.; Skollermo, A.; Steen, J.; Stenvall, M.; Sterky, F.; Stromberg, S.; Sundberg, M.; Tegel, H.; Tourle, S.; Wahlund, E.; Walden, A.; Wan, J. H.; Wernerus, H.; Westberg, J.; Wester, K.; Wrethagen, U.; Xu, L. L.; Hober, S.; Ponten, F., A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & Cellular Proteomics* **2005,** *4* (12), 1920-1932.
23. Bischoff, F. R.; Klebe, C.; Kretschmer, J.; Wittinghofer, A.; Ponstingl, H., RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc Natl Acad Sci U S A* **1994,** *91* (7), 2587-91.

24. Mahajan, R.; Delphin, C.; Guan, T.; Gerace, L.; Melchior, F., A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **1997,** *88* (1), 97-107.
25. Matunis, M. J.; Coutavas, E.; Blobel, G., A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol* **1996,** *135* (6 Pt 1), 1457-70.
26. Shitashige, M.; Satow, R.; Honda, K.; Ono, M.; Hirohashi, S.; Yamada, T., Regulation of Wnt signaling by the nuclear pore complex. *Gastroenterology* **2008,** *134* (7), 1961-71, 1971 e1-4.
27. Mahajan, R.; Gerace, L.; Melchior, F., Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. *J Cell Biol* **1998,** *140* (2), 259-70.
28. Sazer, S.; Dasso, M., The ran decathlon: multiple roles of Ran. *J Cell Sci* **2000,** *113 (Pt 7)*, 1111-8.
29. Oussenko, I. A.; Holland, J. F.; Reddy, E. P.; Ohnuma, T., Effect of ON 01910.Na, an anticancer mitotic inhibitor, on cell-cycle progression correlates with RanGAP1 hyperphosphorylation. *Cancer Res* **2011,** *71* (14), 4968-76.
30. Zhou, C.; Nitschke, A. M.; Xiong, W.; Zhang, Q.; Tang, Y.; Bloch, M.; Elliott, S.; Zhu, Y.; Bazzone, L.; Yu, D.; Weldon, C. B.; Schiff, R.; McLachlan, J. A.; Beckman, B. S.; Wiese, T. E.; Nephew, K. P.; Shan, B.; Burow, M. E.; Wang, G., Proteomic analysis of tumor necrosis factor-alpha resistant human breast cancer cells reveals a MEK5/Erk5-mediated epithelial-mesenchymal transition phenotype. *Breast Cancer Res* **2008,** *10* (6), R105.
31. Gennari, A.; Sormani, M. P.; Pronzato, P.; Puntoni, M.; Colozza, M.; Pfeffer, U.; Bruzzi, P., HER2 status and efficacy of adjuvant anthracyclines in early breast cancer: A pooled analysis of randomized trials. *Journal of the National Cancer Institute* **2008,** *100* (1), 14-20.
32. Moscat, J.; Diaz-Meco, M. T.; Albert, A.; Campuzano, S., Cell signaling and function organized by PB1 domain interactions. *Mol Cell* **2006,** *23* (5), 631-40.

33. Duran, A.; Serrano, M.; Leitges, M.; Flores, J. M.; Picard, S.; Brown, J. P.; Moscat, J.; Diaz-Meco, M. T., The atypical PKC-interacting protein p62 is an important mediator of RANK-activated osteoclastogenesis. *Dev Cell* **2004,** *6* (2), 303-9.

34. Rodriguez, A.; Duran, A.; Selloum, M.; Champy, M. F.; Diez-Guerra, F. J.; Flores, J. M.; Serrano, M.; Auwerx, J.; Diaz-Meco, M. T.; Moscat, J., Mature-onset obesity and insulin resistance in mice deficient in the signaling adapter p62. *Cell Metab* **2006,** *3* (3), 211-22.

35. Duran, A.; Linares, J. F.; Galvez, A. S.; Wikenheiser, K.; Flores, J. M.; Diaz-Meco, M. T.; Moscat, J., The signaling adaptor p62 is an important NF-kappaB mediator in tumorigenesis. *Cancer Cell* **2008,** *13* (4), 343-54.

36. Seibenhener, M. L.; Babu, J. R.; Geetha, T.; Wong, H. C.; Krishna, N. R.; Wooten, M. W., Sequestosome 1/p62 is a polyubiquitin chain binding protein involved in ubiquitin proteasome degradation. *Mol Cell Biol* **2004,** *24* (18), 8055-68.

37. Yu, H. B.; Kielczewska, A.; Rozek, A.; Takenaka, S.; Li, Y.; Thorson, L.; Hancock, R. E.; Guarna, M. M.; North, J. R.; Foster, L. J.; Donini, O.; Finlay, B. B., Sequestosome-1/p62 is the key intracellular target of innate defense regulator peptide. *J Biol Chem* **2009,** *284* (52), 36007-11.

38. Moscat, J.; Diaz-Meco, M. T., p62 at the crossroads of autophagy, apoptosis, and cancer. *Cell* **2009,** *137* (6), 1001-4.

39. Cande, C.; Cohen, I.; Daugas, E.; Ravagnan, L.; Larochette, N.; Zamzami, N.; Kroemer, G., Apoptosis-inducing factor (AIF): a novel caspase-independent death effector released from mitochondria. *Biochimie* **2002,** *84* (2-3), 215-22.

40. Joza, N.; Pospisilik, J. A.; Hangen, E.; Hanada, T.; Modjtahedi, N.; Penninger, J. M.; Kroemer, G., AIF: Not Just an Apoptosis-Inducing Factor. *Natural Compounds and Their Role in Apoptotic Cell Signaling Pathways* **2009,** *1171*, 2-11.

41. Cregan, S. P.; Dawson, V. L.; Slack, R. S., Role of AIF in caspase-dependent and caspase-independent cell death. *Oncogene* **2004,** *23* (16), 2785-96.

42. Daugas, E.; Nochy, D.; Ravagnan, L.; Loeffler, M.; Susin, S. A.; Zamzami, N.; Kroemer, G., Apoptosis-inducing factor (AIF): a ubiquitous mitochondrial oxidoreductase involved in apoptosis. *FEBS Lett* **2000,** *476* (3), 118-23.

43. Millan, A.; Huerta, S., Apoptosis-inducing factor and colon cancer. *J Surg Res* **2009,** *151* (1), 163-70.

44. Ji, H.; Ding, Z.; Hawke, D.; Xing, D.; Jiang, B. H.; Mills, G. B.; Lu, Z., AKT-dependent phosphorylation of Niban regulates nucleophosmin- and MDM2-mediated p53 stability and cell apoptosis. *EMBO Rep* **2012,** *13* (6), 554-60.

45. Liang, K.; Lu, Y.; Jin, W.; Ang, K. K.; Milas, L.; Fan, Z., Sensitization of breast cancer cells to radiation by trastuzumab. *Mol Cancer Ther* **2003,** *2* (11), 1113-20.

46. Siddiqa, A.; Long, L. M.; Li, L.; Marciniak, R. A.; Kazhdan, I., Expression of HER-2 in MCF-7 breast cancer cells modulates anti-apoptotic proteins Survivin and Bcl-2 via the extracellular signal-related kinase (ERK) and phosphoinositide-3 kinase (PI3K) signalling pathways. *BMC Cancer* **2008,** *8*, 129.

47. Brune, B., Nitric oxide: NO apoptosis or turning it ON? *Cell Death Differ* **2003,** *10* (8), 864-9.

48. Oyadomari, S.; Takeda, K.; Takiguchi, M.; Gotoh, T.; Matsumoto, M.; Wada, I.; Akira, S.; Araki, E.; Mori, M., Nitric oxide-induced apoptosis in pancreatic beta cells is mediated by the endoplasmic reticulum stress pathway. *Proc Natl Acad Sci U S A* **2001,** *98* (19), 10845-50.

49. Dhakshinamoorthy, S.; Porter, A. G., Nitric oxide-induced transcriptional up-regulation of protective genes by Nrf2 via the antioxidant response element counteracts apoptosis of neuroblastoma cells. *J Biol Chem* **2004,** *279* (19), 20096-107.

50. Mortensen, K.; Skouv, J.; Hougaard, D. M.; Larsson, L. I., Endogenous endothelial cell nitric-oxide synthase modulates apoptosis in cultured breast cancer cells and is transcriptionally regulated by p53. *J Biol Chem* **1999,** *274* (53), 37679-84.

51. Li, C. Q.; Robles, A. I.; Hanigan, C. L.; Hofseth, L. J.; Trudel, L. J.; Harris, C. C.; Wogan, G. N., Apoptotic signaling pathways induced by nitric oxide in human lymphoblastoid cells expressing wild-type or mutant p53. *Cancer Res* **2004,** *64* (9), 3022-9.

52. Wang, X.; Michael, D.; de Murcia, G.; Oren, M., p53 Activation by nitric oxide

involves down-regulation of Mdm2. *J Biol Chem* **2002,** *277* (18), 15697-702.

53. Dawood, S.; Broglio, K.; Buzdar, A. U.; Hortobagyi, G. N.; Giordano, S. H., Prognosis of Women With Metastatic Breast Cancer by HER2 Status and Trastuzumab Treatment: An Institutional-Based Review. *Journal of Clinical Oncology* **2010,** *28* (1), 92-98.

Chapter 5 - Quantitative Proteomic Analysis of Pig Plasma By Mass Spectrometry for Early Detection of Deep Tissue Injury

5.1 Introduction

Pressure ulcers refer to the breakdown of soft tissue around bony prominences in people with reduced mobility or sensation due to prolonged sitting or lying down. These ulcers are particularly prevalent among the elderly[1-2], people with mobility impairments due to spinal cord injury[3], head trauma or multiple sclerosis, individuals with musculoskeletal diseases, people in coma or those undergoing long surgical procedures[4]. The incidence of pressure ulcers in clinical centers varies, but is as high as 40% in acute care facilities and 39.4% in nursing homes[5-9]. The prevalence of pressure ulcers in people with spinal cord injury is 25–33%[10-15]. Moreover, 80% of people with spinal cord injury (SCI) develop at least one pressure ulcer following their injury[16]. Pressure ulcers can greatly affect the quality of life of those affected[17]. Treatment of these ulcers can be a lengthy and costly process[18-19], in many cases requiring surgery to repair the extensive tissue damage. Furthermore, pressure ulcer recurrence rates are as high as 91%[20]. This fact highlights the importance of preventing pressure ulcers from developing in the first place.

Pressure ulcers can be classified into two categories: those that originate at the level of the skin (outside-in)[21] and those that originate at deep bone-muscle interfaces (inside-out)[22-24]. Outside-in ulcers develop due to skin abrasion, poor nutrition or hygiene, and excessive skin moisture or dryness[25]. These ulcers are detected via skin inspection and can be prevented early on in their stage of development. Inside-out ulcers, known as deep tissue injury (DTI), develop due to prolonged loading and deformation of soft tissue trapped between a bony prominence and an external surface[26-28]. Because of their deep origin, these ulcers' developments are difficult to be noticed to the individuals or their caregivers. Once skin signs are detected, extensive damage of the underlying tissue would have already occurred. Therefore the requirement of effective early detection method for

DTI is mandatory.

Magnetic Resonance Imaging (MRI) has been used to assess the deformation in tissue during sitting in order to estimate the strain levels in the gluteal muscles and build finite element models of the buttocks in humans[29-31]. However, it's not convenient for routine clinical investigation and diagnosis.

Intermittent electrical stimulation (IES) is recently reported by Solis, *et al*. as a novel strategy to prevent the formation of DTI[32-33]. They demonstrated IES was highly effective in redistributing internal pressure by reconfiguring muscle shape and increasing muscle stiffness. However, the molecular mechanism of IES in DTI treatment and prevention is still unknown.

In our project, we hypothesize the presence of specific biomarkers to be associated early in DTI progression prior to the visual detection, and envisage that a simple blood test would confer the ability to monitor recovery, compare efficacy of treatments, and provide prophylactic monitoring method to help prevent DTI. We sought to identify proteomic changes accompanying DTI development and IES in order to i) identify new biomarkers associated with DTI, ii) gauge the magnitude of the proteomic changes induced by IES treatment, and iii) derive a better understanding of the downstream biological changes triggered by IES.

To address these questions, we performed quantitative proteomic analysis of pig plasma samples: DTI with IES treatment, DTI without IES treatment, and intact pig plasma. In the analysis, we applied offline two-dimensional liquid chromatography (2D-LC) coupled with mass spectrometry (MS) based isotopic labeling quantitative proteomic approach. For stable isotopic labeling, we applied the technique of N-terminal dimethylation after lysine guanidination (2-MEGA), which had been developed by our group[34-35].

5.2 Experimental Section

5.2.1 Chemicals and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless stated otherwise. The isotope reagent, d(6), ^{13}CD-formaldehyde (20%, w/w in deuterated H_2O) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). Progenta™ Anionic Acid Labile Surfactant I was obtained from Protea Bioscience, Inc. (Morgantown, WV). Acetonitrile and water were purchased from Fisher Scientific Canada (Edmonton, AB, Canada).

5.2.2 Depletion of High Abundant Proteins

Pig plasma samples of deep tissue injury (DTI) were treated by affinity separation to remove high abundant proteins. To deplete hemoglobin, the hemolyzed serum was mixed with equal volume HemogloBind™ suspension (Biotech Support Group, LLC., Monmouth JCT, NJ, USA) in a microfuge tube. After vortexing for 30 s, the mixture was shaken by inversion for 15 min, followed by centrifuge at 9000 rpm for 2 min. Supernatant was collected as the hemoglobin-depleted proteome sample.

Albumin was removed by SwellGel® Blue Albumin Removal Kit (Pierce Biotechnology, Rockford, IL, USA) as instructed. SwellGel® Blue Discs were hydrated in 380 µL ultrapure water by vortexing 1-2 s before put in a mini spin column. The mini spin column was placed in a 1.5 mL collection tube and spun at 12,000×g for 1 min to remove excess liquid. The flow through liquid was discarded. Plasma was loaded with an equal volume of Tris buffer (pH = 7.4), and incubated for 1-2 min at room temperature. The mini spin column with sample loaded was centrifuged at 12,000×g for 1 min. The flow through was recovered and re-applied to the column. After incubating the column for another 1-2 min to ensure maximum albumin binding, the mini spin column was centrifuged at 12,000×g for 1 min and the flow through was retained. Then the albumin-depleted proteome sample was eluted from the column

with 100 µL of 20 mM Tris buffer (pH = 7.4) by spinning at 12,000×g for 1 min. The elution step was repeated for 6 times. Elutes were pooled together.

Calbiochem® ProteoExtract™ Albumin/IgG Kits (EMD Bioscience, Inc., La Jolla, CA, USA) was applied to remove Albumin and IgG from serum. Before loading proteome sample, the column was rinsed with 0.85 mL binding buffer. Then the diluted sample was loaded onto the column and allowed passing the resin bed by gravity flow. The flow through was collected. The column was later washed with 500 µL of binding buffer by gravity flow. The washing step was repeated twice and the washing fractions were collected and combined. The combined fractions contained the albumin/ IgG depleted samples.

5.2.3 In-solution Digestion

Protein concentration was determined by the BCA assay (Bio-Rad). Proteins were reduced with dithiothreitol (DTT) and alkylated with iodoacetaminde. Then, the reduced and alkylated proteins were precipitated by adding acetone to a final concentration of 80% and incubated in -80℃ for 4 hours, in order to remove salts and other impurities. The mixture was centrifuged at 20,000×g for 15 min. The supernatant was decanted and protein pellets were redissolved with 100 mM ammonium bicarbonate and digested by trypsin at a ratio of 1:50 at 37℃ overnight.

5.2.4 2-MEGA Isotopic Labeling

Peptides were isotopic labeled with the 2-MEGA method using a liquid handler as described previously[34-35] with some modifications. Trypsin in the 500 µL tryptic digest solution (about 1 µg/µL) was irreversibly inactivated by adding 24 µL 2 M sodium hydroxide to adjust pH equal to 11. The ε-amino groups of lysines were blocked by reacting with 100 µL 6 M O-methylisourea at 37℃ for 1 h and the reaction was stopped by adjusting pH to 5 with 48 µL 3 M hydrogen chloride. Dimethylation with 10 µL d(0), ^{12}C-formaldehyde, or d(6), ^{13}CD-formaldehyde (4%, w/w) was carried out at 37℃ for 30 min, using 40 µL 2 M 2-picoline borane as the

reductive reagent. Then, 32 µL 1 M ammonium bicarbonate was added to the reaction mixture followed by incubation at 37℃ for 15 min to consume the excess formaldehyde. Finally, the reaction was stopped by adjusting pH to 2 with 10% TFA. This whole reaction process was carried out on a liquid handler automatically.

5.2.5 Strong Cation Exchange (SCX) Liquid Chromatography

After isotopic labeling, peptide mixtures were fractionated by using strong cation exchange (SCX) liquid chromatography. A gradient for elution was established with solvent A (50 mM KH_2PO_4, pH 2.7) and solvent B (1 M KCl in 50 mM K_2PO_4, pH 2.7): 0-1 min, 0-4% solvent B; 1-12 min, 4-20%; 12-45 min, 20-60%; 45-50 min, 60-100%; 50-53 min, 100%; 53-55 min, 100-0%; and kept flushing the column with 0% solvent B until 62 min. Peptide fractions were collected in 1 min fractions from 2 min to 60 min, followed by desalting and quantification on RPLC. Less abundant neighboring fractions were pooled together to obtain a fraction of >1 µg of peptides.

5.2.6 Liquid Chromatography Mass Spectrometric Analysis

An electrospray ionization (ESI) quadrupole time-of-flight (Q-TOF) Premier mass spectrometer (Waters) equipped with a nanoACQUITY Ultra Performance LC system (Waters) was used to analyze the peptide fractions. Solvent A consisted of 0.1% formic acid in deionized water, and solvent B contained 0.1% formic acid in acetonitile. The following 130 min LC gradient was used to separate the peptides: 0-2 min, 2-7% solvent B; 2-85 min, 7-20%; 85-105 min, 20-30%; 105-110 min, 30-45%; 110-120 min, 45-90%; 120-125 min, 90%; and 125-130 min, 90-2%. The separated peptides were electrosprayed into the mass spectrometer at a flow rate of 350 nl/min. Using the optimized data acquisition conditions, mass spectra were acquired from m/z 350 to 1600 for 1 s followed by four data dependent MS/MS analyses from m/z 50 to 1990 for 0.8 s each, with a precursor ion exclusion strategy to eliminate the redundant identification from two adjacent SCX fractions.

5.2.7 Database Search and Bioinformatics

MS and MS/MS spectral data were processed and searched using MASCOT DISTILLER. The software picked out paired peaks from individual MS spectra, calculated peak areas and worked out peptide ion relative intensity ratios in two comparative samples. MS/MS spectra taken for the paired peptides peaks were searched using the following parameters: database: Sus scrofa (Pig); taxonomy: Entire; enzyme: Trypsin; missed cleavage: 2; fixed modification: Carbamidomethylation (C); variable modification: Guanidinyl (K), Dimethylationd(0) (N-term, + C_2H_4, + 28.0313 Da); Dimethylationd(6) (N-term, + $^{13}C_2D_4$, + 34.0631 Da); MS tolerance: 30 ppm; MS/MS tolerance: 0.2 Da.

5.3 Results and Discussion

5.3.1 Depletion of High Abundant Proteins

Plasma is a challenging biological matrix to work with due to the presence of several high abundance proteins, such as albumin, immunoglobulins, and various components of the complement system. In human, albumin counts for 55% of the total quantity of the serum proteome[36]. Besides these high abundant proteins, the wide dynamic range of protein concentrations makes plasma proteomic analysis more difficult[37]. If lower concentration proteins of interest are to be studied, protein level separation must be performed to remove or separate these high abundant components from other proteins. As an additional step performed prior to isotopic labeling for quantitative proteomics, this can potentially introduce errors into the quantification ratios observed.

The most frequently used method for albumin and high-abundance protein depletion is immunoaffinity columns that use antibodies raised against specific high-abundance plasma proteins[38]. Commercially available products are available for analyzing plasma samples for various research and pharmaceutical applications.

In our experiments, we applied three affinity depletion steps to sequentially remove albumin, hemoglobin and immunoglobin (IgG). After the affinity depletion, their concentrations in the porcine plasma sample were greatly decreased (Data not shown).

5.3.2 2-MEGA Labeling

After immunoaffinity depletion of albumin, hemoglobin and IgG, sample volume was greatly increased. This was not convenient for downstream sample preparation and a too diluted proteome sample may have lower labeling efficiency for 2-MEGA labeling. Normally, the reaction concentration of a proteome sample in 2-MEGA labeling chemistry is from 0.5 to 2 μg/μL. In this range, the labeling efficiency of 2-MEGA can be around 95% as reported[34, 39]. But, will 2-MEGA chemistry work the same in a proteome sample with a concentration of lower than 0.5 μg/μL? This question could be answered by using a linear dilution labeling experiment with MB231 human breast cancer cell.

In the 2-MEGA linear dilution labeling experiment, we manually labeled 10 μg MB231 digests with different proteome concentrations: 0.5 μg/μL, 0.25 μg/μL, 0.125 μg/μL. The labeled MB231 digests were desalted, then separated and quantified on LC-ESI-MS. The quantification results were summarized in Table 5.1. From the results, we can see that, with the decreasing of proteome concentration (equal or lower than 0.5 μg/μL) but keeping the same sample amount (10 μg), the labeling efficiency of 2-MEGA was decreased. The most frequent side reaction was guanidination on the N-termini of peptides in the 2-MEGA labeling reaction. And the frequency of quanidination on N-termini was increased by decreasing the proteome concentration in a reaction solution (Figure 5.1). The reaction mechanism was still not clear. However, the experiment results showed that it's better to perform 2-MEGA labeling chemistry with a proteome concentration of not lower than 0.5 μg/μL.

Table 5.1. Correct reaction rate of 2-MEGA labeling with low peptide concentrations.

	0.5 μg/μL (10 μg)		0.25 μg/μL (10 μg)		0.125 μg/μL (10 μg)	
Correct Labeled Peptide #	645	674	634	624	665	563
Total Identified Peptide #	678	710	691	686	749	628
Correct Labeling Rate	95.13%	94.93%	91.75%	90.96%	88.79%	89.65%

5.3.3 Quantification by Mass Spectrometry

To investigate the reproducibility of quantification, replicate experiments were carried out through forward and reverse labeling on the comparison sample set, including porcine plasma of DTI pig models with or without IES treatment (DTI, +/-IES) and porcine plasma before or after spinal cord hemi-transection surgery (Control, +/-Surgery). After reduction, alkylation, precipitation and trypsin digestion, the DTI and Control porcine plasma samples were divided into two equal parts and labeled by the 2-MEGA method, followed by quantification and desalting on RPLC. Based on the quantification results of the total peptide amount in each proteome sample, the labeled peptides were accurately mixed with 1:1 ratio. The heavy isotopic labeled DTI sample (A_H) was mixed with the light isotopic labeled Control sample (B_L) to produce an A_HB_L mixture, while the mixing of light isotopic labeled DTI porcine plasma sample and heavy isotopic labeled Control sample giving an A_LB_H mixture. The two mixtures were fractionated on SCX respectively and were injected onto LC coupled with ESI-QTOF for separation and MS analysis (Figure 5.2). The precursor ion exclusion (PIE) strategy and optimized MS conditions as descripted in Chapter 4 were applied to increase the quantified peptides number.

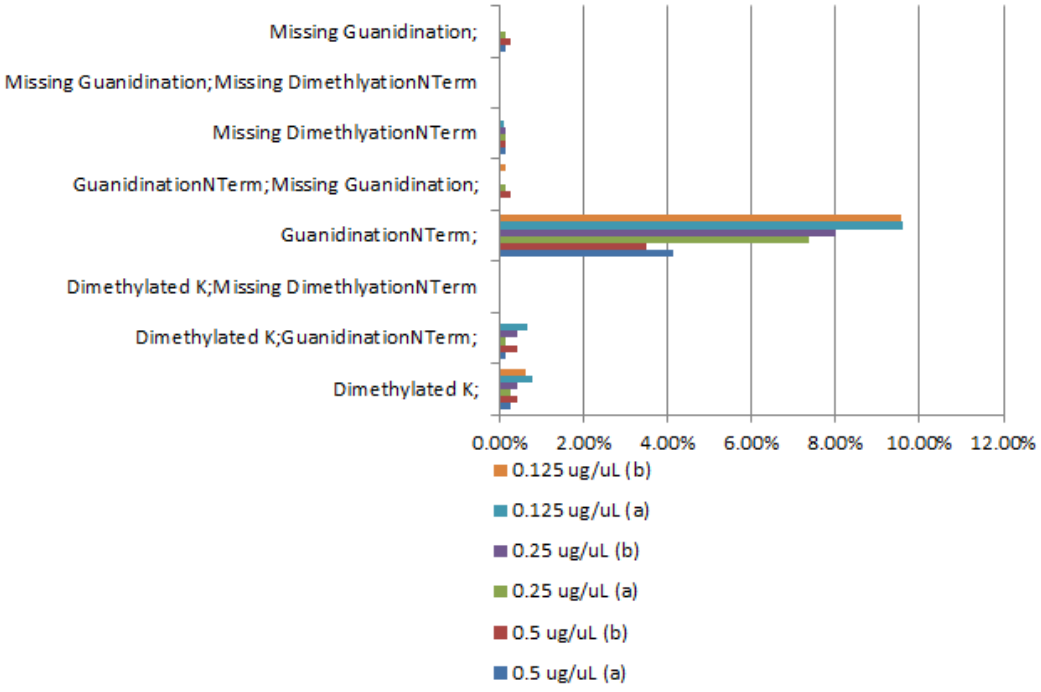


Figure 5.1. Mis-reactions of 2-MEGA labeling with low peptide concentrations.

Peptide quantitation was performed using MASCOT Distiller. Peptide matches found through database searching were processed normally, while peptides identified with single amino acid substitutions were manually entered into a local search database and processed separately. Running through the SCX fractions, a list of quantified peptides would be generated by the program. For the forward (A_HB_L) and reverse (A_LB_H) labeled peptides samples, two different SCX fractions were analyzed and two lists of quantified peptides were produced. The lists of individual quantified peptides were matched against each other to find common peptides. Based on the identified and quantified common peptides, a list of their corresponding proteins was generated.

In total, 126 proteins from pig plasma were quantified. One reason for the limited quantification is ion suppression from highly abundant proteins, like albumin, in plasma samples. Although in our experiment, before analyzed on MS, highly abundant proteins were depleted from the plasma sample by affinity separation, protein depletion was not perfect. As the concentration of albumin in plasma is as high as 34 to 54 mg/mL[40], even with 90% depletion efficiency using immunoaffinity column and twice removals, its concentration would be 340 μg/mL to 540 μg/mL, which is still pretty high for a single protein. The presence of these highly abundant proteins in plasma would cause severe ion suppression on low abundant proteins, making them difficult to be ionized, detected and quantified. Another challenge is the lack of a completed porcine proteome database. These were the two major reasons for the limited quantification in pig plasma. With the improvement of high abundant proteins removal technique and the completion of porcine's proteome database, I believe more information could be obtained from these pig plasma samples. In particular, the MS/MS data collected in this work may be re-searched against a more completed proteome database available in the future to generate a more comprehensive list of quantified proteins.

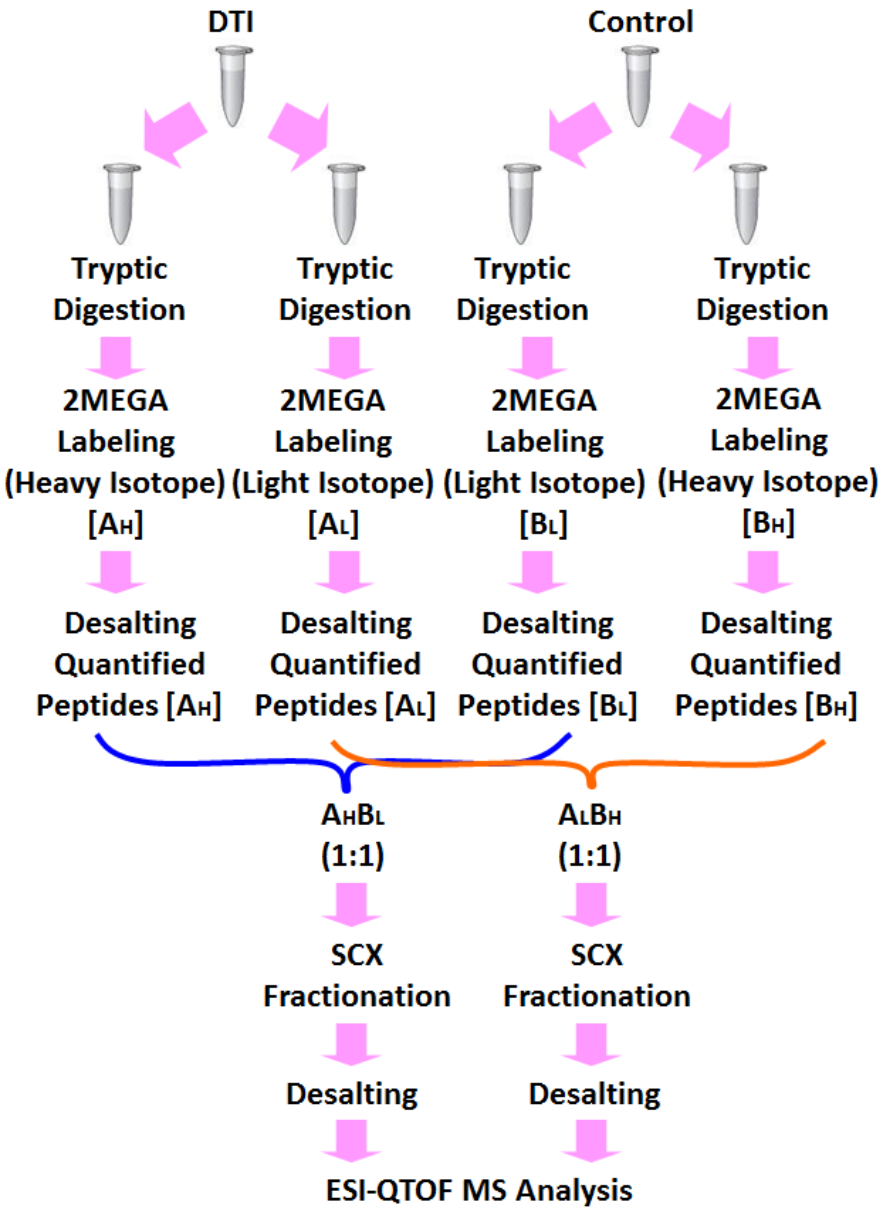


Figure 5.2. Workflow of the pig plasma sample preparation used for quantitative proteome profiling of DTI and Control.

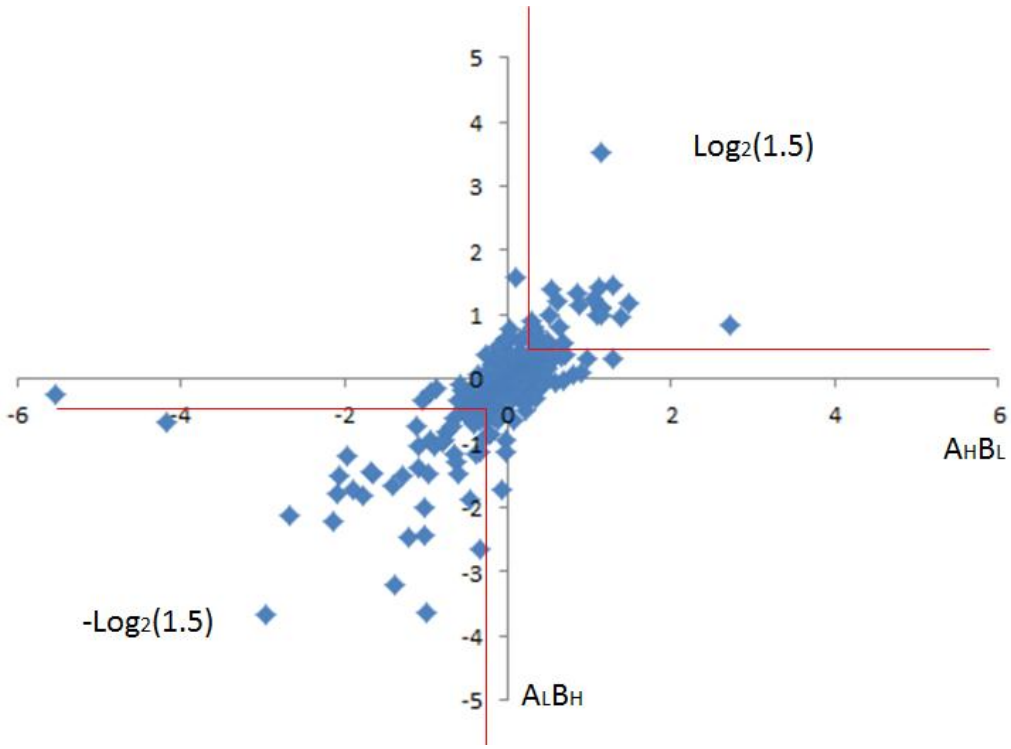


Figure 5.3. Log-log of forward and reverse labeled proteins. The differentiation threshold used was 1.5/0.67.

Based on the quantification results of proteins, we then applied the intensity ratio of 1.50/0.67 as the threshold to determine the proteins that were deemed to be dfferentially expressed in the two samples (Figure 5.3).

5.3.4 Bioinformatics

The information of each protein's biological function was extracted by searching the protein in UniProtKB database under the taxonomy: Sus scrofa (Pig) (www.expasy.org).

In this project, we made comparisons between: i) DTI (L1: 4 times 25% body weight laoding), +IES porcine plasma *vs* Control, +Surgery porcine plasma and DTI (L1), -IES *vs* Control, +Surgery; ii) DTI (L2: 2 times 25% BW laoding), +IES *vs* Control, +Surgery and DTI (L2), -IES *vs* Control, +Surgery; iii) DTI (L2), +IES *vs* Control, +Surgery and DTI (L2), +IES *vs* Control, -Surgery; iv) DTI (L2), -IES *vs* Control, +Surgery and DTI (L2), -IES *vs* Control, -Surgery. Only the common proteins in each comparison with differential expression level changes were searched in the bioinformatics database (Table 5.2). From the list, we could find many interesting proteins had differential expression level changes in the development of disease or before/after IES treatment (Table 5.3). In the future, more bioinformatics study, biological validation and evaluation of their prognostic values would be done based on this list of biomarker candidates, in order to find new biomarkers associated with DTI.

5.4 Conclusion.

Applying 2-MEGA labeling chemistry to isotopically label the tryptic peptides, we successfully quantified 126 proteins from pig plasma samples collected at different stages of deep tissue injury by 2D-LC-MS. In sample preparation, in order to ensure labeling efficiency of 2-MEGA, a linear dilution labeling experiment was done to find out the proper reaction concentration range. From our results, we conclude that 2-MEGA labeling is better to be done with a peptide concentration of not lower than 0.5 μg/μL. From the list of the quantified proteins, we determined the proteins

with differential expression level changes in the development of disease or IES treatment, and made a list of potential biomarker candidates based on bioinformatics study.

5.5 Future Work

More bioinformatics study, biological validation and evaluation of their prognostic values need to be done based on the biomarker candidate list, in order to find new biomarkers associated with DTI. In addition, more biomarker candidates may be generated by re-searching the MS/MS and quantitative data collected in this work when a more complete pig proteome database becomes available in the future.

Table 5.2. List of common proteins from pig plasma collected at different stages of deep tissue injury.

Protein ID	Protein Name	Twice of 25% Body Weight Loading		4 Times of 25% Body Weight Loading	
		+IES	-IES	+IES	-IES
APOA1_PIG		1.145	0.793	0.711	1.108
B3CL06_PIG		1.090	0.732	0.796	1.380
CLUS_PIG	Clusterin, CP40	0.901	0.808	0.628	1.549
F1RHH5_PIG	SERPINF2	0.997	0.708	0.662	1.470
F1RII7_PIG	LOC100515788	0.344	0.895	0.622	0.906
F1RJS2_PIG	Clusterin	0.941	0.942	0.532	1.771
F1RLU2_PIG		1.093	0.980	0.919	1.498
F1RMN7_PIG	Hemopexin	0.657	1.626	5.385	1.592
F1RUM1_PIG	AFM	1.331	0.526	0.923	1.617
F1RUN2_PIG	ALB	1.039	1.192	1.637	1.020
F1RUQ0_PIG	IGJ	1.179	0.766	0.860	2.000
F1RXM6_PIG	SERPINA7	1.325	0.696	0.629	1.331
F1S1G8_PIG	Amine oxidase	0.952	1.078	0.713	1.757
F1S7K2_PIG	LRG1	0.898	0.812	0.610	1.356
F1SC20_PIG	LOC100516980	0.949	0.971	0.789	1.721
F1SCC6_PIG		0.943	0.783	0.677	1.150
F1SCC7_PIG	SERPINA3-1	0.602	0.739	0.655	1.185
F1SCD0_PIG		0.955	0.915	0.716	1.288
F1SCF0_PIG	Alpha-1-antitrypsin	0.795	0.716	0.510	1.537
F1SFI7_PIG	Alpha-2-HS-glyco protein	0.986	0.657	0.711	1.209
F1SH94_PIG	ITIH3	1.344	1.145	0.848	1.769
F1SJT7_PIG	Apolipoprotein A-IV	1.039	0.769	0.468	3.577
F1SKB1_PIG	CP	0.805	0.735	0.529	1.166
F1SLX2_PIG		1.239	0.988	0.788	1.406
F1SN68_PIG	ORM1	1.180	1.187	0.978	2.156
F1SN71_PIG	Trypstatin	0.922	1.136	0.705	2.081
HPT_PIG		0.761	0.931	0.658	0.962
ICA_PIG	Haptoglobin	2.225	0.894	1.608	2.939
Q66RQ0_PIG	Vitamin D-binding protein	0.937	0.650	0.525	1.150
Q9GKP1_PIG	Complement C3	0.787	0.486	0.313	1.625
Q9GMA7_PIG		0.758	0.818	0.674	1.373

Table 5.3. List of proteins having biological functions potentially related to deep tissue injury.

Protein ID	Protein Name	Bio-functions
CLUS_PIG	Clusterin, CP40	cell death, positive regulation of NF-kappaB
F1RHH5_PIG	SERPINF2	serine-type endopeptidase inhibitor activity
F1RII7_PIG	LOC100515788	oxygen transporter activity
F1RJS2_PIG	Clusterin	cell death
F1RMN7_PIG	Hemopexin	positive regulation of humoral immune response mediated by circulating immunoglobulin
F1RUM1_PIG	AFM	
F1RUN2_PIG	ALB	transport
F1RUQ0_PIG	IGJ	
F1RXM6_PIG	SERPINA7	serine-type endopeptidase inhibitor activity
F1S1G8_PIG	Amine oxidase	copper ion binding
F1S7K2_PIG	LRG1	
F1SC20_PIG	LOC100516980	
F1RUQ0_PIG	IGJ	
F1SCC7_PIG	SERPINA3-1	serine-type endopeptidase inhibitor activity
F1SCF0_PIG	Alpha-1-antitrypsin	
F1SFI7_PIG	Alpha-2-HS-glyco protein	cysteine-type endopeptidase inhibitor activity
F1SH94_PIG	ITIH3	hyaluronan metabolic process
F1SJT7_PIG	Apolipoprotein A-IV	lipid transport, lipoprotein metabolic process
F1SKB1_PIG	CP	Oxidoreductase
F1SN68_PIG	ORM1	regulation of immune system process
F1SN71_PIG	Trypstatin	transporter activity
ICA_PIG	Haptoglobin	serine-type endopeptidase activity
Q66RQ0_PIG	Vitamin D-binding protein	vitamin D binding, vitamin transporter activity
Q9GKP1_PIG	Complement C3	complement activation,inflammatory response

5.6 Literature Cited

1. Bergquist-Beringer, S.; Gajewski, B. J., Outcome and assessment information set data that predict pressure ulcer development in older adult home health patients. *Adv Skin Wound Care* **2011,** *24* (9), 404-14.
2. Jaul, E., Assessment and management of pressure ulcers in the elderly: current strategies. *Drugs Aging* **2010,** *27* (4), 311-25.
3. Garber, S. L.; Rintala, D. H., Pressure ulcers in veterans with spinal cord injury: a retrospective study. *J Rehabil Res Dev* **2003,** *40* (5), 433-41.
4. Pham, B.; Teague, L.; Mahoney, J.; Goodman, L.; Paulden, M.; Poss, J.; Li, J.; Sikich, N. J.; Lourenco, R.; Ieraci, L.; Carcone, S.; Krahn, M., Support surfaces for intraoperative prevention of pressure ulcers in patients undergoing surgery: a cost-effectiveness analysis. *Surgery* **2011,** *150* (1), 122-32.
5. Cox, J., Predictors of pressure ulcers in adult critical care patients. *Am J Crit Care* **2011,** *20* (5), 364-75.
6. Keller, B. P.; Wille, J.; van Ramshorst, B.; van der Werken, C., Pressure ulcers in intensive care patients: a review of risks and prevention. *Intensive Care Med* **2002,** *28* (10), 1379-88.
7. Webster, C. I., A pressure care survey in the operating theatres. *Aust Clin Rev* **1993,** *13* (1), 29-37.
8. Woolsey, R. M.; McGarry, J. D., The cause, prevention, and treatment of pressure sores. *Neurol Clin* **1991,** *9* (3), 797-808.
9. Zanca, J. M.; Brienza, D. M.; Berlowitz, D.; Bennett, R. G.; Lyder, C. H., Pressure ulcer research funding in America: creation and analysis of an on-line database. *Adv Skin Wound Care* **2003,** *16* (4), 190-7.
10. Bansal, C.; Scott, R.; Stewart, D.; Cockerell, C. J., Decubitus ulcers: a review of the literature. *Int J Dermatol* **2005,** *44* (10), 805-10.
11. Klotz, R.; Joseph, P. A.; Ravaud, J. F.; Wiart, L.; Barat, M., The Tetrafigap Survey on the long-term outcome of tetraplegic spinal cord injured persons: Part III. Medical complications and associated factors. *Spinal Cord* **2002,** *40* (9),

457-67.

12. Krause, J. S.; Broderick, L., Patterns of recurrent pressure ulcers after spinal cord injury: identification of risk and protective factors 5 or more years after onset. *Arch Phys Med Rehabil* **2004,** *85* (8), 1257-64.
13. Krouskop, T. A.; Williams, R.; Noble, P.; Brown, J., Inflation pressure effect on performance of air-filled wheelchair cushions. *Arch Phys Med Rehabil* **1986,** *67* (2), 126-8.
14. Noreau, L.; Proulx, P.; Gagnon, L.; Drolet, M.; Laramee, M. T., Secondary impairments after spinal cord injury: a population-based study. *Am J Phys Med Rehabil* **2000,** *79* (6), 526-35.
15. Thomas, D. R., Are all pressure ulcers avoidable? *J Am Med Dir Assoc* **2003,** *4* (2 Suppl), S43-8.
16. Salzberg, C. A.; Byrne, D. W.; Cayten, C. G.; van Niewerburgh, P.; Murphy, J. G.; Viehbeck, M., A new pressure ulcer risk assessment scale for individuals with spinal cord injury. *Am J Phys Med Rehabil* **1996,** *75* (2), 96-104.
17. Krouskop, T. A.; Noble, P. C.; Garber, S. L.; Spencer, W. A., The effectiveness of preventive management in reducing the occurrence of pressure sores. *J Rehabil R D* **1983,** *20* (1), 74-83.
18. Grip, J. C.; Merbitz, C. T., Wheelchair-based mobile measurement of behavior for pressure sore prevention. *Comput Methods Programs Biomed* **1986,** *22* (1), 137-44.
19. Rischbieth, H.; Jelbart, M.; Marshall, R., Neuromuscular electrical stimulation keeps a tetraplegic subject in his chair: a case study. *Spinal Cord* **1998,** *36* (6), 443-5.
20. Niazi, Z. B.; Salzberg, C. A.; Byrne, D. W.; Viehbeck, M., Recurrence of initial pressure ulcer in persons with spinal cord injuries. *Adv Wound Care* **1997,** *10* (3), 38-42.
21. Bouten, C. V.; Oomens, C. W.; Baaijens, F. P.; Bader, D. L., The etiology of pressure ulcers: skin deep or muscle bound? *Arch Phys Med Rehabil* **2003,** *84* (4), 616-9.

22. Agam, L.; Gefen, A., Pressure ulcers and deep tissue injury: a bioengineering perspective. *J Wound Care* **2007,** *16* (8), 336-42.

23. Ceelen, K. K.; Stekelenburg, A.; Loerakker, S.; Strijkers, G. J.; Bader, D. L.; Nicolay, K.; Baaijens, F. P.; Oomens, C. W., Compression-induced damage and internal tissue strains are related. *J Biomech* **2008,** *41* (16), 3399-404.

24. Daniel, R. K.; Priest, D. L.; Wheatley, D. C., Etiologic factors in pressure sores: an experimental model. *Arch Phys Med Rehabil* **1981,** *62* (10), 492-8.

25. Lahmann, N. A.; Kottner, J., Relation between pressure, friction and pressure ulcer categories: a secondary data analysis of hospital patients using CHAID methods. *Int J Nurs Stud* **2011,** *48* (12), 1487-94.

26. Guthrie, R. H., Jr.; Goulian, D., Jr., Decubitus ulcers: prevention and treatment. *Geriatrics* **1973,** *28* (8), 67-71.

27. Nola, G. T.; Vistnes, L. M., Differential response of skin and muscle in the experimental production of pressure sores. *Plast Reconstr Surg* **1980,** *66* (5), 728-33.

28. Oomens, C. W.; Loerakker, S.; Bader, D. L., The importance of internal strain as opposed to interface pressure in the prevention of pressure related deep tissue injury. *J Tissue Viability* **2010,** *19* (2), 35-42.

29. Makhsous, M.; Lim, D.; Hendrix, R.; Bankard, J.; Rymer, W. Z.; Lin, F., Finite element analysis for evaluation of pressure ulcer on the buttock: development and validation. *IEEE Trans Neural Syst Rehabil Eng* **2007,** *15* (4), 517-25.

30. Shabshin, N.; Ougortsin, V.; Zoizner, G.; Gefen, A., Evaluation of the effect of trunk tilt on compressive soft tissue deformations under the ischial tuberosities using weight-bearing MRI. *Clin Biomech (Bristol, Avon)* **2010,** *25* (5), 402-8.

31. Sopher, R.; Nixon, J.; Gorecki, C.; Gefen, A., Exposure to internal muscle tissue loads under the ischial tuberosities during sitting is elevated at abnormally high or low body mass indices. *J Biomech* **2010,** *43* (2), 280-6.

32. Solis, L. R.; Gyawali, S.; Seres, P.; Curtis, C. A.; Chong, S. L.; Thompson, R. B.; Mushahwar, V. K., Effects of intermittent electrical stimulation on superficial

pressure, tissue oxygenation, and discomfort levels for the prevention of deep tissue injury. *Ann Biomed Eng* **2011,** *39* (2), 649-63.

33. Solis, L. R.; Hallihan, D. P.; Uwiera, R. R.; Thompson, R. B.; Pehowich, E. D.; Mushahwar, V. K., Prevention of pressure-induced deep tissue injury using intermittent electrical stimulation. *J Appl Physiol* **2007,** *102* (5), 1992-2001.

34. Ji, C.; Guo, N.; Li, L., Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. *J Proteome Res* **2005,** *4* (6), 2099-108.

35. Wang, P.; Lo, A.; Young, J. B.; Song, J. H.; Lai, R.; Kneteman, N. M.; Hao, C.; Li, L., Targeted quantitative mass spectrometric identification of differentially expressed proteins between Bax-expressing and deficient colorectal carcinoma cells. *J Proteome Res* **2009,** *8* (7), 3403-14.

36. Anderson, N. L.; Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **2002,** *1* (11), 845-67.

37. Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G., Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* **2002,** *1* (12), 947-55.

38. Bjorhall, K.; Miliotis, T.; Davidsson, P., Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* **2005,** *5* (1), 307-17.

39. Ji, C.; Zhang, N.; Damaraju, S.; Damaraju, V. L.; Carpenter, P.; Cass, C. E.; Li, L., A study of reproducibility of guanidination-dimethylation labeling and liquid chromatography matrix-assisted laser desorption ionization mass spectrometry for relative proteome quantification. *Anal Chim Acta* **2007,** *585* (2), 219-26.

40. Llewellyn, D. J.; Langa, K. M.; Friedland, R. P.; Lang, I. A., Serum albumin concentration and cognitive impairment. *Curr Alzheimer Res* **2010,** *7* (1), 91-6.

Chapter 6 - Differential Isotope Dansylation Labeling Combined With Liquid Chromatography Mass Spectrometry For Quantification Of Intact And N-Terminal Truncated Proteins

6.1 Introduction

N-terminal truncated protein is a protein form that has missed one amino acid or a stretch of amino acid sequence at the N-terminus of an intact protein, as a result of degradation from a biological process[1-3] or alternative translation initiation[4-6]. The formation of the N-terminal truncated protein may cause the loss of its intended biological function due to the loss of its structure integrity. For example, it has been shown that N-terminal truncated antibodies, enzymes and receptors would lose their activities since their N-terminal sequences are critical to their binding affinities with substrates[7-9]. Storage of protein-based pharmaceutical preparations may also cause the N-terminal degradation, resulting in changes of drug efficacy[10]. Therefore, identification and quantification of N-terminal truncated proteins from their intact forms are of great importance both as potential biomarkers to monitor biological changes in cellular processes and as indicators to control the quality of proteins used in various applications including protein-based drugs (e.g., antibody drugs) or vaccines. However, quantification of the N-terminal truncated proteins in the presence of intact proteins in a solution is a challenging task. Conventional UV or fluorescence based protein quantification methods cannot be readily used, as the N-terminal truncated protein and the intact protein usually have the identical spectroscopic characteristics. Separation of the two forms by liquid chromatography (LC) or electrophoresis is also difficult, because of the minor difference in protein sequences. Quantification of the protein mixture by mass spectrometry (MS) requires the use of protein standards, ideally isotope labeled standards[11], which can be difficult or expense to obtain, particularly for the truncated form.

In this work, we report a relatively simple and rapid method for absolute

quantification of a mixture of intact and N-terminal truncated proteins. It is based on the use of dansylation derivatization of proteins, microwave-assisted acid hydrolysis (MAAH), and LC-MS analysis of the dansyl labeled amino acids. In this strategy, the N-terminal amino acid of the truncated protein as well as the intact protein is first labeled with $^{12}C_2$-dansyl chloride to form protein derivatives. The labeled proteins are then subjected to microwave-assisted acid hydrolysis to degrade into amino acids including the $^{12}C_2$-dansyl labeled N-terminal amino acids. The resultant amino acids are analyzed by LC-MS and the $^{12}C_2$-dansyl labeled amino acids are quantified by using $^{13}C_2$-dansyl labeled amino acid standards. We demonstrate that this method can be used to quantify a protein mixture containing proteins with different N-terminal amino acids with high accuracy and precision.

6.2 Experimental Procedures

6.2.1 Chemicals and Reagents

All chemicals and reagents, except those specifically noted, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). LC/MS grade water and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

6.2.2 Cell Culture and Purification of mCherry and N-terminal Truncated mCherry

E. coli stable transfectants expressing mCherry (vector) (kindly provided by Professor Robert E. Campbell, Department of Chemistry, University of Alberta) were cultured in LB + amp medium. Cells were harvested at ~90% of confluence and lysated. mCherry was purified from the cell lysate by affinity separation with Ni-NTA agarose beads (Qiagen, Toronto, ON, Canada). To obtain the N-terminal truncated form of mCherry, the intact protein, mCherry, was digested with Enterokinase, Light Chain (NEB) at room temperature overnight. The N-terminal truncated protein was separated from mCherry with Ni-NTA agarose affinity separation. Supplemental

Note N6.1 shows the amino acid sequences of mCherry with M as the N-terminal amino acid and truncated mCherry with D as the N-terminal amino acid.

6.2.3 Dansyl Labeling of Amino Acids

$^{13}C_2$-dansyl chloride was synthesized following the protocol reported previously[12]. Amino acid standards (50 µL with 1 mM each) were mixed with 25 µL of 250 mM $NaHCO_3/Na_2CO_3$ buffer solution (pH = 9.4) and 25 µL of ACN. The solution was vortexed, spun down, and mixed with 50 µL of 18 mg/mL dansyl chloride dissolved in ACN, followed by incubation at 60 °C for 1 hr. After dansyl labeling, additional 10 min incubation was carried out at 60 ^{o}C with 10 µl of 250 mM NaOH to quench the excess dansyl chloride. Formic acid in 50% ACN/H_2O (50 µL of 425 mM) was then added to neutralize the solution.

6.2.4 Dansyl Labeling of Proteins

Protein standards were labeled with dansyl chloride as reported[13], with some modifications. In brief, 500 µL of 20 µM protein solution in 8 M urea was mixed with 150 µL of 400 mM Na_2HPO_4/Na_3PO_4 buffer (pH=9.4) and 250 µL of dimethylformamide (DMF). After vortexing and spinning down, 100 µL of 200 mM dansyl chloride in ACN was added. The reaction solution was vortexed at room temperature for 30 min. The resultant dansyl-labeled proteins were precipitated by incubating in 4 times volume of acetone at -80℃ overnight. The protein precipitates were collected by centrifuge at 12000 g at 4℃ for 20 min.

6.2.5 MAAH of Dansyl Labeled Proteins

The N-terminal amino acid of a dansyl labeled protein was generated by hydrolyzing the protein under an acidic condition with the assistance of microwave irradiation. The dansyl labeled protein was suspended in 500 µL of 0.1% phenol in 6 M HCl in 1.5-mL vial (Rose Scientific, Edmonton, AB, Canada). After adding the ^{13}C-labeled amino acid standards with known concentrations to the hydrolysis

solution, the solution vial was placed on a plastic rack floated in a water bath (100 mL glass beaker containing 80 mL of water) at the centre of a household 900 W (2450 MHz) microwave oven (Panasonic) for microwave irradiation[14]. The use of a water bath avoided the effects of hot/cold spots in household microwave oven[14-16], thereby ensuring even distribution of microwave energy for reproducible protein hydrolysis. The irradiation time was optimized to be 50 min. During hydrolysis, water was replenished every 10 min by adding 40 mL room temperature water. After hydrolysis, the protein hydrolysate solution was cooled on ice and dried down in a SpeedVac to remove the excess acid.

6.2.6 Quantification with LC-MS

Before LC-MS analysis, the protein hydrolysate was reconstructed with 10% ACN in 0.1% FA. The solution was analyzed using a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA) linked to an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). The samples were injected onto an a reversed-phase ACQUITY BEH C_{18} column (2.1 mm×50 mm, 1.7 um particle size, 130 Å pore size, from Waters). Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient conditions for the 35-min separation were: 0-7 min, flushing the column with 5% solvent B; 7-25 min, 5-65%; 25-27 min, 65-99%; 27-30 min, 99%; 30-30.01 min, 99-5%; keeping flushing the column with 5% solvent B till 35 min. The flow from RPLC was directed to the electrospray ionization (ESI) source at a flow rate of 180 µL/min. All mass spectra were collected in positive ion mode.

6.3 Results and Discussion

Figure 6.1 shows the overall workflow of the protein quantification method described in this work. One key procedure in the workflow is the dansylation of proteins. Dansylation is a chemical derivatization method commonly used to react with the primary amines of proteins and peptides to enhance fluorescent signals for

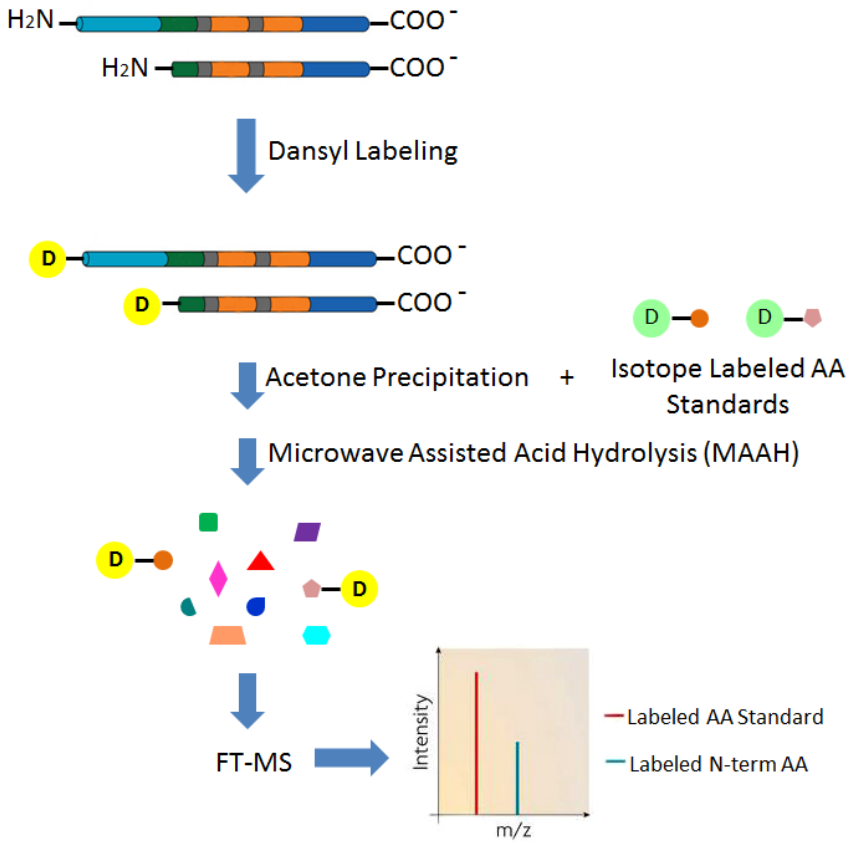


Figure 6.1. Workflow of the dansyl labeling LC-MS method for absolute quantification of protein mixtures containing intact and N-terminal truncated proteins.

N-terminal sequencing[17-18]. More recently, it has also been used as a chemical derivatization strategy for MS-based quantitative analysis of amine- and phenol-containing metabolites[12]. The dansyl modification was found to improve the overall detection sensitivity of metabolites including amino acids by 10- to 1000-fold[12, 19]. In addition, the two carbon atoms in the two methyl group in the dansyl group can be either $^{12}C_2$ (light labeling reagent) or $^{13}C_2$ (heavy labeling reagent), providing a convenient means of generating the isotopic standards for quantification. In this work, amino acid standards were labeled using $^{13}C_2$-dansyl chloride ($^{13}C_2$-Dns-Cl) to create the isotope internal standards. The proteins were derivatized on their N-terminus with $^{12}C_2$-Dns-Cl. After reaction, the dansylated proteins were hydrolyzed by a strong acid with the assistance of microwave irradiation to release the dansyl labeled N-terminal amino acids. The $^{13}C_2$-Dns-amino-acid standards were spiked into the sample as the internal standards for LC-MS analysis. Several experimental parameters in the overall workflow were investigated and optimized. They are described below, followed by the illustration of the analytical performance of this method for protein quantification.

6.3.1 Stability of Dansyl Label under MAAH

To evaluate the stability of the dansyl labeling group under microwave-assisted acid hydrolysis (MAAH), 20 amino acids standards were labeled with $^{12}C_2$- and $^{13}C_2$-dansyl chloride. The 20 $^{12}C_2$-dansyl labeled amino acids were hydrolyzed in 0.1% (v/v) phenol containing 6 M HCl using different irradiation times: 0 min, 0.5 h, 1 h, 1.5 h and 2 h. After MAAH, the hydrolysate solution was mixed with the 20 $^{13}C_2$-dansyl labeled amino acid standards at a calculated molar ratio of 1 to 1. After evaporating the excess acid on Speedvac, the ^{12}C-/^{13}C-mixture was quantified using LC-FTICR-MS. By comparing the peak intensity ratio of the isotopic peak pair of each amino acid at different MAAH time points (the mass difference between the ion pair is 2.0067 Da for a singly labeled amino acid), a plot of peak ratio vs. irradiation time for each amino acid was constructed (see Figure 6.2). Figure 6.2 shows that the recovery rate of each amino acid after MAAH was greater than 90%, indicating that

the dansyl labeling moiety was stable under MAAH for at least up to 2 hr. From the triplicate experiments, the relative standard deviations (RSDs) were found to be less than 5% for all the quantification results (the error bars are purposely not shown in Figure 6.2 for clarity).

For Asn, Gln, Trp and His, these four amino acids have been reported to undergo chemical transformation during the process of acid hydrolysis[20-22]. Asn and Gln can be deaminated to Asp and Glu, respectively, while Trp and His can be oxidized to 5-hydroxytryptophan or ox-indole alanine (Oia) and 2-oxo-histidine (see Supplemental Figure S6.1 for the reaction schemes). These transformations were observed in our experiment in the process of MAAH even with the presence of 0.1% (v/v) phenol as protective reagent in the acid hydrolysis solution[23-24]. Fortunately, the chemical transformation only happened on their amino acid moieties and the dansyl labeling moiety was still stable. To avoid the interference of chemical transformation on quantification of these four amino acids, the $^{13}C_2$-dansyl labeled internal standards were added to the labeled protein solution before acid hydrolysis. As an example, panels (a-c) of Figure 6.3 show the extracted ion chromatograms (EICs) of dansyl histidine obtained under different conditions. As Figure 6.3(a) shows, without MAAH, both $^{12}C_2$- and $^{13}C_2$-dansyl labeled histidine gave similar EIC profiles. Figure 6.3(b) shows that, with MAAH, both labeled histidine were transformed, resulting in no detectable signals. However, by monitoring Dns-oxo-Histidine [Figure 6.3(c)], the chromatographic signals could be recovered for quantification of histidine. Similarly, in the case of tryptophan, as Figure 6.3(d) shows, by monitoring both transformed products, Dns-5-hydroxytryptophan and Dns-ox-indole alanine, the amino acid, tryptophan, after MAAH, could be quantified.

6.3.2 Optimization of MAAH for Protein Hydrolysis

To hydrolyze a protein into amino acids, the conventional acid hydrolysis method is to heat a protein at 110 ℃ with 6 M HCl for 24 h[25]. Releasing of the dansyl labeled N-terminal amino acid from a protein was reported to be faster, only

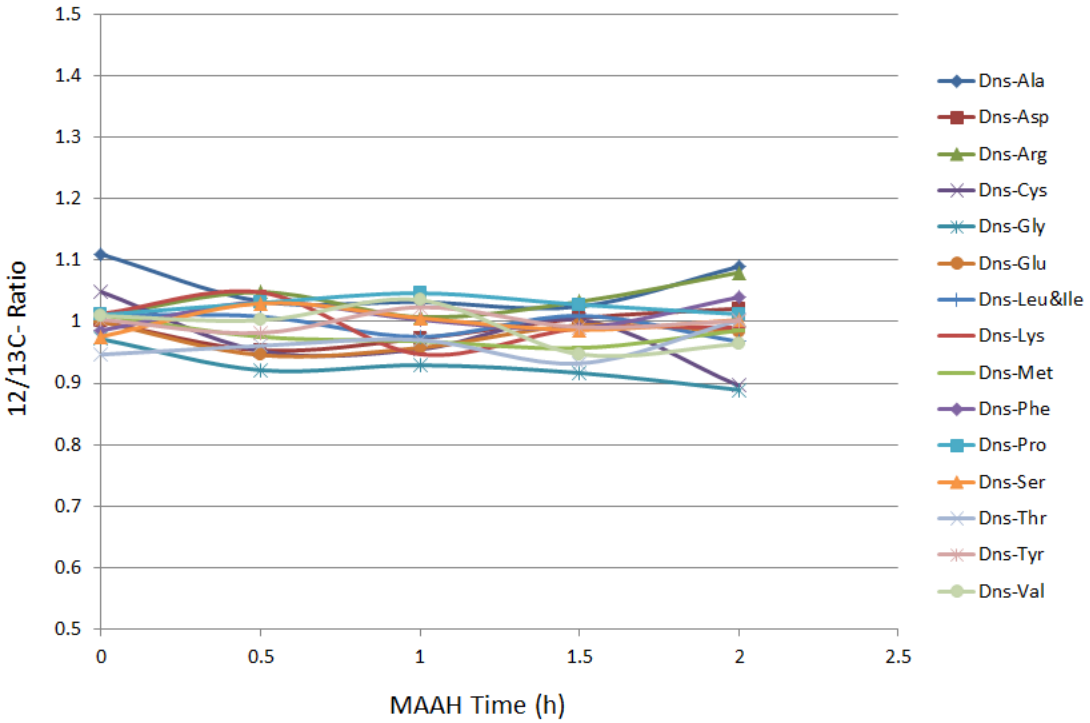


Figure 6.2. The ratio of the peak areas from $^{12}C_2$- and $^{13}C_2$-dansyl labeled amino acid standards as a function of microwave irradiation time. The $^{12}C_2$-labeled standards were hydrolyzed in 6 M HCl with microwave irradiation for 0 h, 0.5 h, 1 h, 1.5 h or 2 h. The resultant solutions were mixed with $^{13}C_2$-dansyl labeled amino acid standards at a calculated ratio of 1:1 and then analyzed by LC-MS.

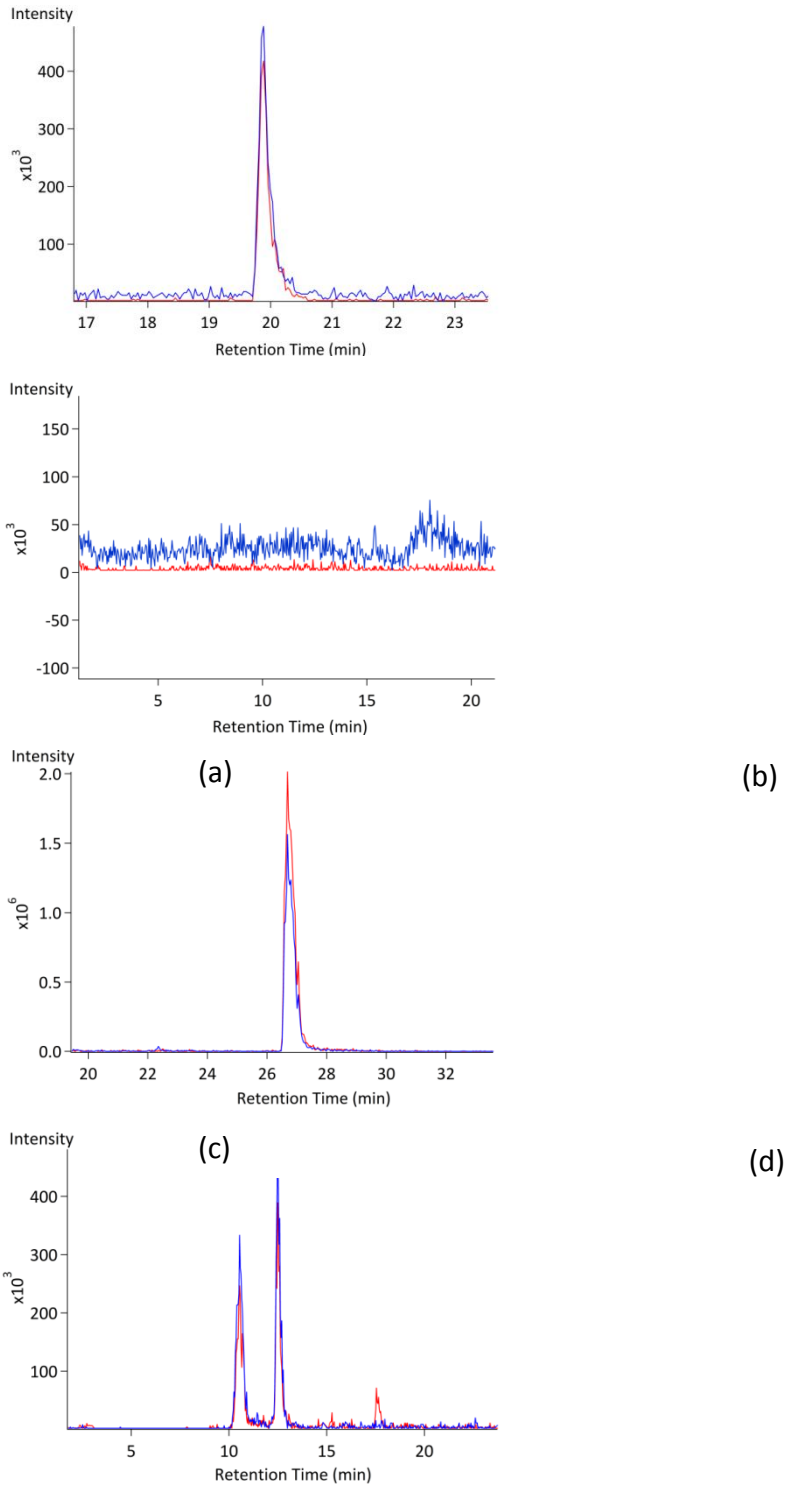


Figure 6.3. Extracted ion chromatograms of (a) Dns-Histidine without MAAH (red trace: $^{13}C_2$-dansyl labeled histidine; blue trace: $^{12}C_2$-dansyl labeled histindine), (b) Dns-Histidine with MAAH, (c) Dns-oxo-Histidine with MAAH, and (d) Dns-5-hydroxytryptophan and Dns-ox-indole alanine with MAAH.

requiring heating at 110 ℃ with 6 M HCl for 4 h to reach >70% conversion of proteins[13]. Although it is faster, complete release of the dansyl labeled N-terminal amino acid by heating is still time consuming. Microwave irradiation can speed up the acid hydrolysis process and has been used in recent years as an alternative to heating for protein hydrolysis[26-27]. It can reduce the hydrolysis reaction time from 24 h to about 30-60 min[27], since the energy transmission is more efficient than traditional heating[28-29].

In our experiment, the effect of microwave irradiation time on the releasing of dansyl labeled N-terminal amino acid from a protein was investigated. Two protein standards, myoglobin (16.7 kDa) and BSA (66.5 kDa), were labeled with dansyl chloride. The labeled proteins were hydrolyzed using different microwave irradiation times (10, 20, 30, 40, 50, 60, 90 and 120 min). $^{13}C_2$-dansyl labeled amino acid standards were spiked in as internal standards before acid hydrolysis. The resultant solutions were analyzed by LC-MS. Representative ion chromatograms of the hydrolysates of labeled myoglobin and BSA mixed with equal molar $^{13}C_2$-dansyl labeled amino acid standards are shown in Figure 6.4. The extracted ion chromatograms (EIC) of the hydrolysates of dansyl labeled myoglobin and BSA after 10 min, 50 min and 90 min MAAH provide us a general idea of the releasing process of the dansyl labeled N-terminal amino acids from proteins (Dns-Glycine from myoglobin and Dns-Arginine from BSA). Using 10 min MAAH [Figure 6.4(a) and (d)], only a small portion of myoglobin or BSA were hydrolyzed to release the N-terminal amino acids. When the hydrolysis time was increased from 10 min to 50 min [Figure 6.4(b) and (e)] or 90 min [Figure 6.4(d) and (f)], almost all proteins were hydrolyzed, resulting in similar signal responses in EICs of the released amino acids and their corresponding $^{13}C_2$-dansyl labeled amino acid standards.

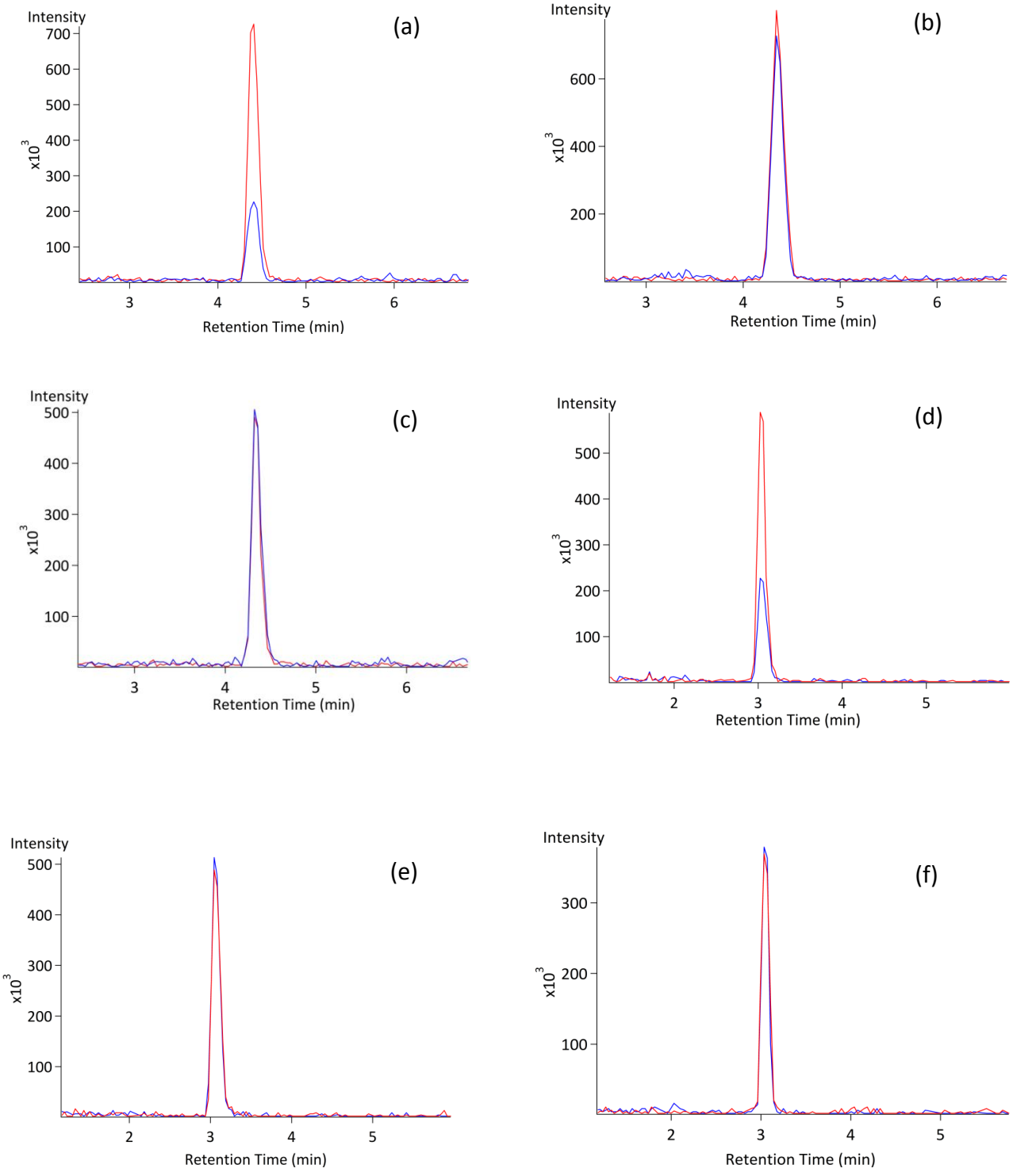


Figure 6.4. Extracted ion chromatograms of Dns-Glycine from MAAH of dansyl labeled myoglobin for (a) 10 min (red trace: $^{13}C_2$-dansyl labeled glycine standard; blue trace: $^{12}C_2$-dansyl labeled glycine from dansyl labeled myoglobin), (b) 50 min, (c) 90 min, and Dns-Arginine from MAAH of dansyl labeled BSA for (d) 10 min, (e) 50 min, and (f) 90 min.

The releasing rates of the N-terminal amino acid from each protein at different microwave irradiation times are shown in Figure 6.5. After only 10 min of MAAH, both proteins could have ~40% dansyl labeled N-terminal amino acids released. The release rate could reach 90% to 103% after 50 min of acid hydrolysis with microwave irradiation. Interestingly, for the smaller protein, myoglobin, the releasing rate was lower compared to that of BSA between 20 and 50 min. To ensure all proteins with different molecular masses reach the maximum release of the N-terminal amino acids, 50 min MAAH appears to be sufficient.

6.3.3 LOD and LOQ of the Absolute Protein Quantification Method

To determine of the limit of detection (LOD) and limit of quantification (LOQ), a set of quantitative analyses were carried out. In this case, 20 amino acid standards after dansylation were subjected to hydrolysis in 6 M HCl with 0.1% (v/v) phenol by MAAH for 50 min. The dansylated amino acid standard solutions were diluted to different concentrations before analyzing by LC-MS. Based on the peak intensities of each dansylated amino acid in the mass spectra obtained from a series of concentrations, calibration curves of the dansylated amino acid standards were constructed. The LOD and LOQ for each standard were then calculated and the results are shown in Table 6.1. Table 6.1 shows that, even with a conventional micro-flow LC-MS, dansylated amino acids can be quantified down to a few picomoles per analysis.

It should be noted that, for quantifying Asn, Gln, Trp and His, their calibration curves were built on the signal responses of their chemical transformed products on FT-MS. In brief, Asn, Gln and His were calibrated on peak intensities of Asp, Glu and 2-oxo-histidine, respectively, after MAAH. In the case of Trp, because it would be oxidized into two products simultaneously from acid hydrolysis, its responses on MS with different concentrations were determined by summing the peak intensities of the two oxidized products, 5-hydroxytryptophan and ox-indole alanine (Oia) [see Figure 6.3(d)].

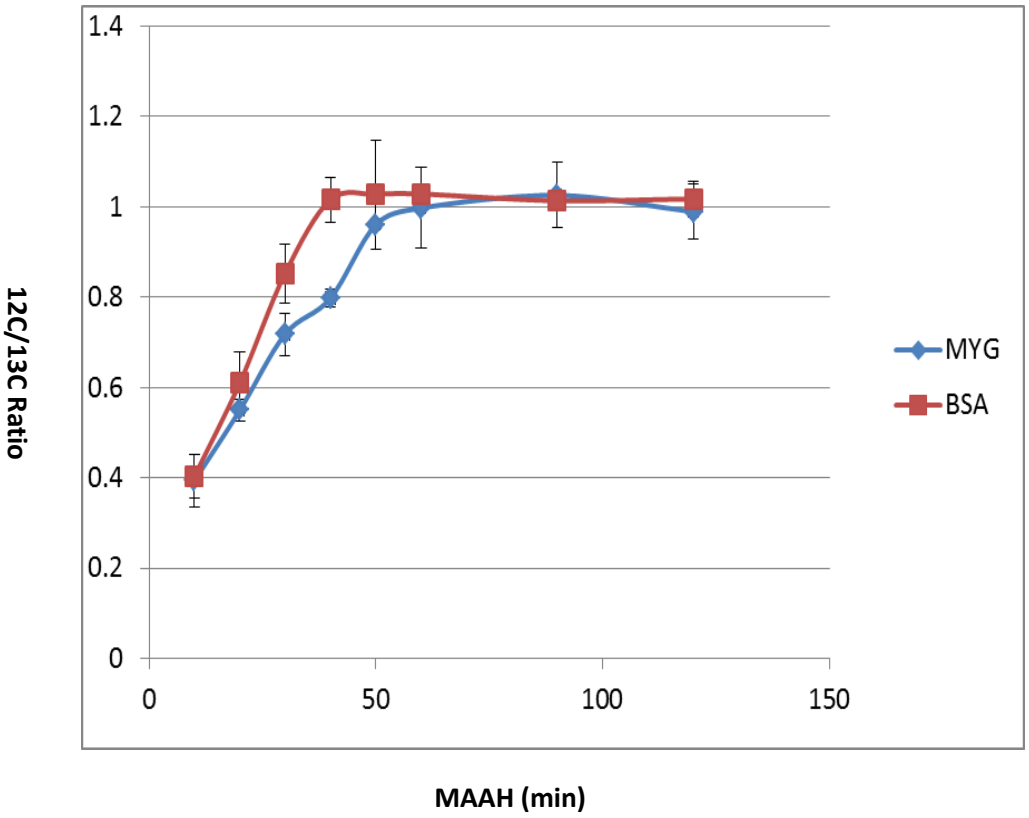


Figure 6.5. The ratio of the peak areas from the $^{12}C_2$-dansyl labeled amino acid released from dansyl labeled myoglobin or BSA and its corresponding $^{13}C_2$-dansyl labeled amino acid standard as a function of microwave irradiation time.

Table 6.1: The LODs and LOQs of 20 dansylated amino acid standards after MAAH on LC-coupled FT-MS.

Amino Acid	Regression Equation	R Square	LOD (pmole)	LOQ (pmole)
Dns-Ala	y = 63583x - 291907	0.9968	4.58	15.25
Dns-Asp	y = 310159x - 1E+06	0.9985	4.47	14.88
Dns-Arg	y = 5731.4x - 90289	0.9998	4.98	16.61
Dns-Cys	y = 43321x - 657207	0.9925	8.26	27.53
Dns-Glu	y = 51999x - 324268	0.9994	6.47	21.57
Dns-Gly	y = 127174x - 767737	0.9985	2.62	8.73
Dns-2-oxo-His	y = 58606x - 3E+06	0.9958	4.54	15.14
Dns-Leu&Ile	y = 58072x - 216752	0.9990	8.29	27.62
Dns-Lys	y = 23456x - 1E+06	0.9979	2.40	7.99
Dns-Met	y = 81173x - 770873	0.9952	9.85	32.82
Dns-Phe	y = 58024x - 310636	0.9996	2.84	9.46
Dns-Pro	y = 649280x - 1046360	0.9995	3.09	10.29
Dns-Ser	y = 104509x - 954997	0.9953	3.13	10.45
Dns-Thr	y = 127656x - 2E+06	0.9961	0.87	2.88
Dns-Trp (oxidized products)	y = 1264.6x - 25524	0.9965	14.62	48.74
Dns-Tyr	y = 2109.4x - 36440	0.9912	2.09	6.96
Dns-Val	y = 42334x - 284191	0.9986	5.93	19.75

*The linear dynamic range for each amino acid was from LOQ to 1250 pmol which was the highest sample concentration tested. This range is sufficient for protein quantification and the upper limit of the quantification range may be higher.

6.3.4 Accuracy and Precision

Lysozyme (14.3 kDa), myoglobin (16.7 kDa), β-casein (23.6 kDa) and BSA (66.5 kDa) were selected as protein standards to evaluate the accuracy and precision of this quantification method. These four proteins were labeled at their N-terminal free amines with $^{12}C_2$-dansyl chloride and hydrolyzed under the optimized MAAH condition as described above. Lysozyme has lysine as its N-terminal amino acid, while myoglobin has glycine, β-casein has aspartic acid and BSA has arginine at its N-terminus. After dansyl labeling and MAAH, $^{12}C_2$-dansyl-lysine, $^{12}C_2$-dansyl-glycine, $^{12}C_2$-dansyl-aspartic acid and $^{12}C_2$-dansyl-arginine were present in the protein hydrolysate which could be quantified to determine the correspondent protein concentrations. Again, $^{13}C_2$-dansyl labeled amino acids standards were spiked in the solution as internal standards for quantification.

Note that lysine has two free amine groups in its structure if not modified. When lysine is present in the middle of the protein sequence, only the amine group on its side chain is free and can be labeled with dansyl chloride. However, if lysine is at the N-terminus of a protein, like lysozyme, both of its free amine groups can be dansylated, resulting in two dansyl groups attached to lysine. Thus, the singly dansyl labeled lysine released from the middle of a protein sequence can be readily differentiated from the doubly labeled lysine from the N-terminus.

Figure 6.6 shows three representative mass spectra obtained from FT-MS of the protein hydrolysates displaying the peak pairs: one from the N-terminal amino acids of proteins and the other from the amino acid standards. The relative intensity of each peak pair, along with the absolute concentration information of the $^{13}C_2$-dansyl labeled amino acids standards can be used to determine the absolute concentration of a protein. Table 6.2 shows the protein amounts determined from this analysis, compared to the standard sample amount used. The recovery ratio, defined as the ratio of the quantified amount vs. the starting amount calculated from the protein concentration, was found to be ranging from about 94% to 112% with the relative

standard deviation ranging from 2.1% to 8.3%. These results illustrate that, despite several steps involved in the overall workflow, good accuracy and precision can be obtained using the method described.

Table 6.2: Recovery Rate of protein standards with dansyl labeling MS-based quantification method.

Protein Standards	Dansylated N-Terminal Amino Acid	Starting Amount (nmole)	Quantification Results (nmole)	RSD (%)	Recovery Rate (%)
Myoglobin	Dns-Gly	10	9.41	4.7%	94.1%
β- Casein	Dns-Arg	10	10.31	2.1%	103.1%
Lysozyme	Dns-Lys	10	10.87	8.3%	108.7%
BSA	Dns-Asp	10	11.19	2.7%	111.9%

6.3.5 Quantification of mCherry and N-terminal truncated mCherry

To demonstrate the application of the method for quantifying intact and N-terminal truncated proteins, a mixture of a fluorescent protein, mCherry, and its truncated form, were analyzed. Each protein was dissolved in 8 M urea and its concentration was determined by their absorbance at 587 nm, with extinction coefficient of 72,000 $M^{-1}cm^{-1}$ [30]. mCherry and N-truncated mCherry were mixed at different ratios: 10:1, 5:1, 1:1, 1:5, and 1:10. The protein mixture was labeled at their N-terminus by dansylation, followed by protein acetone precipitation. The protein precipitates along with the $^{13}C_2$-dansyl labeled amino acid standards were subjected to MAAH and the resultant hydrolysates were analyzed by LC-MS. In this case, the N-terminal of mCherry is methionine while the truncated form has aspartic acid as its N-terminal (see Supplemental Note N1 for the sequences). After dansyl labeling and MAAH, mCherry released dansyl-methionine and N-truncated mCherry had dansyl-aspartic acid cleaved from its terminus. Comparing to the isotope labeled internal standards, the absolute amounts of mCherry and N-truncated mCherry were determined from the ratios between ^{12}C-/^{13}C-dansyl-methionine and ^{12}C-/^{13}C- dansyl-

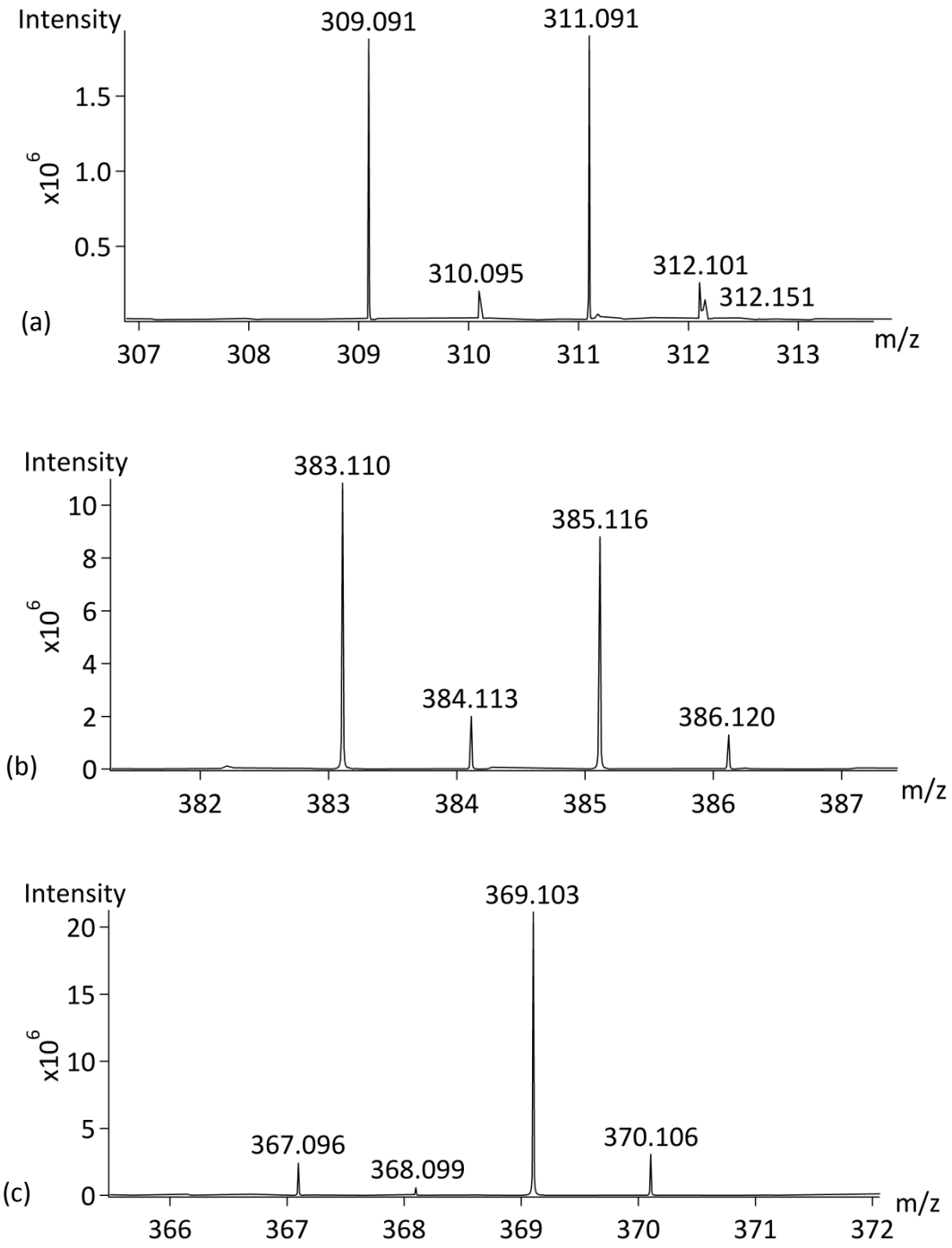


Figure 6.6. MS peak pairs of (a) $^{12}C_2$-Dns-Gly released from the N-terminus of myoglobin and $^{13}C_2$-Dns-Gly standard, (b) $^{12}C_2$-Dns-Met released from the N-terminus of mCherry and $^{13}C_2$-Dns-Met standard, and (c) $^{12}C_2$-Dns-Asp released from the N-terminus of N-truncated mCherry and $^{13}C_2$-Dns-Asp standard.

aspartic acid. The relative amounts of N-truncated mCherry and intact mCherry from the different samples were calculated and the results are shown in Table 6.3. The replicate experimental results indicate that the recovery rate ranges from 99% to 106% with a precision of better than 3.1%.

Table 6.3: Quantification of mCherry and N-truncated mCherry mixture.

mCherry/N-Truncated mCherry					
Theoretical Ratio	0.1	0.2	1	5	10
Observed Ratio	0.102	0.196	0.995	4.96	10.63
RSD (%)	3.10%	2.20%	1.50%	0.80%	1.70%
Recovery Rate (%)	102.00%	98.00%	99.50%	99.20%	106.30%

6.4 Conclusions

We have developed a relatively rapid method for absolute quantification of protein mixtures containing intact and N-terminal truncated proteins based on LC-MS quantification of the dansyl labeled N-terminal amino acids released from the proteins by MAAH after the proteins are labeled with dansyl chloride. The total analysis time required is less than 8 hrs. By analyzing several protein standards and a mixture of different amounts of mCherry and N-terminal truncated mCherry, it was demonstrated that the quantification error of the method was less than 12% with a precision of better than 8%. This method should be generally applicable to quantify individual proteins in a protein mixture except in cases where the two or more proteins, after acid hydrolysis, generate the same N-terminal amino acids or the N-terminus of a protein is blocked with a modification group.

6.5 Acknowledgements

This work was funded by Genome Canada, the Canada Research Chairs program, and Sanofi. We thank Mr. Ruokun Zhou for assistance in preparing the isotope reagents.

6.6 Literature Cited

1. Van Damme, P.; Arnesen, T.; Gevaert, K., Protein alpha-N-acetylation studied by N-terminomics. *Febs J.* **2012,** *278* (20), 3822-3834.

2. Chevigne, A.; Fievez, V.; Schmit, J. C.; Deroo, S., Engineering and screening the N-terminus of chemokines for drug discovery. *Biochem. Pharmacol.* **2012,** *82* (10), 1438-1456.

3. Keller, U. A. D.; Overall, C. M., CLIPPER: an add-on to the Trans-Proteomic Pipeline for the automated analysis of TAILS N-terminomics data. *Biological Chemistry* **2012,** *393* (12), 1477-1483.

4. Kobayashi, R.; Patenia, R.; Ashizawa, S.; Vykoukal, J., Targeted mass spectrometric analysis of N-terminally truncated isoforms generated via alternative translation initiation. *FEBS Lett.* **2009,** *583* (14), 2441-2445.

5. Gite, S.; Lim, M.; Carlson, R.; Olejnik, J.; Zehnbauer, B.; Rothschild, K., A high-throughput nonisotopic protein truncation test. *Nat. Biotechnol.* **2003,** *21* (2), 194-197.

6. Petrakis, E. C.; Trantakis, I. A.; Kalogianni, D. P.; Christopoulos, T. K., Screening for Unknown Mutations by a Bioluminescent Protein Truncation Test with Homogeneous Detection. *J. Am. Chem. Soc.* **2010,** *132* (14), 5091-5095.

7. Galvez, S.; Hodges, M.; Bismuth, E.; Samson, I.; Teller, S.; Gadal, P., Purification and characterization of a fully active recombinant tobacco cytosolic NADP-dependent isocitrate dehydrogenase in Escherichia coli: evidence for a role for the N-terminal region in enzyme activity. *Arch Biochem Biophys* **1995,** *323* (1), 164-8.

8. Vafai, A., Antibody-binding sites on truncated forms of varicella-zoster virus gpI(gE) glycoprotein. *Vaccine* **1994,** *12* (14), 1265-9.

9. Verrall, S.; Hall, Z. W., The N-terminal domains of acetylcholine receptor subunits contain recognition signals for the initial steps of receptor assembly. *Cell* **1992,** *68* (1), 23-31.

10. Christiansen, C.; Skotland, T., Changes of protein solutions during storage: a study of albumin pharmaceutical preparations. *Biotechnol. Appl. Biochem.* **2010,** *55*, 121-130.

11. Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J., Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2005,** *2* (8), 587-589.

12. Guo, K.; Li, L., Differential (12)C/(13)C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome. *Analytical Chemistry* **2009,** *81* (10), 3919-3932.

13. Gros, C.; Labouess.B, Study of Dansylation Reaction of Amino Acids Peptides and Proteins. *European Journal of Biochemistry* **1969,** *7* (4), 463-&.

14. Wang, N.; Li, L., Reproducible Microwave-Assisted Acid Hydrolysis of Proteins Using a Household Microwave Oven and Its Combination with LC-ESI MS/MS for Mapping Protein Sequences and Modifications. *J. Am. Soc. Mass Spectrom.* **2011,** *21* (9), 1573-1587.

15. Giberson, R. T.; Demaree, R. S., Microwave Fixation - Understanding the Variables to Achieve Rapid Reproducible Results. *Microscopy Research and Technique* **1995,** *32* (3), 246-254.

16. Kok, L. P.; Boon, M. E.; Smid, H. M., The Problem of Hot-Spots in Microwave Equipment Used for Preparatory Techniques - Theory and Practice. *Scanning* **1993,** *15* (2), 100-109.

17. Walker, J. M., The dansyl method for identifying N-terminal amino acids. *Methods Mol Biol* **1994,** *32*, 321-8.

18. Walker, J. M., The Dansyl-Edman method for peptide sequencing. *Methods Mol Biol* **1994,** *32*, 329-34.

19. Chiappetta, G.; NDiaye, S.; Demey, E.; Haddad, I.; Marino, G.; Amoresano, A.; Vinh, J., Dansyl-peptides matrix-assisted laser desorption/ionization mass spectrometric (MALDI-MS) and tandem mass spectrometric (MS/MS) features improve the liquid chromatography/MALDI-MS/MS analysis of the proteome. *Rapid Communications in Mass Spectrometry* **2010,** *24* (20), 3021-3032.

20. Ji, J. A.; Zhang, B. Y.; Cheng, W.; Wang, Y. J., Methionine, Tryptophan, and Histidine Oxidation in a Model Protein, PTH: Mechanisms and Stabilization. *Journal of Pharmaceutical Sciences* **2009,** *98* (12), 4485-4500.

21. Mason, B. D.; McCracken, M.; Bures, E. J.; Kerwin, B. A., Oxidation of Free L-histidine by tert-Butylhydroperoxide. *Pharmaceutical Research* **2010,** *27* (3), 447-456.

22. Schoneich, C., Mechanisms of metal-catalyzed oxidation of histidine to 2-oxo-histidine in peptides and proteins. *Journal of Pharmaceutical and Biomedical Analysis* **2000,** *21* (6), 1093-1097.

23. Fountoulakis, M.; Lahm, H. W., Hydrolysis and amino acid composition analysis of proteins. *Journal of Chromatography A* **1998,** *826* (2), 109-134.

24. Munoz, A.; Kral, R.; Schimmel, H., Quantification of protein calibrants by amino acid analysis using isotope dilution mass spectrometry. *Analytical Biochemistry* **2011,** *408* (1), 124-131.

25. Hirs, C. H. W.; Stein, W. H.; Moore, S., The Amino Acid Composition of Ribonuclease. *Journal of Biological Chemistry* **1954,** *211* (2), 941-950.

26. Kroll, J.; Rawel, H.; Krock, R., Microwave digestion of proteins. *Zeitschrift Fur Lebensmittel-Untersuchung Und-Forschung a-Food Research and Technology* **1998,** *207* (3), 202-206.

27. Zhang, X.; Yang, L.; Mester, Z., Determination of amino acids in selenium-enriched yeast by gas chromatographyâ€“mass spectrometry after microwave assisted hydrolysis. *Analytica Chimica Acta* **2012,** *744* (0), 54-59.

28. Chen, S. T.; Chiou, S. H.; Chu, Y. H.; Wang, K. T., Rapid Hydrolysis of Proteins and Peptides by Means of Microwave Technology and Its Application to Amino-Acid-Analysis. *International Journal of Peptide and Protein Research* **1987,** *30* (4), 572-576.

29. Marconi, E.; Panfili, G.; Bruschi, L.; Vivanti, V.; Pizzoferrato, L., Comparative-Study on Microwave and Conventional Methods for Protein Hydrolysis in Food. *Amino Acids* **1995,** *8* (2), 201-208.

30. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nature Biotechnology* **2004,** *22* (12), 1567-1572.

Chapter 7 - Conclusions and Future Work

Quantitative proteomics techniques have made rapid progress in recent years to enumerate and quantify proteins expressed in a biological sample. Identification and quantification of differentially expressed proteins of cells, tissues and biofluids are increasingly being recognized as a key objective in proteomics research. Mass spectrometry (MS) is the most comprehensive and versatile tool in large-scale proteomics. However, due to the complexity of most biological samples, it's still challenging to detect the proteome very comprehensively. First, proteins in biological sample usually present in a large dynamic range of concentration. Signals from low abundant proteins are difficult to detect due to the suppression of high abundant proteins. Second, inherent limitations of biological MS^1 require several different approaches to protein analysis. Implementation of these strategies (e.g., sample preparation, front-end separation and chemistry derivatization) differs depending on the sample complexity and the goals of the analysis. Therefore, the overall goal of my thesis work is to develop and optimize MS technique for quantitative proteome or protein analysis.

In Chapter 2, a microbore LC/UV method was developed with step gradient elution to desalt and, at the same time, rapidly quantify the total peptide amount in proteomic digest samples. By coupling with LC-ESI MS/MS, this method was applied to identify the proteome of a limited number of cells. In this method, RPLC columns (C18, particle size: 30 µm, pore size: 300 Å, length: 50 mm) with different inner diameters: 4.6-mm, 2-mm and 1-mm were firstly calibrated with a four-protein mixture digest (myoglobin, cytochrome C, lysozyme and ß-casein) on an Agilent 1100 LC-UV system. The flow rates of RPLC columns with 4.6-mm, 2-mm and 1-mm inner diameters were 1 mL/min, 200 µL/min and 100 µL/min, respectively. Peptides were eluded from columns by 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. The UV absorbance of eluted peptides was detected at 214 nm. After building the calibration curves for

the columns with different inner diameters, the column with 1.0 mm inner diameter was found to be able to quantify peptide amount from 0.03 μg to 0.3 μg, 0.6 μg to 5 μg with linear regression. This linear dynamic range was shown to be suitable for quantifying peptides extracted from limited cell samples, like samples < 1000 cells.

For this type of few-cell samples, sample loss of any type in sample preparation can be detrimental to the proteome coverage, since the amount of sample will be limited and will often not meet the optimal sample amount required for peptide sequencing in LC-MS/MS (e.g., < 1 μg). Thus it's very necessary to evaluate sample loss in preparation and optimize LC-MS/MS conditions for the limited amount samples before analysis. The LC-UV quantification method we developed is nondestructive and good for peptide sample evaluation before MS analysis. Using this method, we quantified the peptides generated from MCF-7 cell extracts with limited cell numbers, like 250, 500 and 1000. Based on the quantified results of extracted peptides, LC-MS analysis gradient was optimized for each cell samples, from 90 min gradient to 150 min. In MS analysis, this method was demonstrated to be capable of identifying an average of 103±23, 138±16, and 269±13 proteins from 250, 500, and 1000 cells, respectively. This LC-UV method should be useful in proteome profiling of small numbers of cells for disease diagnosis and prognosis.

Besides the LC-UV method used for sample integrity evaluation in proteomic sample preparation process, in order to further optimize the proteomic sample preparation process, in Chapter 3, we tested the compatibility of a stable isotopic labeling reaction, 2-MEGA, with commonly used cell lysis buffers, salts and detergents for protein solubilization.

2-MEGA (dimethylation after guanidation) was developed and demonstrated by our group[2-3] to be a competent alternative for commercial isotopic peptide labeling reagents. However, to apply this method to a wide range of samples, we need to determine whether 2-MEGA reaction is compatible with cell lysis buffers, salts and detergents usually used for proteome extraction and solubilization. In Chapter 3, *E.*

coli samples were lysed with CelLytic M™ or TM buffer. Then the *E. coli* sampls were divided into two vials, one of which was precipitated with acetone incubation at -80℃ to remove cell lysis buffers. Samples were labeled with 2-MEGA chemistry, and analyzed by LC-MS. The peptides identified were checked to see whether the labeling happened on the right position. Correctly labeling peptides and correct labeling rates for each sample were compared. In these cases, with or without cell lysis buffer, the correct labeling efficiency of 2-MEGA on peptides was over 94%. And peptides identified from each lysis buffer or sample before or after acetone precipitation had larger than 60% overlaps, indicating that similar sample components were obtained in the sample preparations.

The same type of comparison experiments were carried out with seven commonly used salts and detergents for protein solubilization. *E. coli* lysate was precipitated with acetone incubation, and redissolved with NH_4HCO_3, MeOH, Urea, SDS, Rapigest, ProteaseMax and AALS. The dissolved proteome samples were digested, labeled by 2-MEGA reaction, and purified with necessary separation steps before MS analysis. The peptide identification results show similar 2-MEGA labeling efficiency (~95%) in the presence of the salts and detergents, except urea. The correct labeling rate for urea was only around 92%. The most frequently happened side reaction in 2-MEGA labeling was found to be guanidination on peptides' N-terminal in all cases. In most solutions this side reaction was over 2.5% in total peptide identities. However, in urea, another mis-reaction happened with high frequency, that's missing guanidination on lysine (K). In urea solution, the percentage of missing guanidination on lysine (K) went up to 1.85%, while in other solutions, it's less than 0.2%. By looking closer to modifications on identified peptides, we found carbamylation was one reason of the lower correct 2-MEGA labeling efficiency in urea peptide solutions. As the urea molecules may react with the free amine group on the side chain of K and N-terminus, carbamylation blocked the reaction spots for 2-MEGA labeling, thus causing missing labeling or side reactions.

Till now, we demonstrated 2-MEGA labeling could be compatible with the tested

two cell lysis buffers and six salts and detergents, except urea. In Chapters 4 and 5, this stable isotope labeling chemistry was applied in real biological samples to find potential biomarkers of diseases.

In Chapter 4, we sought to identify proteomic changes accompanying *ERBB2* gene amplification in Michigan Cancer Foundation-7 (MCF-7) human breast cancer cells in order to i) identify new biomarkers associated with the HER2 phenotype, ii) gauge the magnitude of the proteomic changes triggered by amplification of this single gene, and iii) derive a better understanding of the downstream biological changes triggered by HER2 overexpression. We performed quantitative proteomic analysis of wild-type (parental) and overexpressing (stably transfected) MCF-7 cell lines using offline two-dimensional liquid chromatography (2D-LC) MS coupled with the 2-MEGA labeling quantitative proteomic approach.

In our experiments, 2455 unique proteins were quantified for the bio-triplicate samples and 1278 proteins were identified to be differentially expressed, with 231 proteins exhibiting the same expression level changes in the bio-triplicates of the two clones. The multiplicity of proteins differentially expressed in these isogenic cell lines as a result of transfection of additional copies of the HER2 gene suggests that the biological effects of HER2 are mediated by multiple pathways and various effector molecules. Of the 231 proteins, five proteins were selected and validated by western blotting. The direction of change in protein levels observed in the MCF-7 cell lines was recapitulated in clinical samples, as determined by immunohistochemical analysis, for four of the five proteins. While these biomarkers did not demonstrate prognostic value in this study, the clinical population received aggressive adjuvant therapy including taxane and anthracycline chemotherapy, endocrine therapy for estrogen receptor positive disease, and in the majority of the HER2 population, adjuvant anti-HER2 therapy. Nonetheless, identification of the proteins by which HER2 positivity confers its biological effects may provide a better understanding of the mediators of clinically aggressive disease and potential mechanisms for resistance to anti-HER2 therapies.

In Chapter 5, the proteomic change at different stages of deep tissue injury (DTI) was studied in pig model. In our project, we envisioned specific biomarkers to be associated early in DTI progression prior to the visual detection, and a simple blood test would confer the ability to monitor recovery, compare efficacy of treatments, and provide prophylactic monitoring method to help prevent DTI. To achieve this goal, we performed quantitative proteomic analysis of pig plasma samples: DTI with IES treatment, DTI without IES treatment, and intact pig plasma. In the analysis, we applied offline two-dimensional liquid chromatography (2D-LC) MS with the 2-MEGA method.

In total, 126 proteins were quantified from bio-replicate pig plasma samples. One reason for the limited quantification was ion suppression from highly abundant proteins, like albumin, in plasma samples. Although before analyzed on MS, highly abundant proteins were depleted from the plasma sample by affinity separation, the depletion cannot be 100% and the presence of these highly abundant proteins left in plasma sample would still cause severe ion suppression on low abundant proteins, making them difficult to be ionized, detected and quantified. Another challenge was the lack of a completed porcine proteome database. These may be two major reasons for the limited quantification in pig plasma. However, from the limited number of quantified proteins, we found the proteins with differential expression level change in the development of disease or IES treatment, and made a biomarker candidate list based on bioinformatics study.

In the future, more bioinformatics study, biological validation and evaluation of their prognostic values will be needed, based on the biomarker candidate list, in order to find new biomarkers associated with DTI.

In proteomic quantitative analysis, absolute quantification is another research topic besides comparative quantification. In Chapter 6, a novel MS-based protein absolute quantification method for N-terminal truncated protein and intact protein mixtures is described.

The N-termini amino acids of proteins are important structure units to maintain the bio-function, localization, and interaction networks of proteins. In biological conditions, N-termini amino acids could be cleaved from a normal protein due to many reasons, e.g., proteolysis, resulting in the loss of the protein structure integrity. As a result, the intended biological function of the protein can be affected or even terminated. Thus, the ability to quantify the N-terminal truncated forms of certain proteins is of great importance for clinical, biological and protein production industry. As described in Chapter 6, a dansylation reaction was applied as chemical labeling strategy to label the free amine at the N-terminal of proteins. The N-terminal labeled proteins were hydrolyzed into amino acids by microwave-assisted acid hydrolysis (MAAH). Then the protein hydrolysate was separated and analyzed on LC-MS with dansyl labeled amino acids as internal standards. By comparing the peak intensity of dansyl labeled N-termini amino acids released from protein in MAAH and the dansyl labeled amino acid standards, for which the concentration was known, the absolute concentration of dansyl labeled proteins could be calculated from the concentration of the released labeled amino acids. The overall sample preparation process took only a few hours.

Using this method, four proteins, myoglobin (MYG, 16.7 kDa), lysozyme (LYSC, 14.3 kDa), β-casein (23.6 kDa) and BSA (66.5 kDa), were quantified in a mixture. The error of protein quantification could be less than 6%. In addition to the ability to absolutely quantify different proteins in a mixture, we also demonstrated that this method could be applied to quantify the N-terminal truncated proteins in a mixture containing the intact one. In our experiment, a recombinant protein, mCherry, and its N-terminal truncated form, N-truncated mCherry with 31 amino acids cleaved from the N-termini, were analyzing using our method. These two proteins were mixed with different ratios and quantified with good precision and accuracy.

In summary, although there are still many issues needed to be addressed in proteomic quantitative analysis, the techniques developed or optimized in my thesis work improved sample preparation, chemical labeling reaction and applied these

technique in complex biological samples, aiming to find potential biomarkers for diseases. The comparative and absolute protein quantification techniques developed in this thesis work may also have the potential to be applied into quantifying a wide range of proteome samples and protein products in a more comprehensive and efficient way.

Literature Cited

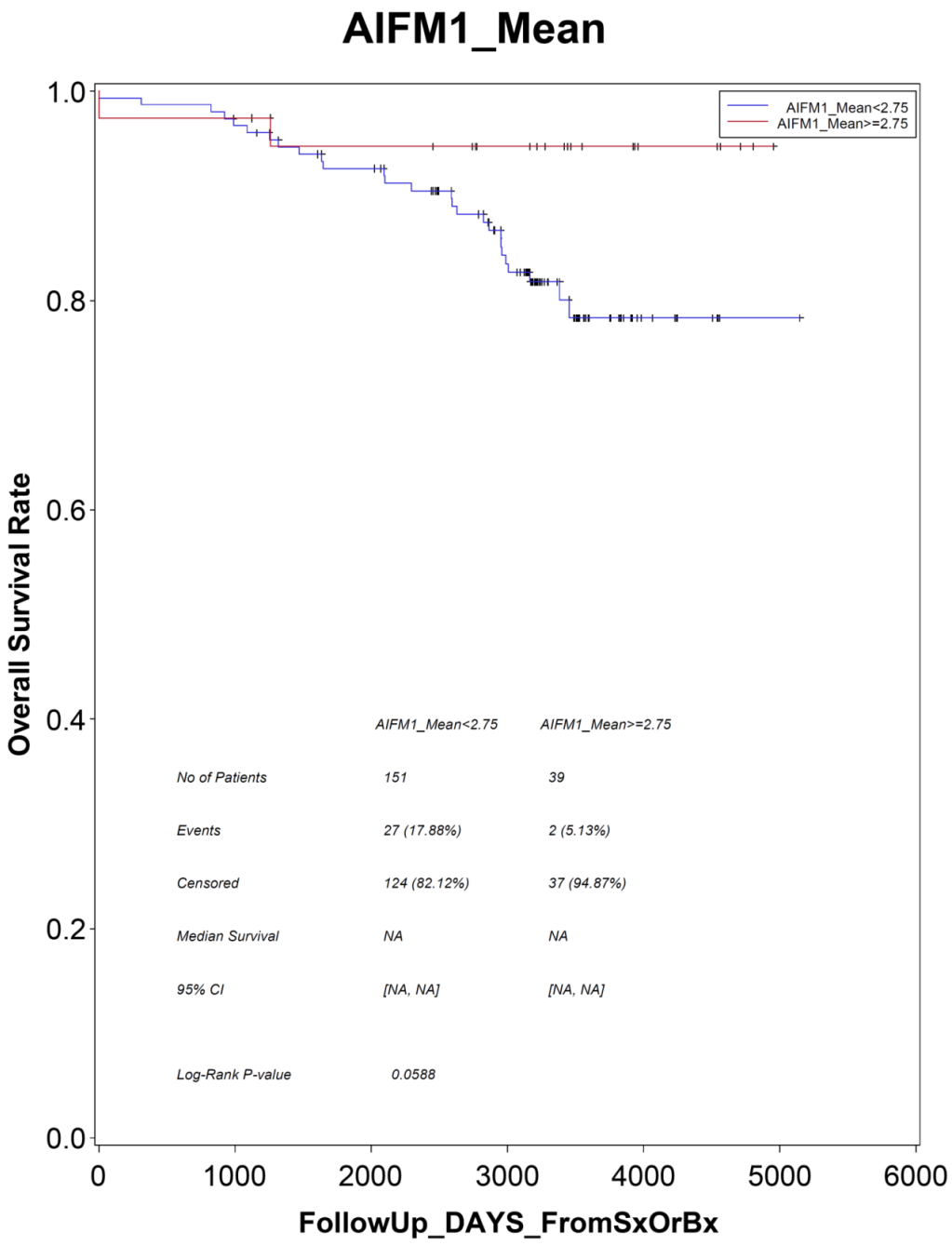
1. Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S., Global analysis of protein expression in yeast. *Nature* **2003,** *425* (6959), 737-41.

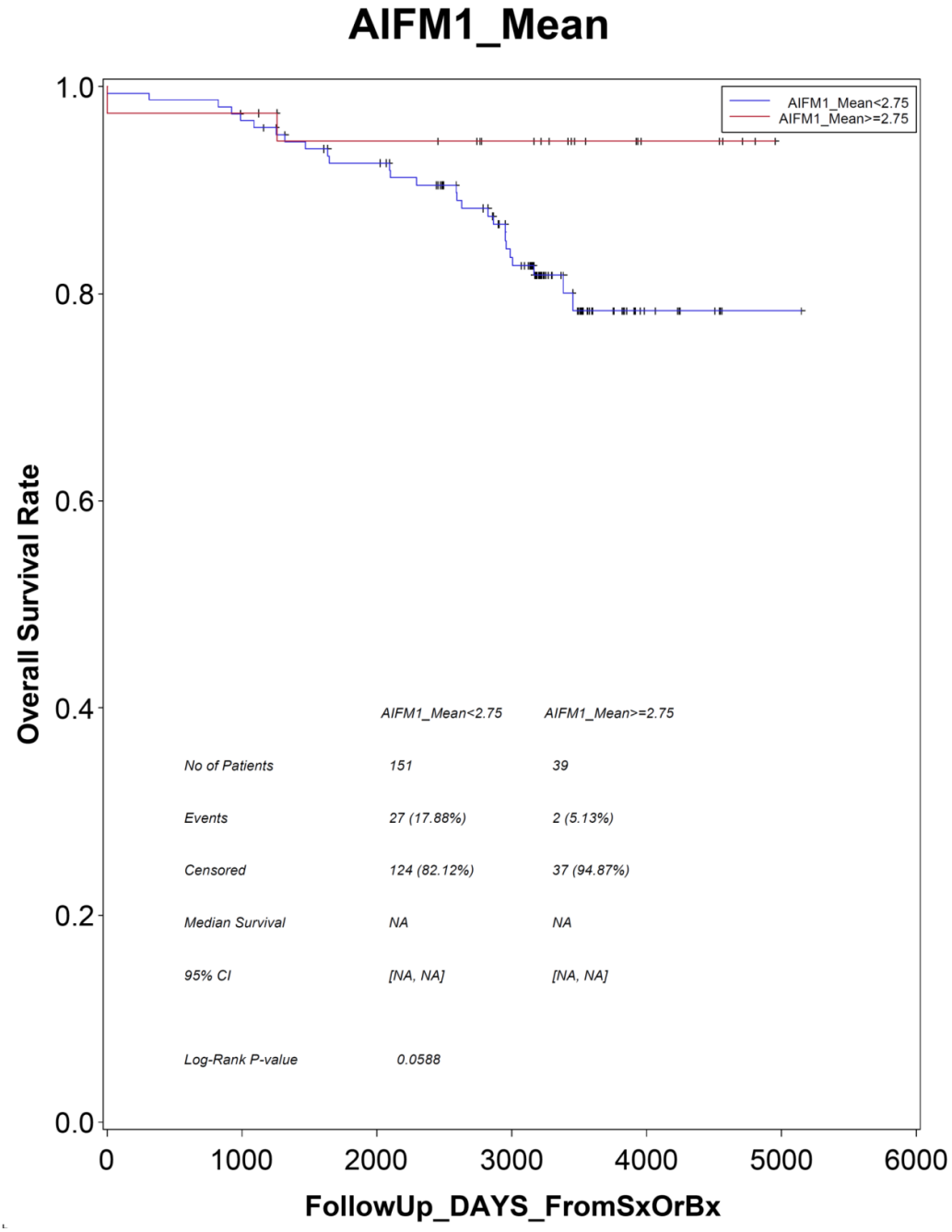
2. Ji, C.; Li, L., Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS. *J Proteome Res* **2005,** *4* (3), 734-42.

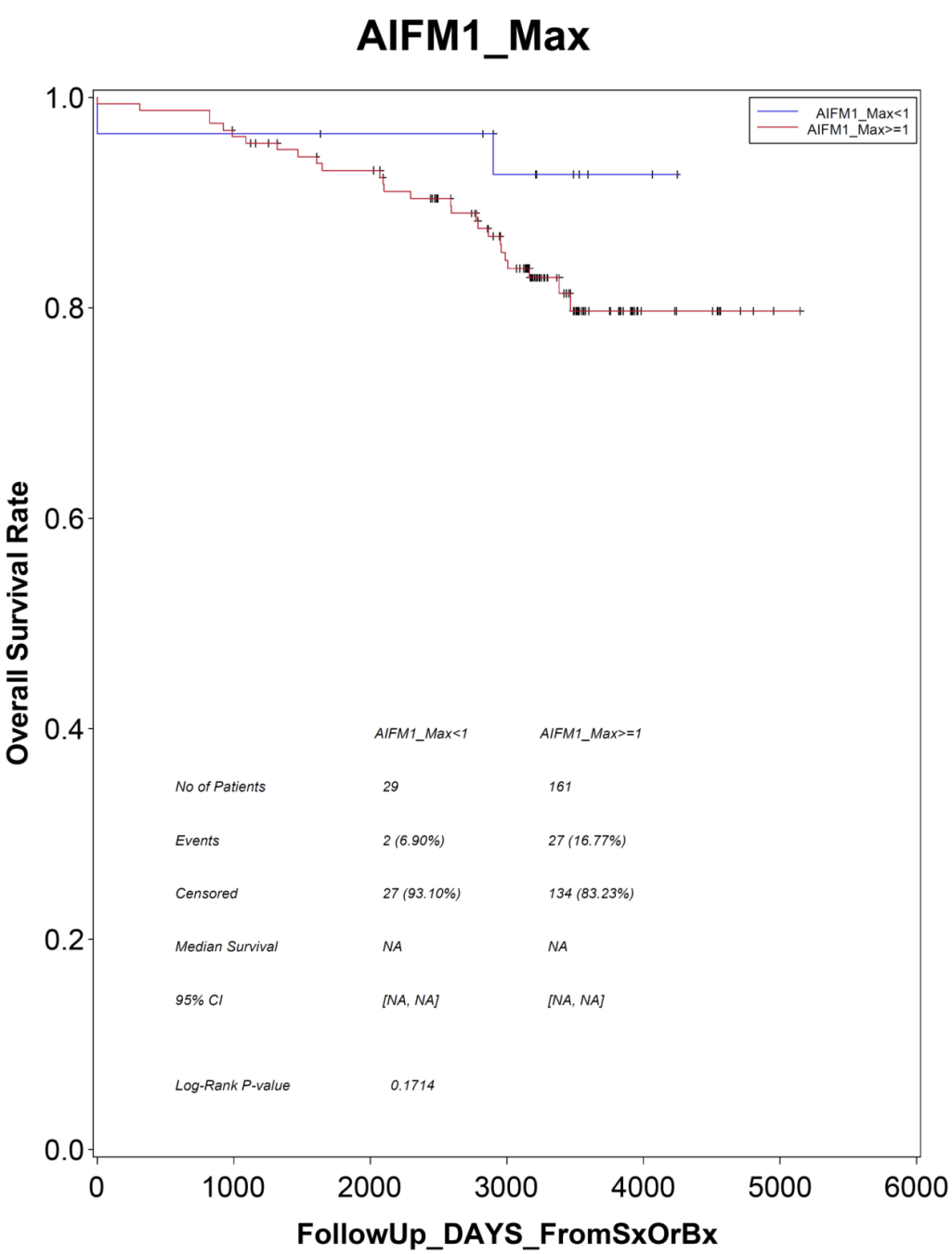
3. Ji, C.; Zhang, N.; Damaraju, S.; Damaraju, V. L.; Carpenter, P.; Cass, C. E.; Li, L., A study of reproducibility of guanidination-dimethylation labeling and liquid chromatography matrix-assisted laser desorption ionization mass spectrometry for relative proteome quantification. *Anal Chim Acta* **2007,** *585* (2), 219-26.

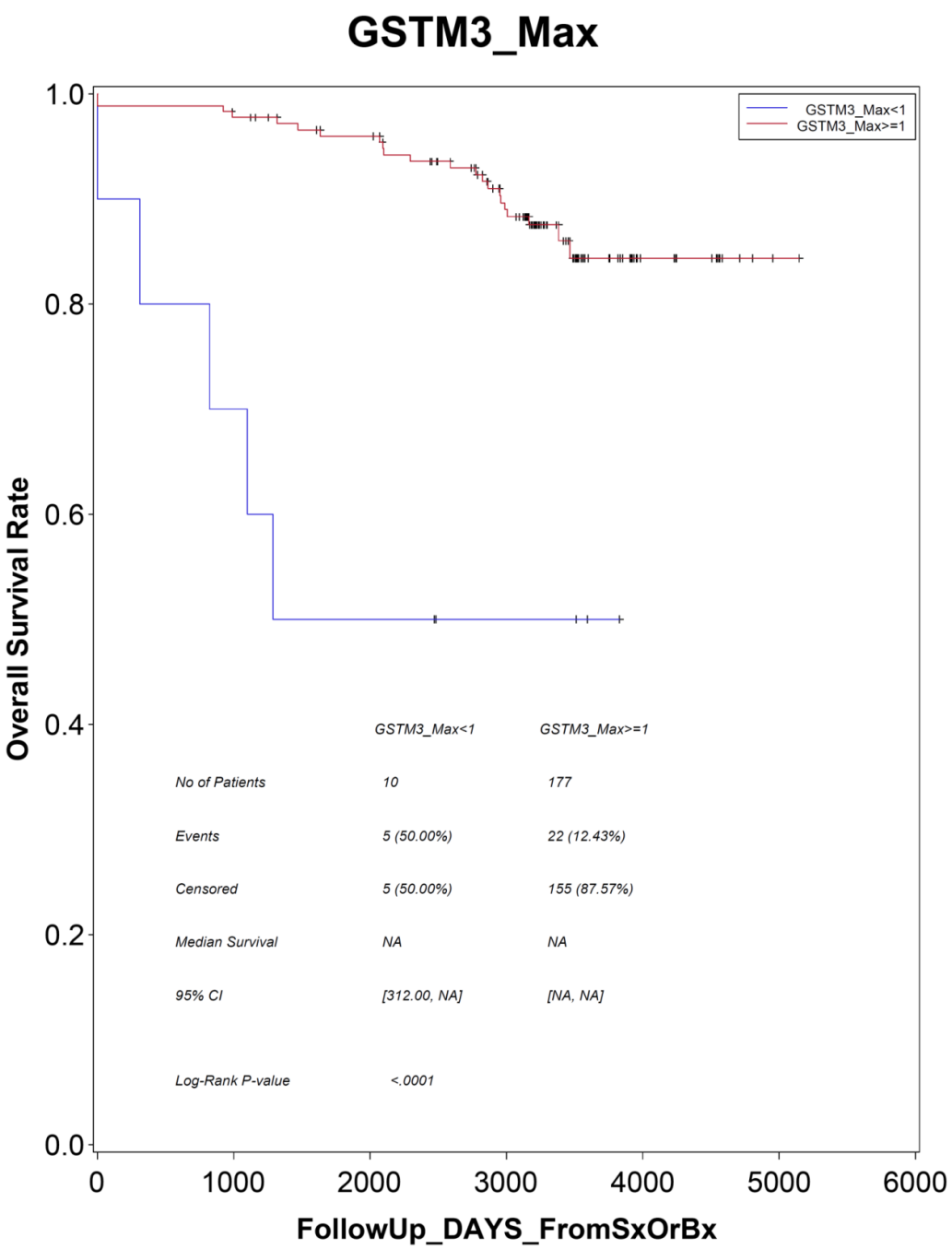
Appendices

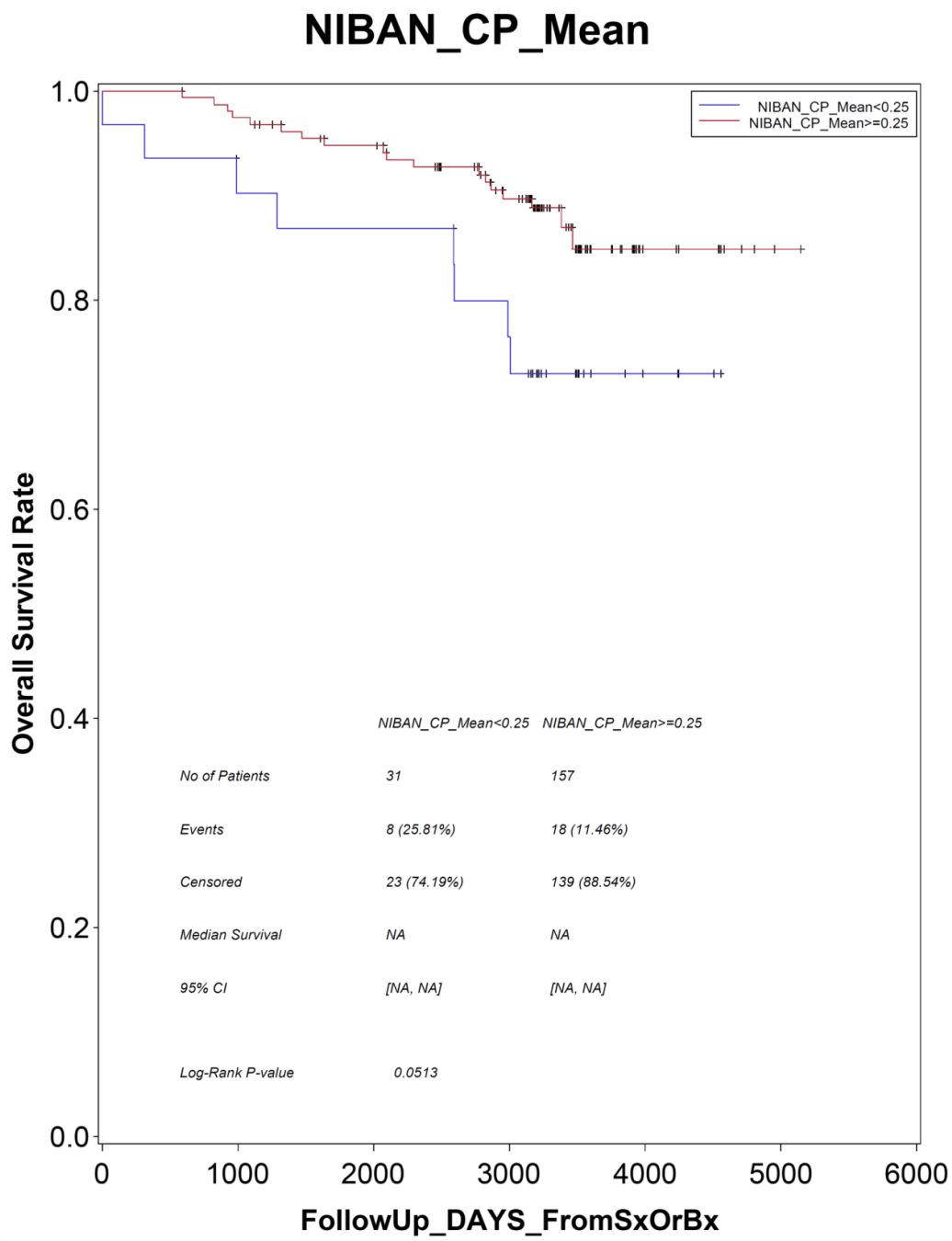
Supplemental Figure S4.1. 5 proteins' overall survival curves in breast cancer patients calculated from the day of surgery or biopsy.

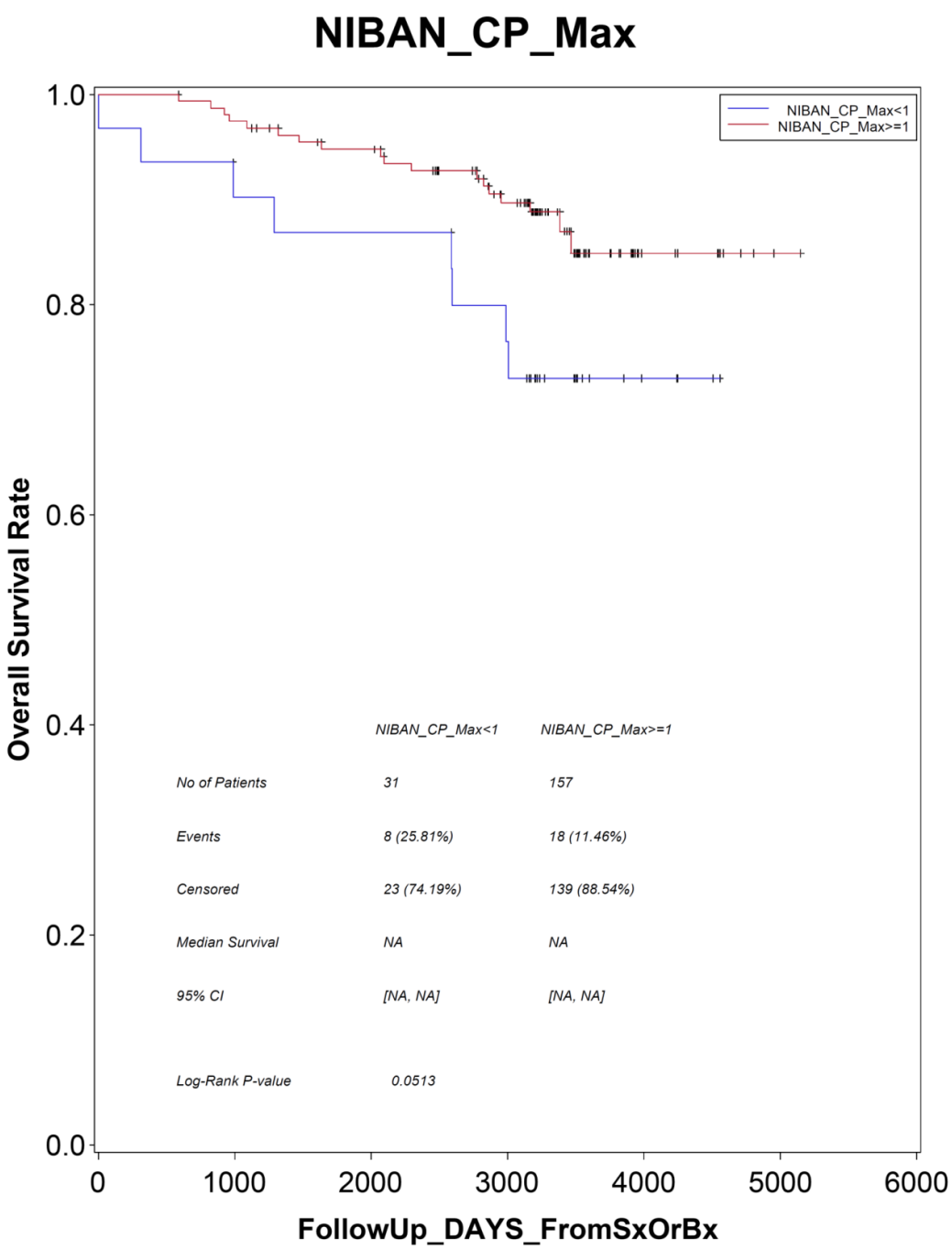


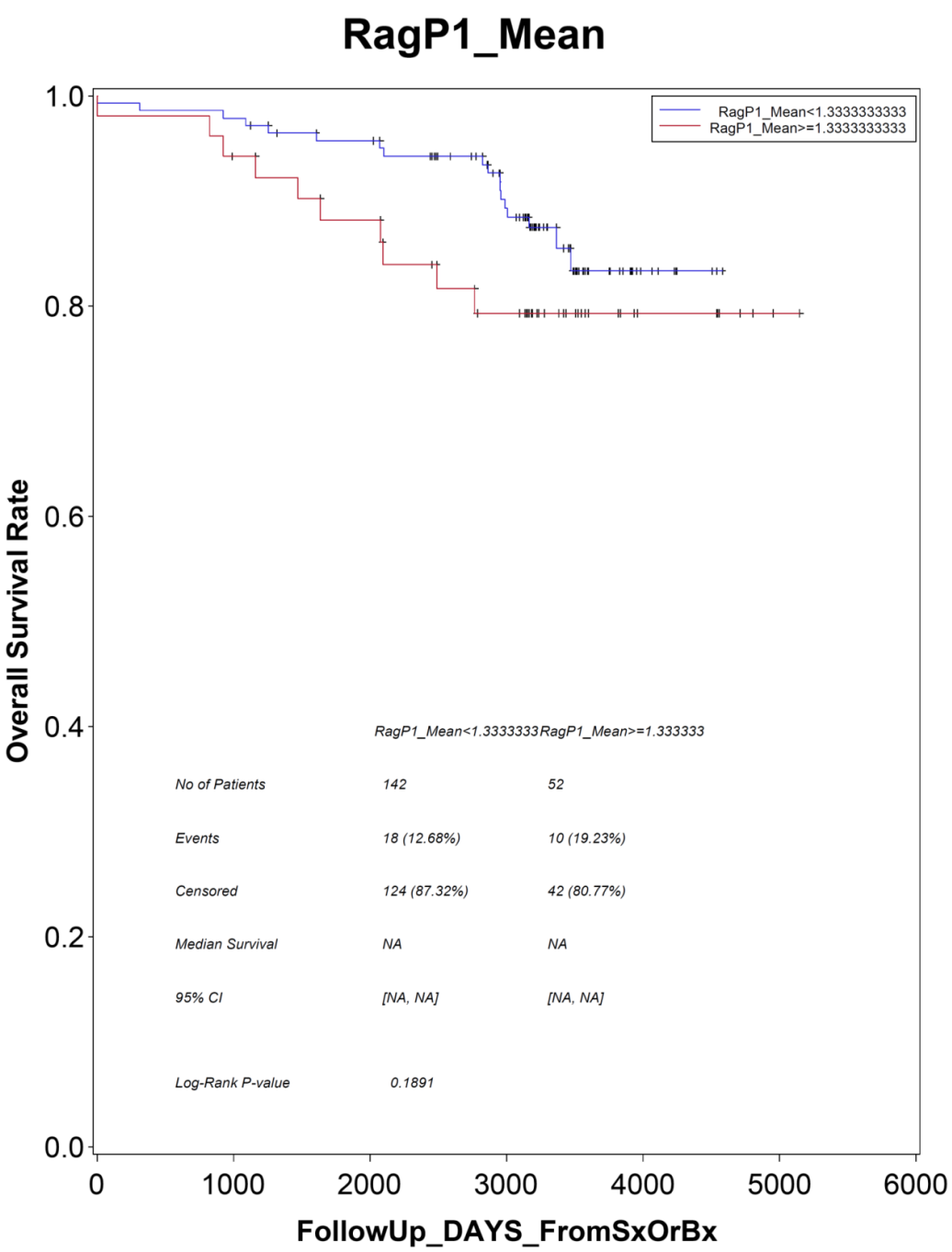

AIFM1_Mean
AIFM1_Mean<2.75
AIFM1_Mean>=2.75
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
AIFM1_Mean<2.75
AIFM1_Mean>=2.75
No of Patients
151
39
Events
27 (17.88%)
2 (5.13%)
Censored
124 (82.12%)
37 (94.87%)
Median Survival
NA
NA
95% CI
[NA, NA]
[NA, NA]
Log-Rank P-value
0.0588

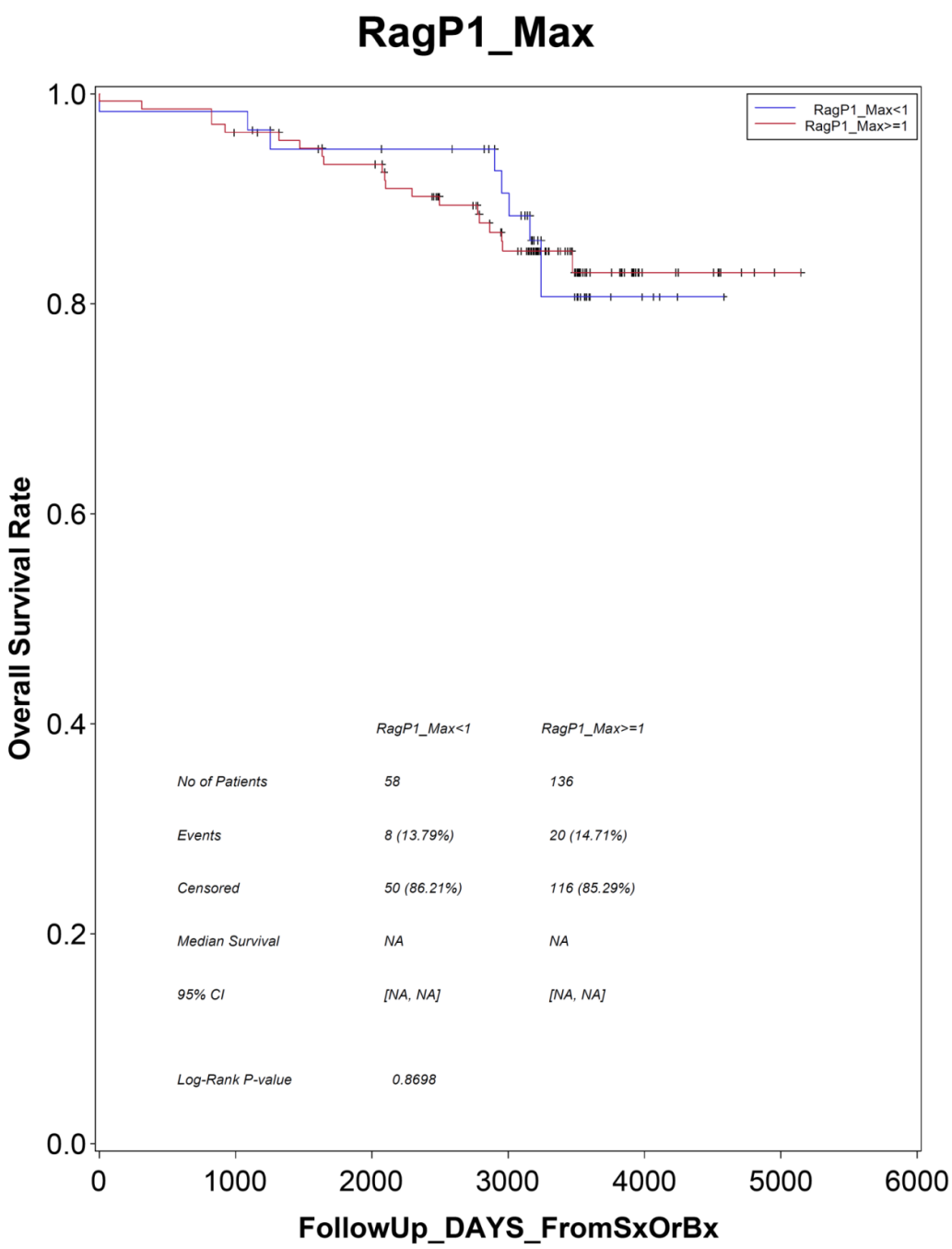

AIFM1_Max
AIFM1_Max<1
AIFM1_Max>=1
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
AIFM1_Max<1
AIFM1_Max>=1
No of Patients 29 161
Events 2 (6.90%) 27 (16.77%)
Censored 27 (93.10%) 134 (83.23%)
Median Survival NA NA
95% CI [NA, NA] [NA, NA]
Log-Rank P-value 0.1714

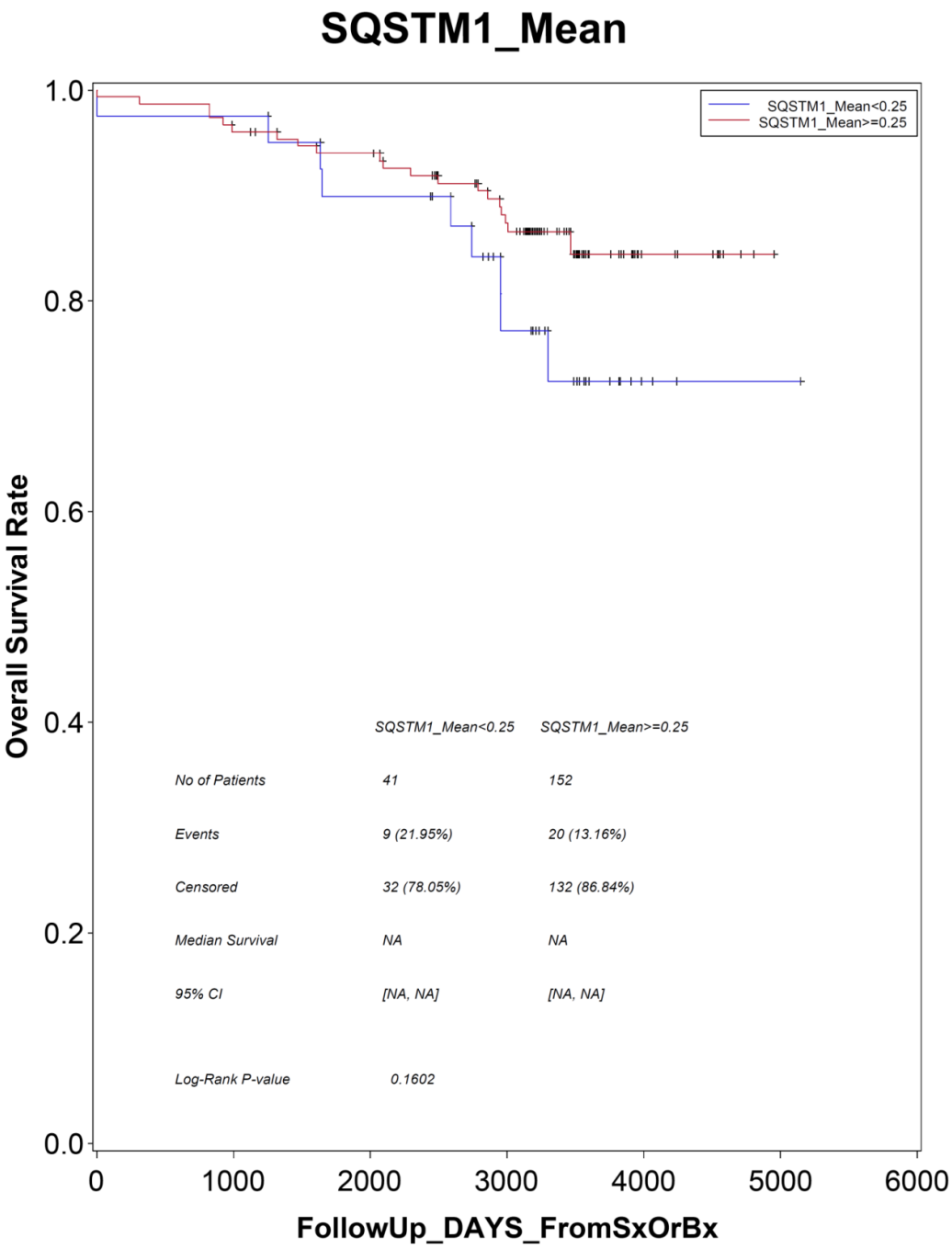

GSTM3_Max
GSTM3_Max<1
GSTM3_Max>=1
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
GSTM3_Max<1
GSTM3_Max>=1
No of Patients
10
177
Events
5 (50.00%)
22 (12.43%)
Censored
5 (50.00%)
155 (87.57%)
Median Survival
NA
NA
95% CI
[312.00, NA]
[NA, NA]
Log-Rank P-value
<.0001

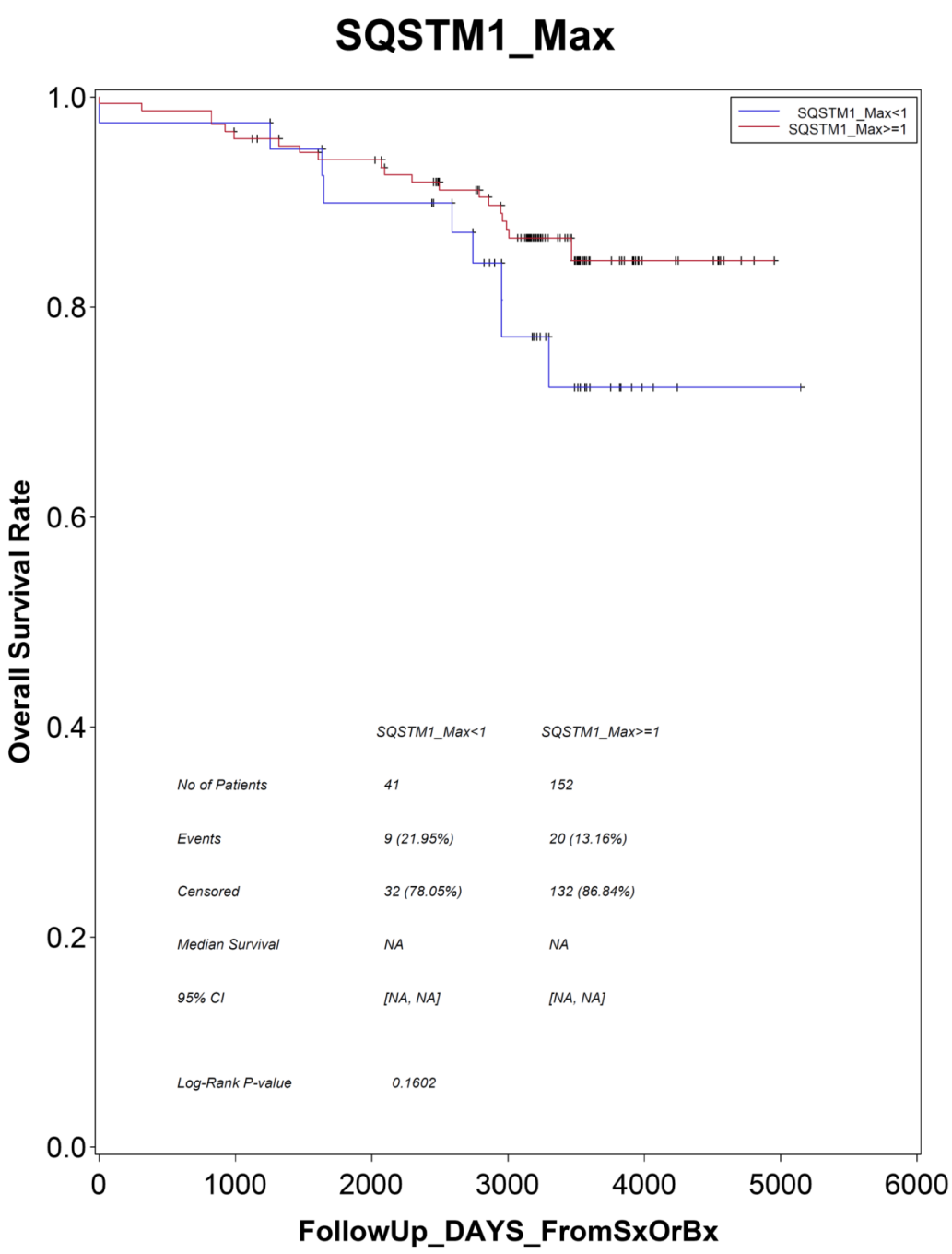

NIBAN_CP_Mean
NIBAN_CP_Mean<0.25
NIBAN_CP_Mean>=0.25
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
NIBAN_CP_Mean<0.25 NIBAN_CP_Mean>=0.25
No of Patients 31 157
Events 8 (25.81%) 18 (11.46%)
Censored 23 (74.19%) 139 (88.54%)
Median Survival NA NA
95% CI [NA, NA] [NA, NA]
Log-Rank P-value 0.0513

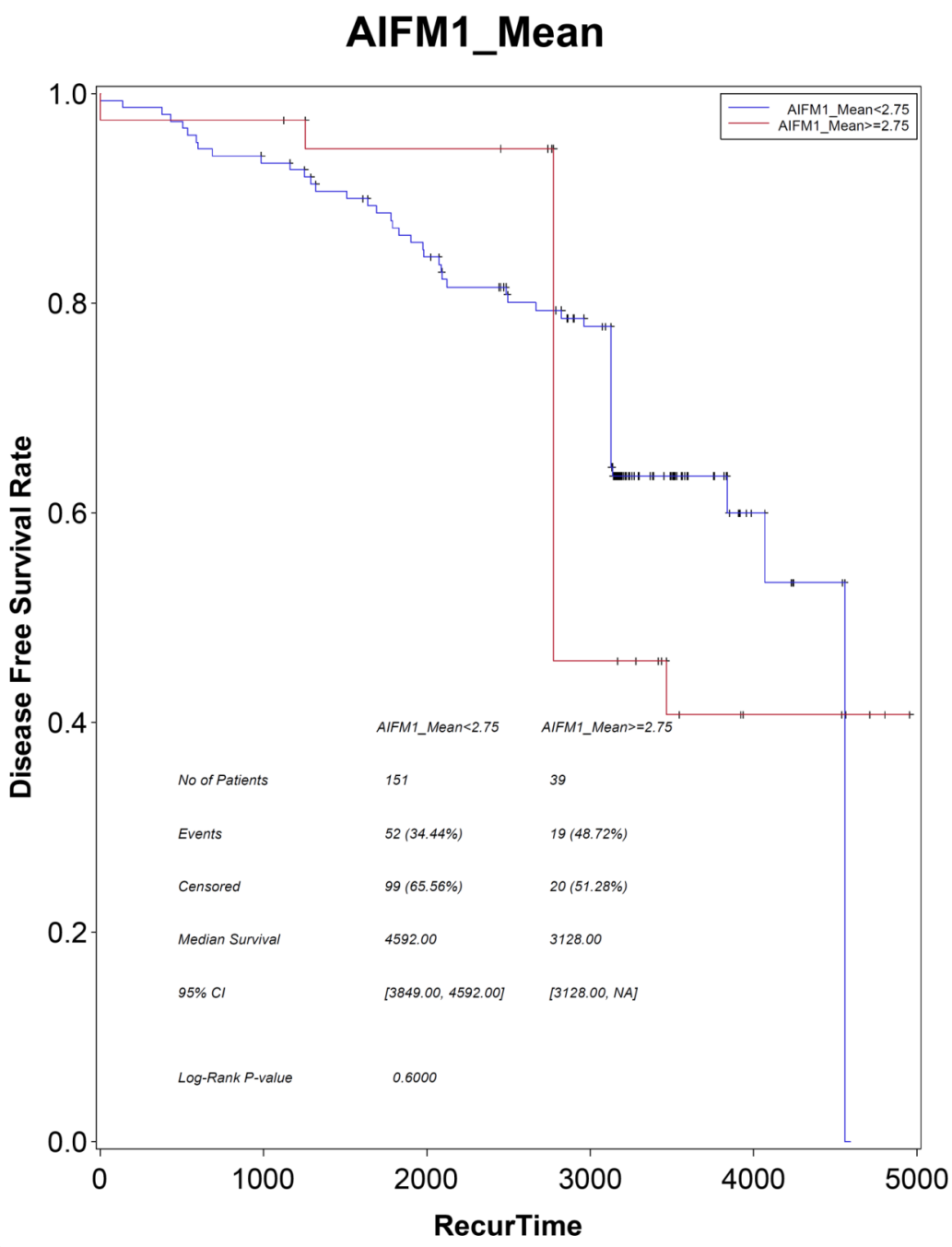

NIBAN_CP_Max
NIBAN_CP_Max<1
NIBAN_CP_Max>=1
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
NIBAN_CP_Max<1
NIBAN_CP_Max>=1
No of Patients 31 157
Events 8 (25.81%) 18 (11.46%)
Censored 23 (74.19%) 139 (88.54%)
Median Survival NA NA
95% CI [NA, NA] [NA, NA]
Log-Rank P-value 0.0513

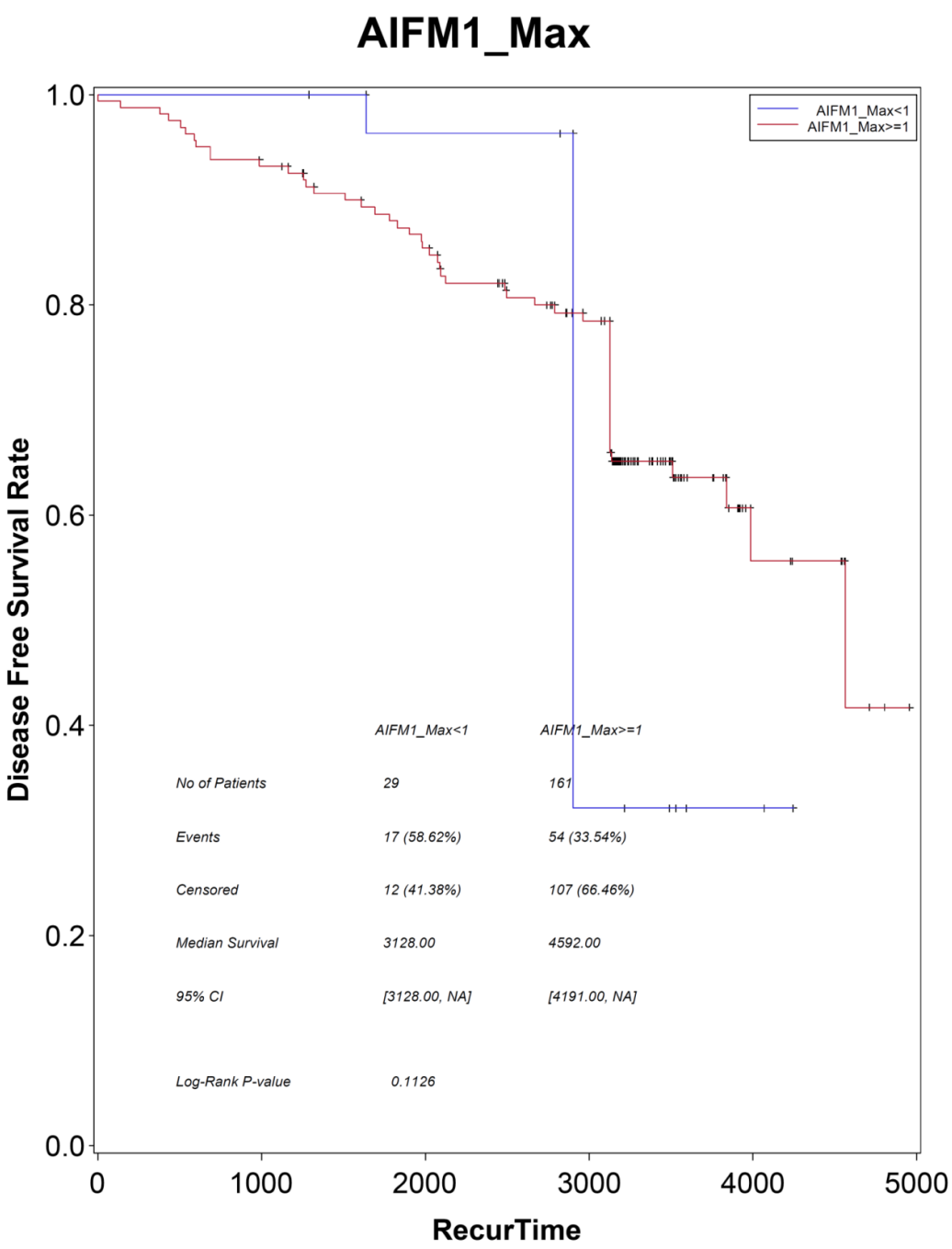

RagP1_Mean
RagP1_Mean<1.3333333333
RagP1_Mean>=1.3333333333
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
RagP1_Mean<1.3333333
RagP1_Mean>=1.333333
No of Patients 142 52
Events 18 (12.68%) 10 (19.23%)
Censored 124 (87.32%) 42 (80.77%)
Median Survival NA NA
95% CI [NA, NA] [NA, NA]
Log-Rank P-value 0.1891

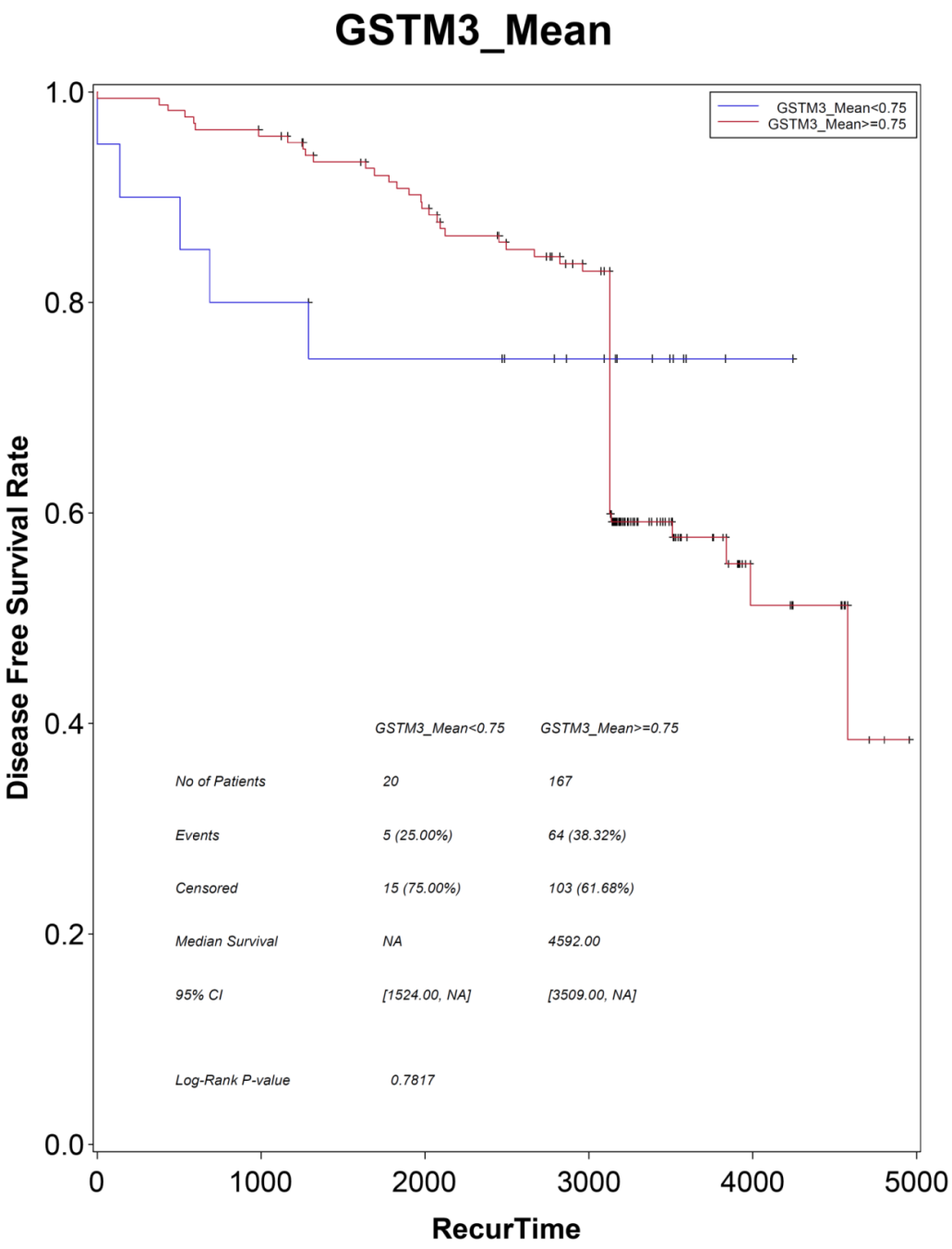

RagP1_Max
RagP1_Max<1
RagP1_Max>=1
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
RagP1_Max<1
RagP1_Max>=1
No of Patients 58 136
Events 8 (13.79%) 20 (14.71%)
Censored 50 (86.21%) 116 (85.29%)
Median Survival NA NA
95% CI [NA, NA] [NA, NA]
Log-Rank P-value 0.8698

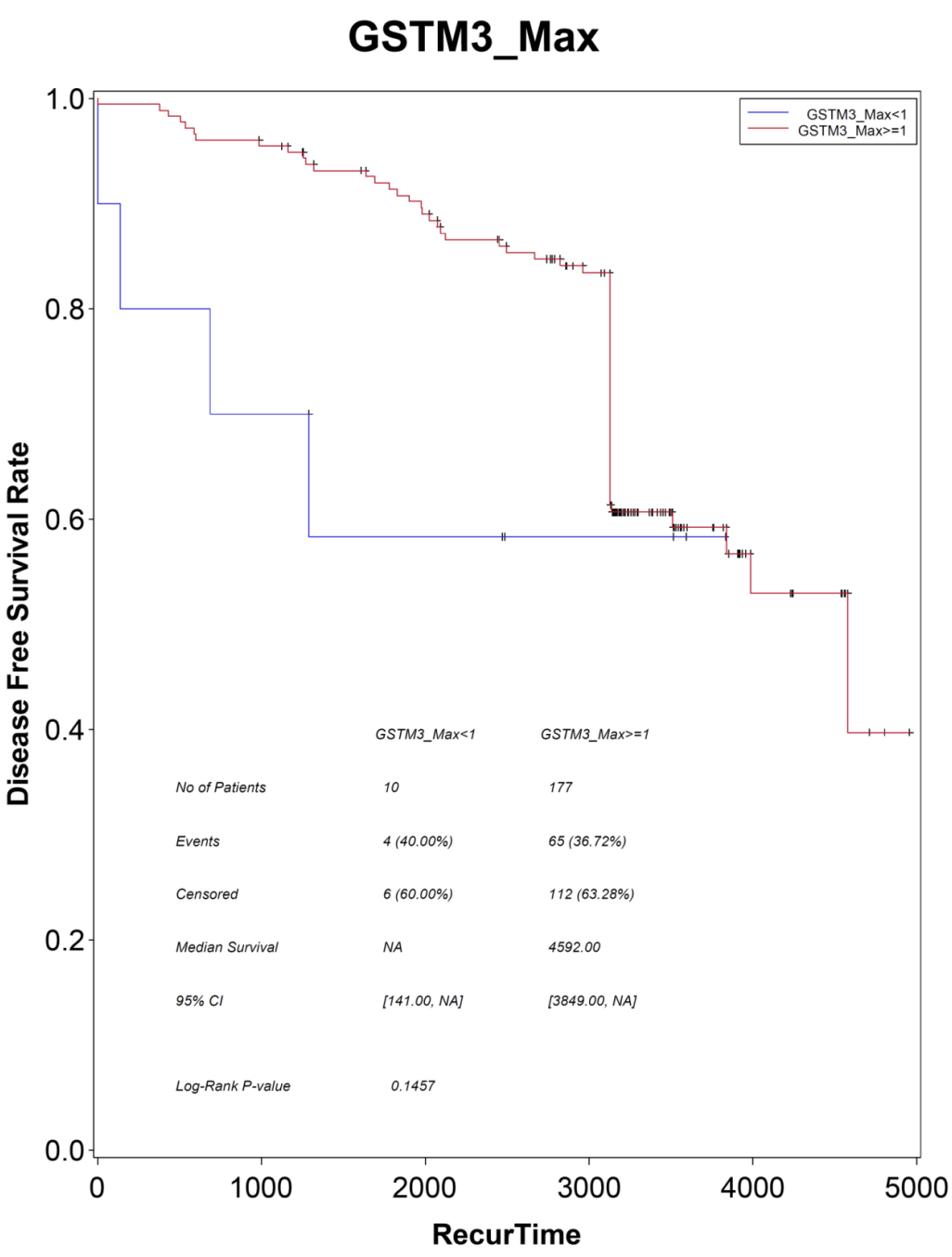

SQSTM1_Mean
SQSTM1_Mean<0.25
SQSTM1_Mean>=0.25
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
SQSTM1_Mean<0.25 SQSTM1_Mean>=0.25
No of Patients 41 152
Events 9 (21.95%) 20 (13.16%)
Censored 32 (78.05%) 132 (86.84%)
Median Survival NA NA
95% CI [NA, NA] [NA, NA]
Log-Rank P-value 0.1602

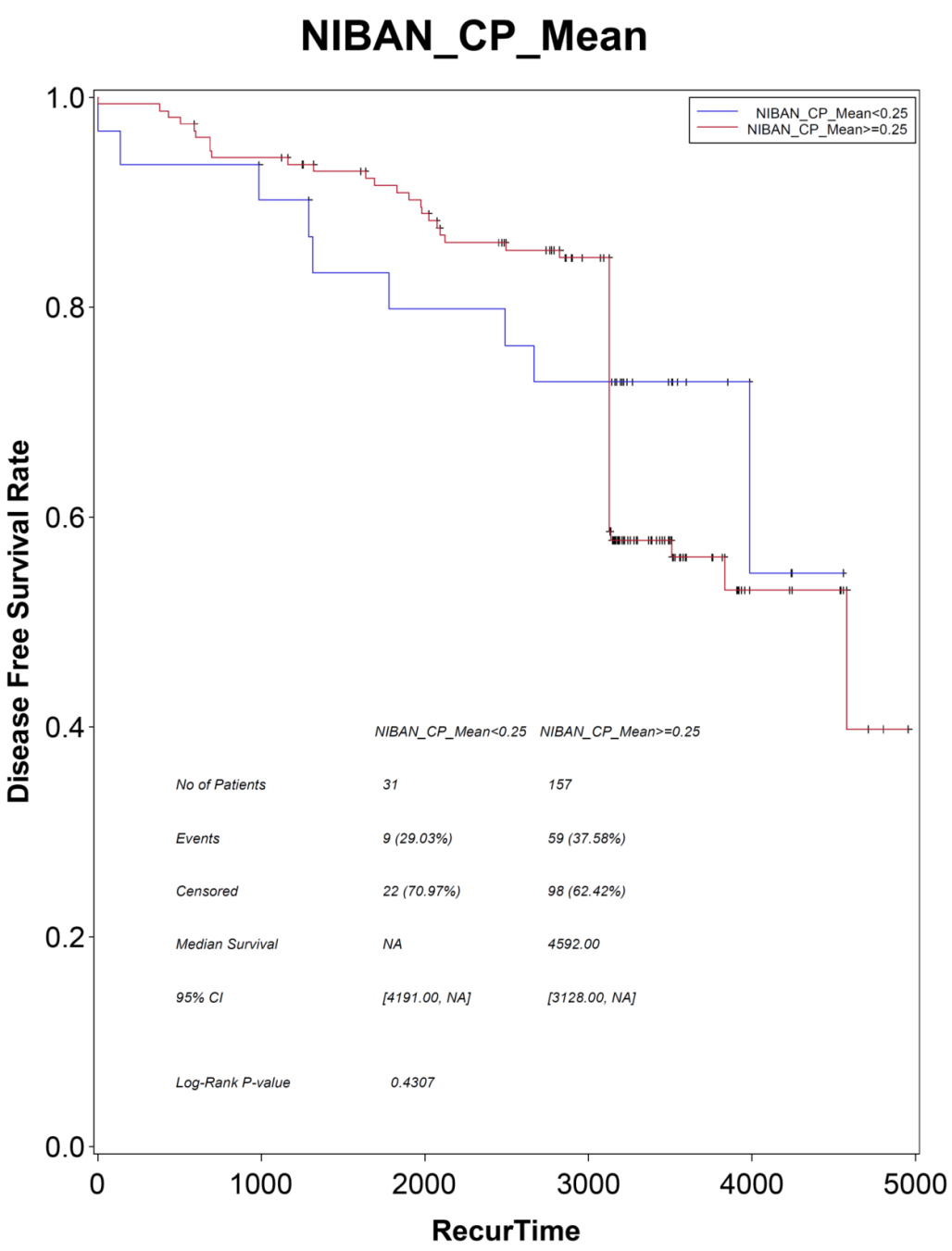

SQSTM1_Max
SQSTM1_Max<1
SQSTM1_Max>=1
Overall Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
6000
FollowUp_DAYS_FromSxOrBx
SQSTM1_Max<1
SQSTM1_Max>=1
No of Patients
41
152
Events
9 (21.95%)
20 (13.16%)
Censored
32 (78.05%)
132 (86.84%)
Median Survival
NA
NA
95% CI
[NA, NA]
[NA, NA]
Log-Rank P-value
0.1602

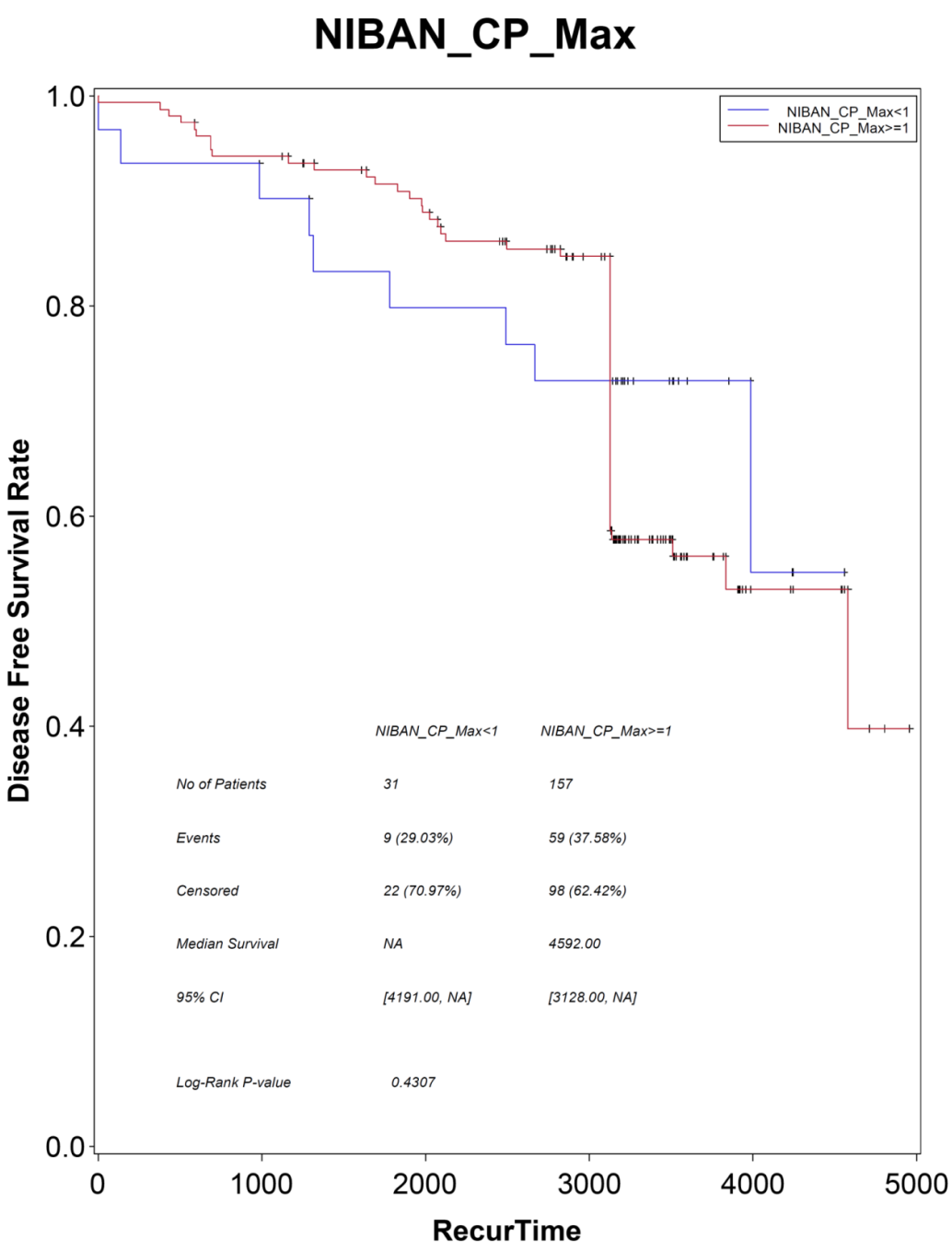

AIFM1_Mean
AIFM1_Mean<2.75
AIFM1_Mean>=2.75
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
AIFM1_Mean<2.75
AIFM1_Mean>=2.75
No of Patients
151
39
Events
52 (34.44%)
19 (48.72%)
Censored
99 (65.56%)
20 (51.28%)
Median Survival
4592.00
3128.00
95% CI
[3849.00, 4592.00]
[3128.00, NA]
Log-Rank P-value
0.6000

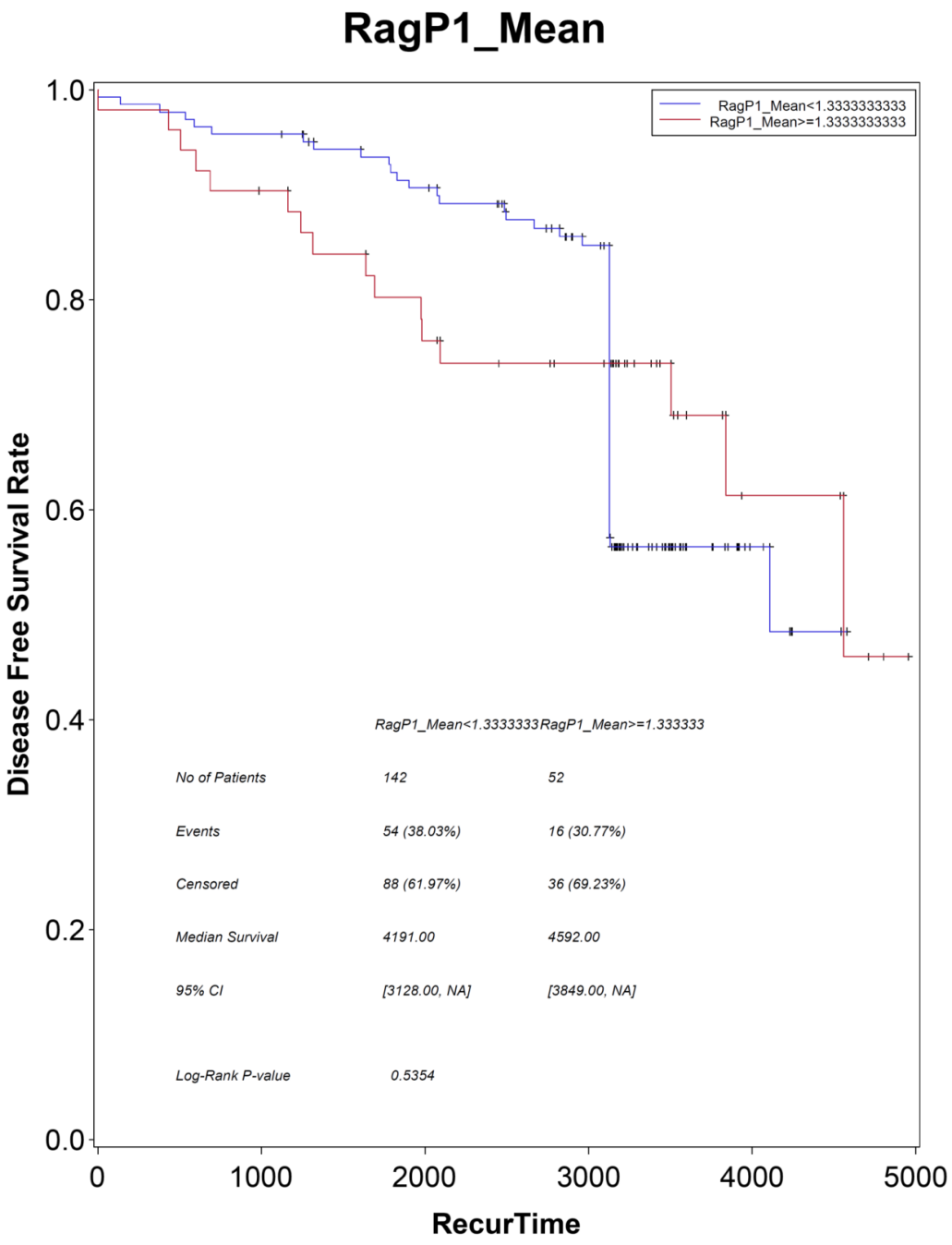

AIFM1_Max
AIFM1_Max<1
AIFM1_Max>=1
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
AIFM1_Max<1
AIFM1_Max>=1
No of Patients 29 161
Events 17 (58.62%) 54 (33.54%)
Censored 12 (41.38%) 107 (66.46%)
Median Survival 3128.00 4592.00
95% CI [3128.00, NA] [4191.00, NA]
Log-Rank P-value 0.1126

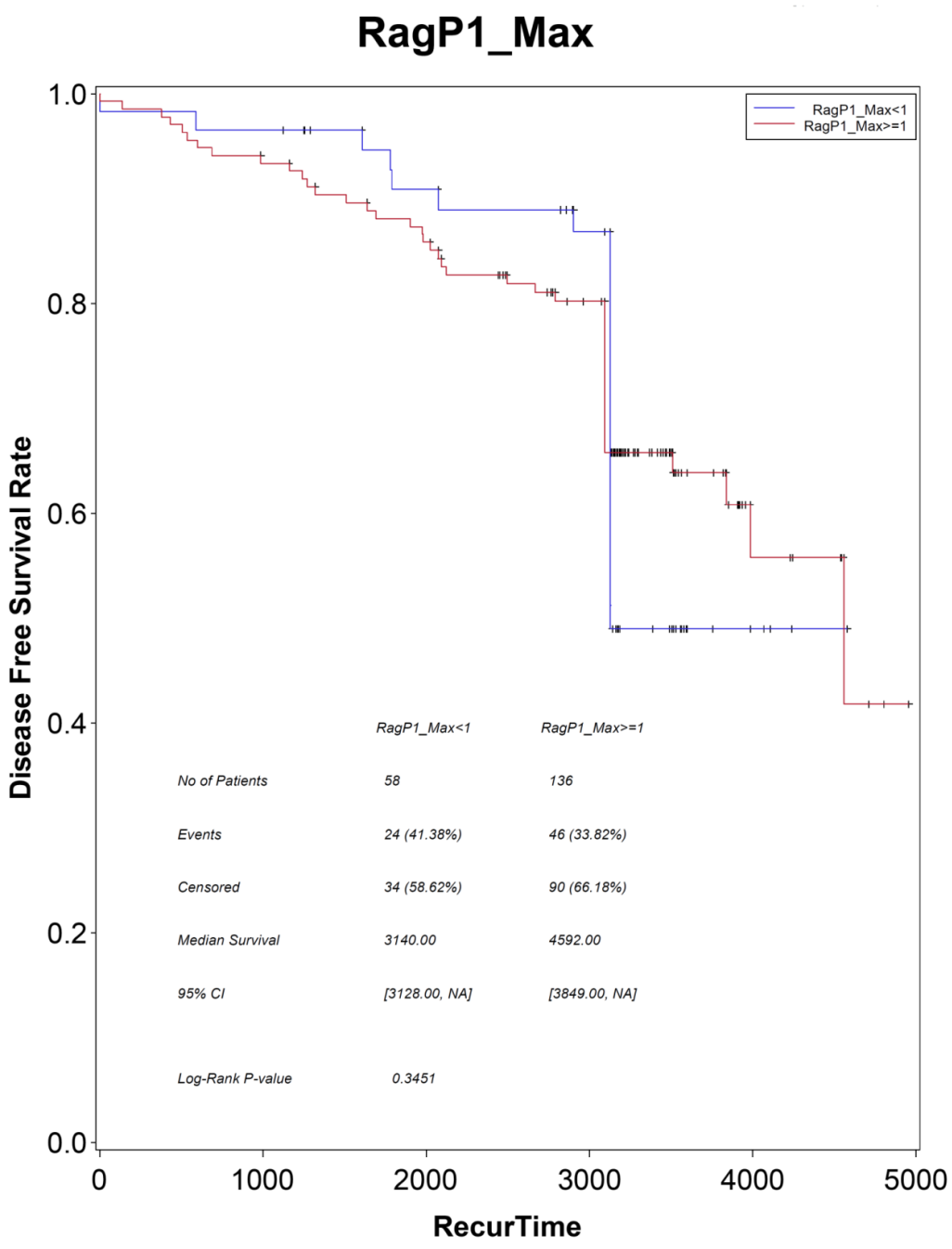

GSTM3_Mean
GSTM3_Mean<0.75
GSTM3_Mean>=0.75
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
GSTM3_Mean<0.75 GSTM3_Mean>=0.75
No of Patients 20 167
Events 5 (25.00%) 64 (38.32%)
Censored 15 (75.00%) 103 (61.68%)
Median Survival NA 4592.00
95% CI [1524.00, NA] [3509.00, NA]
Log-Rank P-value 0.7817

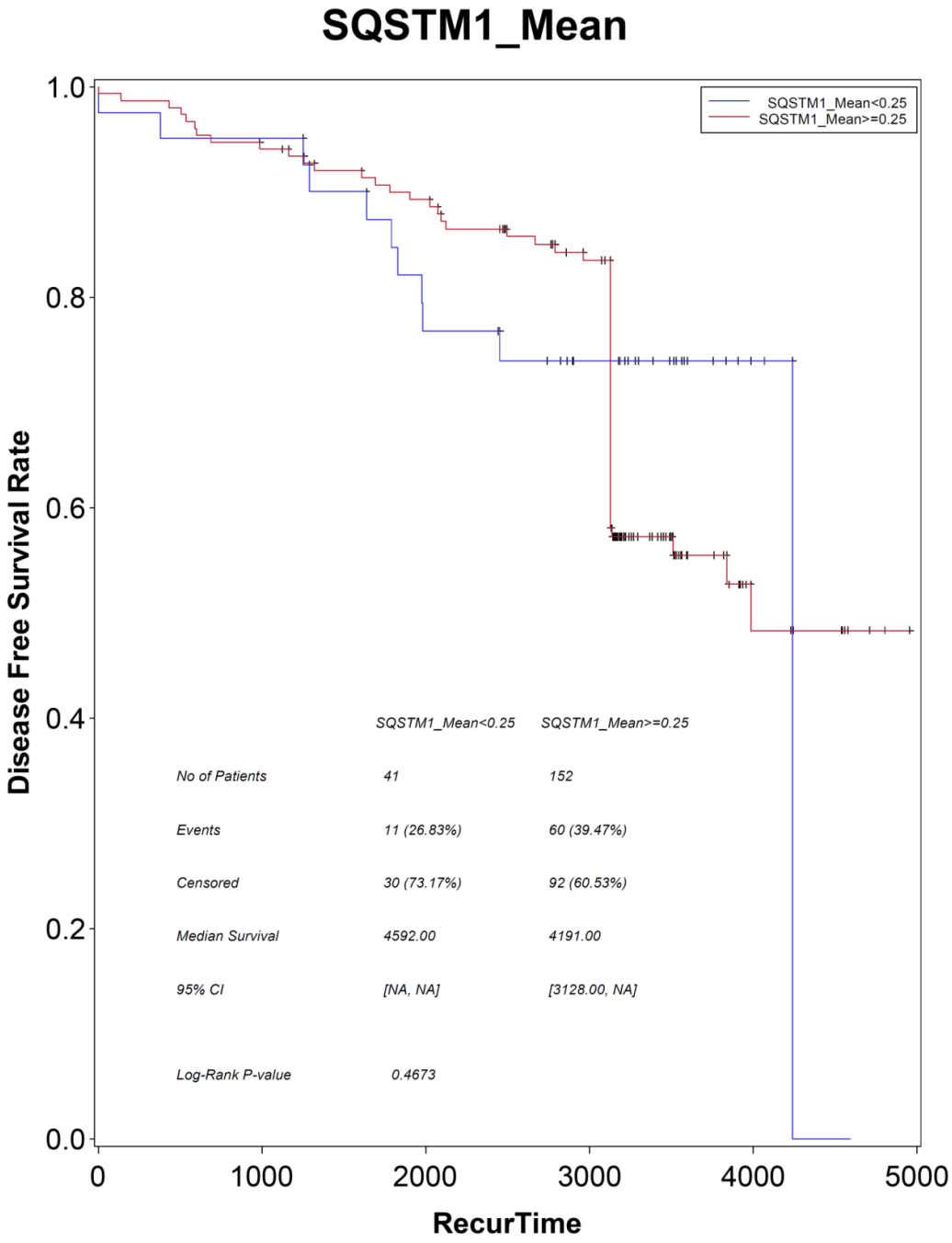

GSTM3_Max
GSTM3_Max<1
GSTM3_Max>=1
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
GSTM3_Max<1
GSTM3_Max>=1
No of Patients
10
177
Events
4 (40.00%)
65 (36.72%)
Censored
6 (60.00%)
112 (63.28%)
Median Survival
NA
4592.00
95% CI
[141.00, NA]
[3849.00, NA]
Log-Rank P-value
0.1457

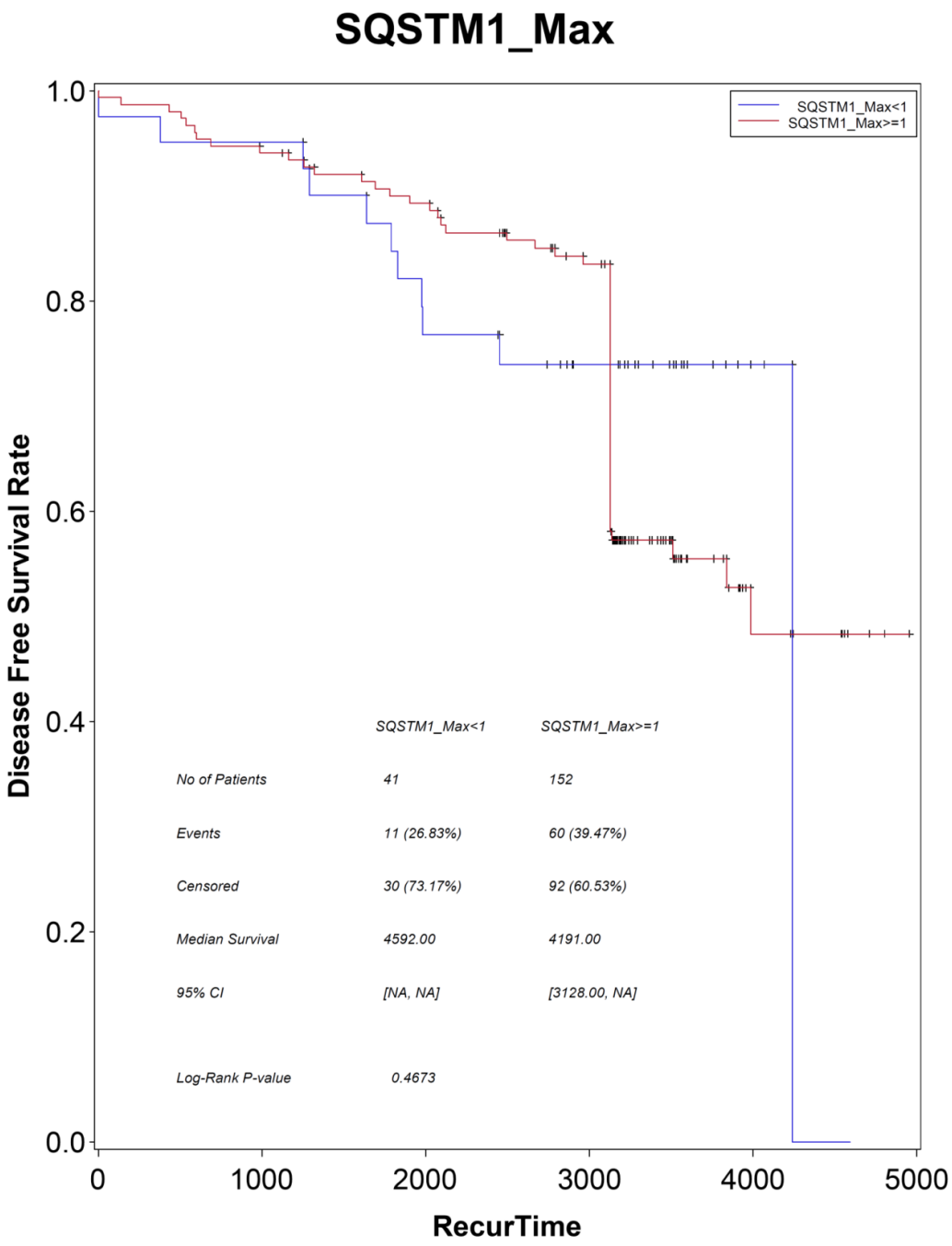

NIBAN_CP_Mean
NIBAN_CP_Mean<0.25
NIBAN_CP_Mean>=0.25
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
NIBAN_CP_Mean<0.25 NIBAN_CP_Mean>=0.25
No of Patients 31 157
Events 9 (29.03%) 59 (37.58%)
Censored 22 (70.97%) 98 (62.42%)
Median Survival NA 4592.00
95% CI [4191.00, NA] [3128.00, NA]
Log-Rank P-value 0.4307


NIBAN_CP_Max
NIBAN_CP_Max<1
NIBAN_CP_Max>=1
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
NIBAN_CP_Max<1
NIBAN_CP_Max>=1
No of Patients 31 157
Events 9 (29.03%) 59 (37.58%)
Censored 22 (70.97%) 98 (62.42%)
Median Survival NA 4592.00
95% CI [4191.00, NA] [3128.00, NA]
Log-Rank P-value 0.4307

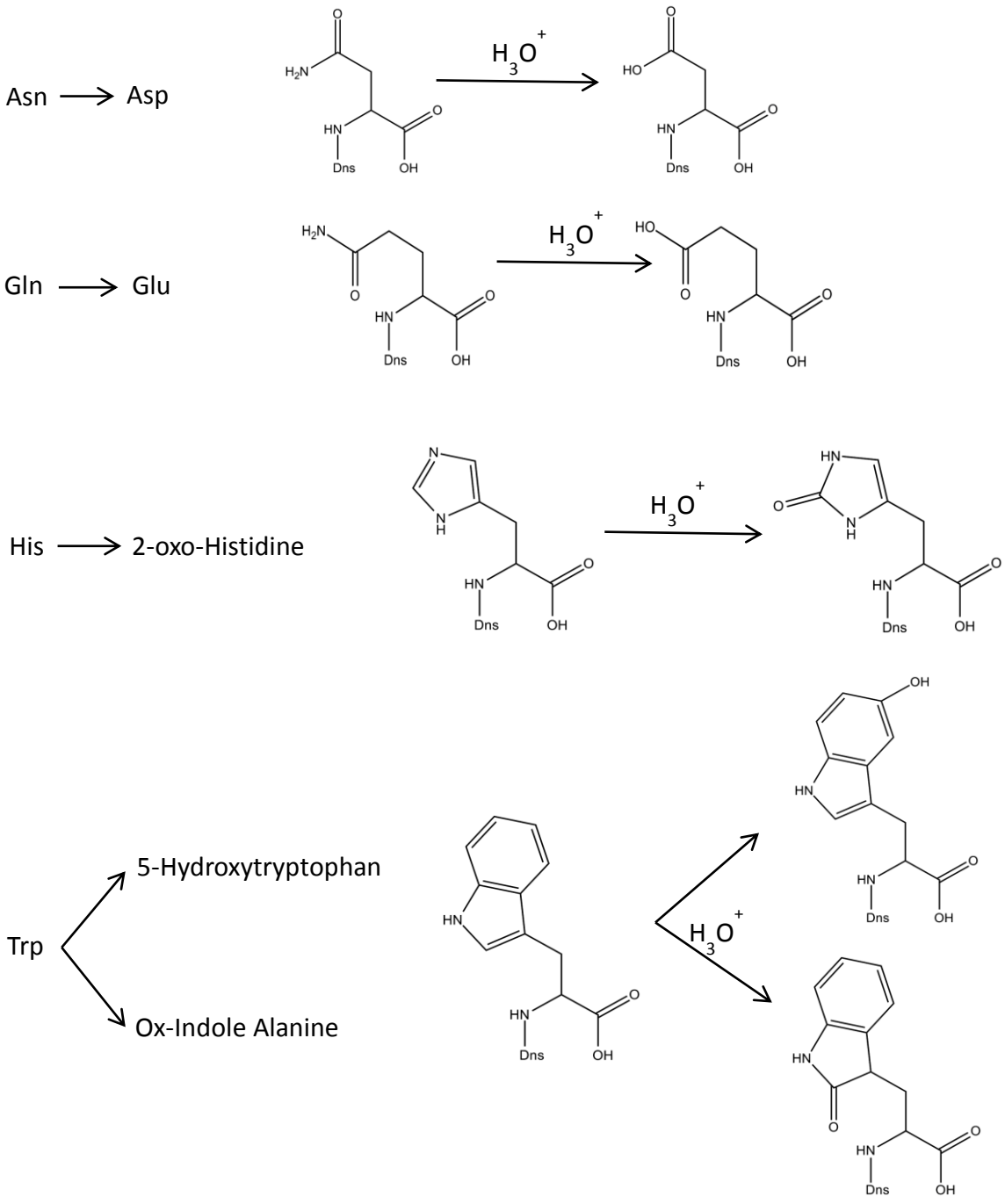

RagP1_Mean
RagP1_Mean<1.3333333333
RagP1_Mean>=1.3333333333
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
RagP1_Mean<1.3333333
RagP1_Mean>=1.333333
No of Patients 142 52
Events 54 (38.03%) 16 (30.77%)
Censored 88 (61.97%) 36 (69.23%)
Median Survival 4191.00 4592.00
95% CI [3128.00, NA] [3849.00, NA]
Log-Rank P-value 0.5354


RagP1_Max
RagP1_Max<1
RagP1_Max>=1
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
RagP1_Max<1
RagP1_Max>=1
No of Patients
58
136
Events
24 (41.38%)
46 (33.82%)
Censored
34 (58.62%)
90 (66.18%)
Median Survival
3140.00
4592.00
95% CI
[3128.00, NA]
[3849.00, NA]
Log-Rank P-value
0.3451


SQSTM1_Mean
SQSTM1_Mean<0.25
SQSTM1_Mean>=0.25
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
SQSTM1_Mean<0.25
SQSTM1_Mean>=0.25
No of Patients 41 152
Events 11 (26.83%) 60 (39.47%)
Censored 30 (73.17%) 92 (60.53%)
Median Survival 4592.00 4191.00
95% CI [NA, NA] [3128.00, NA]
Log-Rank P-value 0.4673


SQSTM1_Max
SQSTM1_Max<1
SQSTM1_Max>=1
Disease Free Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
RecurTime
SQSTM1_Max<1
SQSTM1_Max>=1
No of Patients 41 152
Events 11 (26.83%) 60 (39.47%)
Censored 30 (73.17%) 92 (60.53%)
Median Survival 4592.00 4191.00
95% CI [NA, NA] [3128.00, NA]
Log-Rank P-value 0.4673

Supplemental Figure S6.1. Chemical transformation of amino acids during acid hydrolysis.



Supplemental Note N6.1

The amino acid sequence of intact mCherry:

MRGSH HHHHH GMASM TGGQQ MGRDL YDDDD KDPMV SKGEE DNMAI IKEFM RFKVH MEGSV NGHEF EIEGE GEGRP YEGTQ TAKLK VTKGG PLPFA WDILS PQFMY GSKAY VKHPA DIPDY LKLSF PEGFK WERVM NFEDG GVVTV TQDSS LQDGE FIYKV KLRGT NFPSD GPVMQ KKTMG WEASS ERMYP EDGAL KGEIK QRLKL KDGGH YDAEV KTTYK AKKPV QLPGA YNVNI KLDIT SHNED YTIVE QYERA EGRHS TGGMD ELYK

The amino acid sequence of N-terminal truncated mCherry:

DPMVS KGEED NMAII KEFMR FKVHM EGSVN GHEFE IEGEG EGRPY EGTQT AKLKV TKGGP LPFAW DILSP QFMYG SKAYV KHPAD IPDYL KLSFP EGFKW ERVMN FEDGG VVTVT QDSSL QDGEF IYKVK LRGTN FPSDG PVMQK KTMGW EASSE RMYPE DGALK GEIKQ RLKLK DGGHY DAEVK TTYKA KKPVQ LPGAY NVNIK LDITS HNEDY TIVEQ YERAE GRHST GGMDE LYK