University of Alberta

Development of Mass Spectrometric Methods
for Membrane Proteome Analysis and Protein Sequence Mapping

by

Difei Sun

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

Spring 2014
Edmonton, Alberta

To my parents and my husband

Abstract

This thesis work mainly focused on the development of mass spectrometric methods for membrane proteome analysis and protein sequence mapping. First, the comparison of the performance of RapiGest-, PPS- and SDS-based sample preparation methods for shotgun membrane proteome analysis was investigated. We demonstrated that the use of RapiGest allows identification of more peptides and proteins than the use of PPS or SDS. Second, based on the RapiGest-assisted membrane protein analysis method, a high throughput plasma membrane protein identification and quantitation strategy was developed for the analysis of ALK+ ALCL cell lines. Using this method, 561 and 552 unique plasma membrane proteins and extracellular proteins were identified from Karpas 299 cell line and SUPM2 cell line, respectively; 48 unique plasma membrane proteins and extracellular proteins were found to be differentially expressed between NPM-ALK expressing HEK 293 cells and NPM-ALK absent HEK 293 cells by quantification analysis; and six putative biomarkers were chosen for further biological validation. Third, an integrated strong-cation exchange liquid chromatographic procedure for SDS removal and peptide separation for SDS-assisted shotgun proteome analysis was developed, which allowed effective SDS removal while keeping a high peptide recovery rate when the low-cost

SDS-assisted sample preparation method was used. We have shown that the performance of this method is better than that of the RapiGest-assisted method. Fourth, a protein sequence mapping method was developed based on the use of gel electrophoresis to separate proteins, followed by in-gel MAAH of the gel-separated proteins using TFA to produce peptides and LC-MS/MS of the resultant peptides. This method provided high sequence coverage of proteins separated from complex protein mixture. To demonstrate the applications of this method, 19 relatively high abundance human plasma proteins were mapped with high sequence coverage. In addition, bovine alpha-S1-casein phospho-isoforms were characterized and six new phosphorylation sites were identified. The methods developed in this thesis work have been shown to be useful in proteomics research. The membrane proteome analysis methods are applicable to study any cell lines or tissues and the protein sequence mapping method can be applied to characterize a variety of proteins with virtually no molecular weight limit.

Acknowledgements

I would like to express my sincere thanks to my supervisor, Dr. Liang Li, for his invaluable guidance, encouragement and advice throughout my Ph.D. studies. I have learned a lot from him on the technical, professional, and personal aspects. His enthusiasm to research will leave a long lasting impression to me.

My sincere gratitude also goes to my committee members, Dr. Richard L. McCreey, Dr. Christopher W. Cairo, Dr. Charles A. Lucy, Dr. Larry Fliegel, and Dr. Daniel Figeys for their time and contribution to the examination process and their valuable advice on my research.

The work in this thesis was carried out with the support of many other individuals, including Dr. Raymond Lai, Dr. Nan Li and other researchers at Sanofi Pasteur, Toronto, Canada. I thank them for their contributions and helpful discussion. I also would like to thank all the staff in the Chemistry Department who have given support to my graduate study.

I would like to express special thanks to Dr. Fang Wu for her training at the beginning of my Ph.D. studies and to Dr. Nan Wang for her help and advice to my projects. And many thanks are given to all group members in Dr. Liang Li's research group.

Finally, I would like to express my deepest thanks to my parents, Mr. Jiyu Sun and Mrs. Yanying Liu, and my husband, Mr. Yi Gu, for their long term support, encouragement, and understanding.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2-DE	two-dimensional gel electrophoresis
2MEGA	dimethylation after guanidination
ACN	acetonitrile
ALK+ ALCL	anaplastic lymphoma kinase-positive anaplastic large cell lymphoma
BSA	bovine serum albumin
CHCA	α-cyano-4-hydroxycinnamic acid
CID	collision-induced dissociation
CMC	critical micelle concentration
CNBr	cyanogen bromide
DDA	data directed analysis
DHB	2, 5-dihydroxy benzoic acid
DOX	doxycycline
DTT	dithiothreitol
ECD	electron-capture dissociation
ESI	electrospray ionization
ETD	electron-transfer dissociation

FA	formic acid
FTICR	fourier-transform ion cyclotron resonance
GC	gas chromatography
GRAVY	grand average of hydropathy
HCD	higher-energy collisional dissociation
HILIC	hydrophilic interaction liquid chromatography
IAA	iodoacetamide
ICAT	isotope-coded affinity tag
IDA	iminodiacetic acid
IMAC	immobilized metal affinity chromatography
iTRAQ	isobaric tag for relative and absolute quantification
LB	Lysogeny Broth
m/z	mass to charge ratio
MAAH	microwave-assisted acid hydrolysis
MALDI	matrix-assisted laser desorption ionization
MDLC	multidimensional liquid chromatography
MOAC	metal oxide affinity chromatography
MS/MS	tandem mass spectrometry
NPM	nucleophosmin
NTA	nitrilotriacetic acid

P	probability
PBS	phosphate-buffered saline
pI	isoelectric point
PIE	precursor ion exclusion
QTOF	quadrupole time-of-flight
RP	reversed-phase
RPLC	reverse phase liquid chromatography
SA	sinapinic acid
SAX	strong anion exchange
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SILAC	stable isotope labeling by amino acids in cell culture
TFA	trifluoroacetic acid
WAX	weak anion exchange
WCX	weak cation exchange

Chapter 1 Introduction

Proteomics is the large-scale study of an entire set of proteins, produced or modified by an organism or system. It is a relatively young research field and plays an important role in many areas such as biology and pharmacy. Depending on the type of biological questions to be addressed, the goal of proteomics can be the identification of proteins in a biological sample or the quantitative analysis of protein varying at different times or under different conditions.

In proteomics analysis, ideally all the proteins or the entire proteome in a sample would be profiled and quantified. However, this is a very challenging task, especially for membrane proteins which are difficult to handle due to their high hydrophobicity. Also, the protein sequence mapping from complex mixture is another remaining challenge. Therefore, the overall goal of this thesis research is to develop new mass spectrometric techniques for comprehensive membrane proteome analysis and protein sequence mapping. To achieve this goal, mass spectrometry and other basic techniques are used in my thesis work. In this chapter, the most relevant topics will be introduced, including techniques for mass spectrometric protein identification, quantification, and analysis of protein phosphorylation. The scope of this thesis will be discussed at the end.

1.1 Mass spectrometric protein identification

1.1.1 Overview of mass spectrometric protein identification methods

In recent decades, two different MS-based approaches, top-down and bottom-up (shotgun), have been developed and widely used for protein identification.[1] Top-down method is based on the separation of proteins and subsequent MS/MS analysis of individual proteins.[2-4] The protein identification and amino acid sequence are determined by the m/z values of the fragment ions. One advantage of top-down strategy is the high sequence coverage of the identified protein, which could reduce the ambiguities of the protein identification.[5] Another advantage of top-down method is the improved characterization of post-translational modifications and reliability of protein quantification. However, this method still needs further development for complex proteomic samples analysis.[6,7] At present, the throughput of top-down method is much lower than the shotgun method. In the shotgun strategy, chemical or enzyme digestion of proteins is used to degrade all of the proteins in the sample to peptides. Liquid chromatography separation is applied to separate the peptides and followed by mass spectrometric analysis. This method is faster and more sensitive as compared to top-down method and is becoming the most popular method for large scale protein identification.[8]

There is another method for proteome analysis using mass spectrometry, called gel-based method. This method separates the proteins in one sample by gel electrophoresis and each individual protein is then analyzed by MS.[9] Liquid chromatography can also be used to separate proteins, but the separation resolution is not as good as gel electrophoresis.[10] Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a very common practice used for one-dimensional gel separation of proteins.[11] In this method, SDS is an anionic detergent that binds and denatures proteins and imparts negative charges to the linearized proteins. Because the charge-to-mass ratio is nearly the same among SDS-denatured proteins, the gel electrophoresis separation of proteins is almost entirely dependent on the protein mass differences. Two-dimensional gel electrophoresis (2-DE) can provide higher separation efficiency compared to one-dimensional gel electrophoresis because the separation is based on two distinct properties of the proteins. The most commonly used 2-DE technique combines a first dimensional separation based on different pI values and a second dimensional separation based on molecular weight differences.[12]

In this thesis, Chapters 2-4 are based on the shotgun approach and Chapter 5 is based on the gel-based method in order to separate individual proteins from complex protein mixture efficiently.

1.1.2 Protein sample preparation

1.1.2.1 Protein extraction and purification

In shotgun proteome analysis, the protein sample preparation includes several steps. The first step is protein extraction from variable proteomic samples, such as cell lines, tissues, and body fluids. Great care needs to be taken in this step to avoid sample loss. The second step involves protein purification. Acetone precipitation is widely used in this step to precipitate proteins from sample solutions. This treatment can effectively remove buffers, salts, and other contaminants.[13]

1.1.2.2 Protein solubilization

After efficient protein extraction and purification, the third step in protein sample preparation is protein solubilization. The overall performance of the shotgun proteome analysis is largely dependent on this step. Complete solubilization of the proteins is required because protein degradation and the following steps need to be performed in solution. Many different solvents are available for protein solubilization. The hydrophobicity of proteins, as indicated by their GRAVY indexes, needs to be considered to determine the type of solvent to be used. Generally, hydrophilic proteins are relatively easy to dissolve in

aqueous solution, such as ammonium bicarbonate buffer. In basic buffer solution (pH=8), the terminal carboxylic acid and the side chain of acidic amino acids are negative charged. Therefore, the protein solubility is increased since the net charges at the protein surface make the protein prefer to interact with water rather than with other protein molecules. Since pH 8 is required for the next step tryptic digestion normally used for protein degradation, no pH adjusting is needed. On the other hand, hydrophobic proteins, such as membrane proteins, cannot be readily dissolved in ammonium bicarbonate buffer. To overcome the hydrophobic interactions between proteins, buffers, chaotropic agents (e.g., urea), aqueous-organic solvents (e.g., methanol-water), organic acids (e.g., formic acid), and surfactants (e.g., sodium dodecyl sulfate (SDS)) could be used. Of these reagents, surfactants are the most powerful in dissolving proteins. Surfactants can be classified into two categories: ionic and non-ionic. The structure of both types of surfactants contains a hydrophobic group and a hydrophilic group. The hydrophobic group in a given surfactant can interact with the hydrophobic domain of the proteins and the hydrophilic group can interact with the aqueous solution, leading to the solubilization of proteins. There are many different types of surfactants that could be used, such as SDS, NP-40, and CHAPS. Among those, SDS, an ionic detergent with the strongest solubilization capability, is widely used to dissolve and denature hydrophobic proteins, including integral membrane

proteins.[14] However, the samples containing SDS cannot be analyzed directly by reversed-phase (RP) LC-MS, as SDS causes interference with the RPLC separation[15,16] and MS analysis.[17,18] Therefore, efficient removal of SDS is required before RPLC tandem MS (MS/MS) analysis. Alternatively, SDS-mimic surfactants, such as acid labile surfactants (e.g., RapiGest from Waters, PPS from Protein Discovery, and Invitrosol from Invitrogen), could be used.[19-26] These reagents rapidly decompose into two easily removed products upon adding an acid, such as trifluoroacetic acid (TFA), to the sample. However, the cost of these reagents is generally high, which could be a major concern in large scale proteome analysis (e.g., proteome profiling of multiple organ or tissue samples) or in situations where large quantity of surfactants is required to process a sample (e.g., working with proteins electro-eluted from a polyacrylamide gel in SDS solution).[27]

1.1.2.3 Protein digestion

After protein solubilization, samples are digested into peptides by using enzymes (e.g., trypsin) or chemical reagents (e.g., cyanogen bromide (CNBr) and acids) prior to the LC-MS/MS analysis. Trypsin is the most widely used among a number of enzymes that are available for protein digestion.[28] It cleaves peptide chains mainly at the carboxyl side of the amino acids lysine or arginine, except

when either is followed by a proline. The primary advantage of using trypsin for protein digestion is the high specificity of this enzyme compared to other enzymes such as chymotrypsin. The second advantage of using trypsin is that the peptides generated would have the ideal molecular weights for MS analysis. Moreover, the peptides generated from tryptic digestion contain the C-terminal lysine or arginine, which would be more readily protonated in electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) compared to other types of peptides with no lysine or arginine in their sequences. For protein samples that do not dissolve well in enzyme digestion buffers or are resistant to enzymatic digestions, chemical reagents could be applied for protein solubilization and degradation. Among the chemical reagents that could be used for specific cleavage of the protein amide bonds, CNBr is one of the most commonly used reagents.[29] CNBr hydrolyzes peptide bonds at the C-terminus of methionine residues with high specificity. Acids could also be employed to facilitate protein degradation. The microwave-assisted acid hydrolysis (MAAH) method was first introduced in 2004.[30] By using a strong acid (6 M HCl) combined with less than 90 s of microwave irradiation, the terminal peptide ladders of a protein could be generated for MALDI MS protein sequencing. Alternatively, by using 25% TFA combined with 7.5 to 10 min irradiation time, a protein could be hydrolyzed into smaller peptides with mass less than 3000 Da, ideal for shotgun proteome

analysis.[31] Many of these small peptides generated by non-specific cutting of a protein have overlapping sequences, which can be very useful to confirm peptide identities from the database search results and generate redundant sequence information to map the protein sequence with high confidence.[32,33] One advantage of MAAH is the shorter reaction time, usually less than 10 min, compared to using the conventional enzyme or chemical reagent. Another advantage is that the protein can be present in any solvent and does not require a specific buffer for the reaction. Therefore, the MAAH method could be used for dealing with very hydrophobic proteins.

While protein samples are fragmented in solution in shotgun proteomics, in the gel-based method, the intact proteins are separated using gel electrophoresis followed by in-gel digestion or degradation. In-gel tryptic digestion is a well-established method for protein identification and sequence analysis.[34,35] In in-gel tryptic digestion, reduction and alkylation are commonly performed following dehydration of the gel. Two additional dehydration steps are sometimes applied to ensure efficient alkylation and in-gel digestion. The digestion time and condition are similar to those of the in-solution method.

1.1.3 Peptide separation methods

The peptide separation is another very important step in shotgun

proteomics. Since the peptide mixture generated from digestion of a given proteomic sample (e.g., whole cell lysate) could be very complex, the resolving power of any single separation procedure will likely not be enough and ion suppression effect among different peptides could be severe. Thus, prior to MS and MS/MS analysis, multidimensional liquid chromatography (MDLC) separation of the peptide mixture is required in shotgun proteome analysis. This technique combines two or more forms of LC to increase the peak capacity, and thus the resolving power, of separations to better fractionate the peptides.[36] The selectivity of the different forms of LC used in MDLC should be orthogonal to each other, and therefore, the peak capacity becomes a product of those in each dimension.

There are many different types of peptide separation techniques available for multidimensional peptide separation, such as reverse phase liquid chromatography (RPLC), RPLC with high-pH elution, strong cation exchange (SCX), strong anion exchange (SAX), weak cation exchange (WCX), weak anion exchange (WAX), hydrophilic interaction liquid chromatography (HILIC), and size exclusion chromatography (SEC). Among these, the combination of SCX and RPLC is the most widely used in proteomics. The online setup of this MDLC technique was first introduced by John Yates in 2001.[29] SCX was used as the first dimensional separation and RPLC served as the second dimensional separation,

followed by MS analysis. In SCX separation of peptides, the surface of the stationary phase in the column contains anions (e.g., $-SO_3^-$ groups). Therefore，the interaction between the positive charged peptides and the column stationary surface is mainly the ionic interaction. Orthogonally, in the RPLC separation, the stationary phase surface of the column contains non-polar functional groups (e.g., C_8 or C_{18}). Hence, the interaction between peptides and column stationary surface is mainly hydrophobic. Since RPLC can provide relatively high separation efficiency compared to other techniques and the mobile phases used are compatible with the peptide ionization techniques (ESI and MALDI) in MS analysis, it is normally used as the last dimensional separation in MDLC. Instead of the on-line setup, off-line SCX-RPLC (Figure 1-1) was used in Chapter 2, Chapter 3, and Chapter 4 of this thesis. In this setup, the complex peptide mixture is first fractionated by SCX. Then, each individual fraction collected is desalted by C_{18} column and quantified by UV absorbance simultaneously.[37] In the end, the optimal amount of peptide from each SCX fraction could be loaded onto the second dimensional RPLC separation followed by MS analysis.

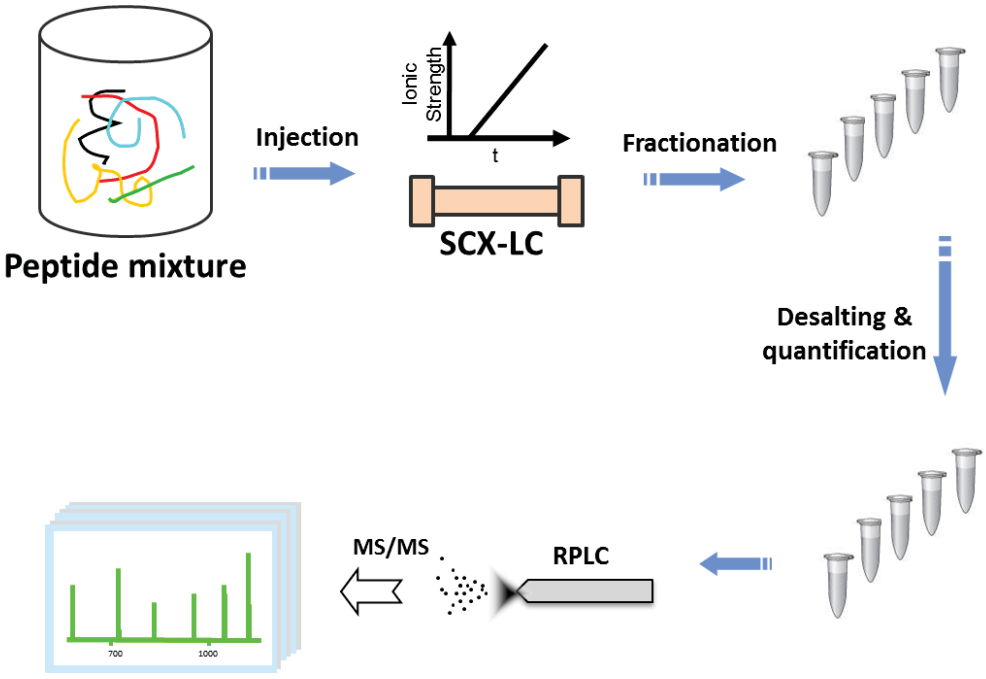


Figure 1-1. Schematic diagram of off-line two dimensional separation of peptide mixture (SCX-RP).

1.1.4 Tandem mass spectrometry (MS/MS)

1.1.4.1 Peptide ionization techniques

To analyze the peptide mixture by tandem mass spectrometry, the peptides need to be ionized and transferred into the gas phase. Electrospray ionization (ESI)[38] and matrix-assisted laser desorption ionization (MALDI)[39] are the two major soft ionization techniques used for peptide ionization.

In ESI, as shown in Figure 1-2 (a), a high voltage (2000 to 4000 V) is applied to a metal capillary through which the liquid flows. The exit of the capillary is placed close to the entrance of a mass spectrometer. The charged liquid forms a Taylor cone at the exit of the metal capillary and then becomes a fine filament. Finally, it breaks into a spray of fine droplets and then gas-phase ions. Since the ESI technique converts the peptides in the solution into charged gas-phase ions without the help of applying heat, the gas-phase ions are generated without thermal-induced decomposition. The advantage of ESI is that the ions generated are mainly multiply charged, which effectively extends the mass range that an analyzer can accommodate. In this thesis, ESI was used as the ionization technique.

Alternatively, as shown in Figure 1-2 (b), MALDI is another soft

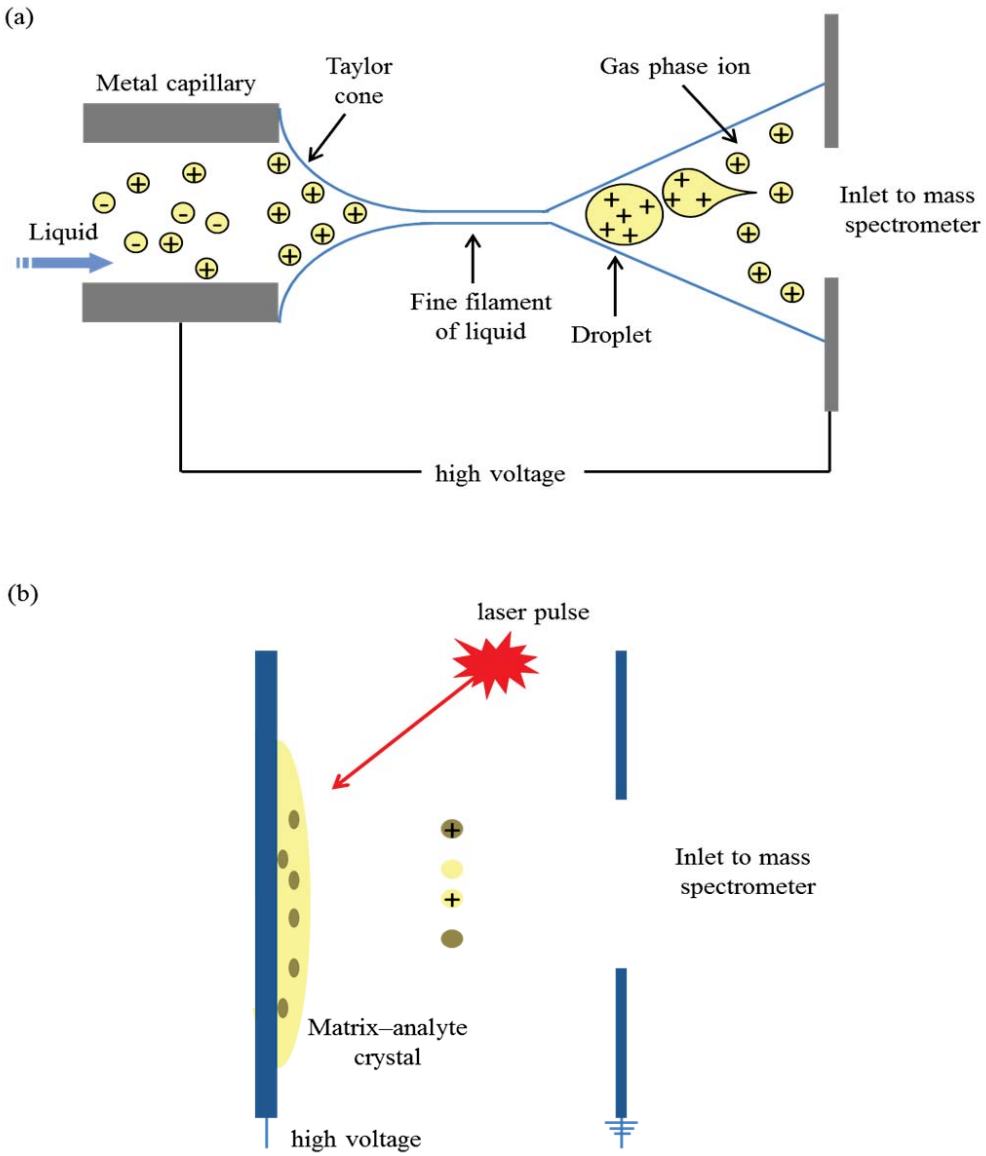


Figure 1-2. Schematic diagram of (a) electrospray ionization (ESI), positive mode; and (b) matrix-assisted laser desorption ionization (MALDI).

ionization technique that could be used to produce peptide ions. In MALDI, peptide samples are mixed with matrix molecules in a solution and a small droplet of the solution is deposited on to a metal plate to form a solid sample. The most commonly used matrix compounds are 2, 5-dihydroxy benzoic acid (DHB),[40] sinapinic acid (SA)[41] and α-cyano-4-hydroxycinnamic acid (CHCA).[42] They are all small organic compounds that can absorb the light from the laser pulse. Therefore, when a laser beam strikes the solid sample on the metal plate, the matrix molecules absorb the photons or energy and are desorbed into the gas phase. At the same time, the peptides which were surrounded by matrix molecules are lifted into gas phase. Since some matrix molecules are ionized during the desorption process, the peptides are converted into ions through collision with the matrix ions in the gas phase. The ions generated by MALDI are predominantly singly charged. This technique provides higher sensitivity and better tolerance to the presence of contaminants such as salts and surfactants than ESI.

1.1.4.2 Instrumentation

After the peptide ions are introduced into gas phase, they will be analyzed by tandem mass spectrometer. There are many different types of instruments available for proteomics analysis such as quadrupole time-of-flight (QTOF), linear ion trap (LTQ)-orbitrap, and triple TOF. Each mass spectrometer

has its unique properties, such as mass range, sensitivity, resolving power, and speed. In this thesis work, the QTOF from Waters is the instrument used for sample analysis. Figure 1-3 shows the schematic diagram of the ESI Q-TOF mass spectrometry from Waters. It is a hybrid orthogonal acceleration time-of-flight mass spectrometer that couples a quadrupole mass filter with an orthogonal acceleration time-of-flight mass spectrometer.[43] In this instrument, the peptides are ionized by ESI. The ZSpray source technology used in the instrument provides high ion transmission efficiency. Since both the analyte spray and the lock mass spray are orthogonal to the sample cone, this design efficiently reduces noise and contamination. After the peptide ions are introduced into the mass spectrometer, they are analyzed in both the MS mode and MS/MS mode. In the MS mode, the quadrupole resolving DC is off (RF only). Therefore, all peptide ions pass through the quadrupole and are directly analyzed by the TOF without fragmentation. In the MS/MS mode, the quadrupole resolving DC is on. Hence, the quadrupole can select one specific precursor ion for fragmentation in the collision cell and then be analyzed by the TOF. In both modes, a high voltage pulse is used in the TOF analyzer to orthogonally accelerate the ions down the flight tube. The reflectron is also used to increase the resolving power of the analysis. In this thesis work, normally, the instrument is set to automatically switch between the MS scan and MS/MS scan mode depending on the data generated from the MS scan,

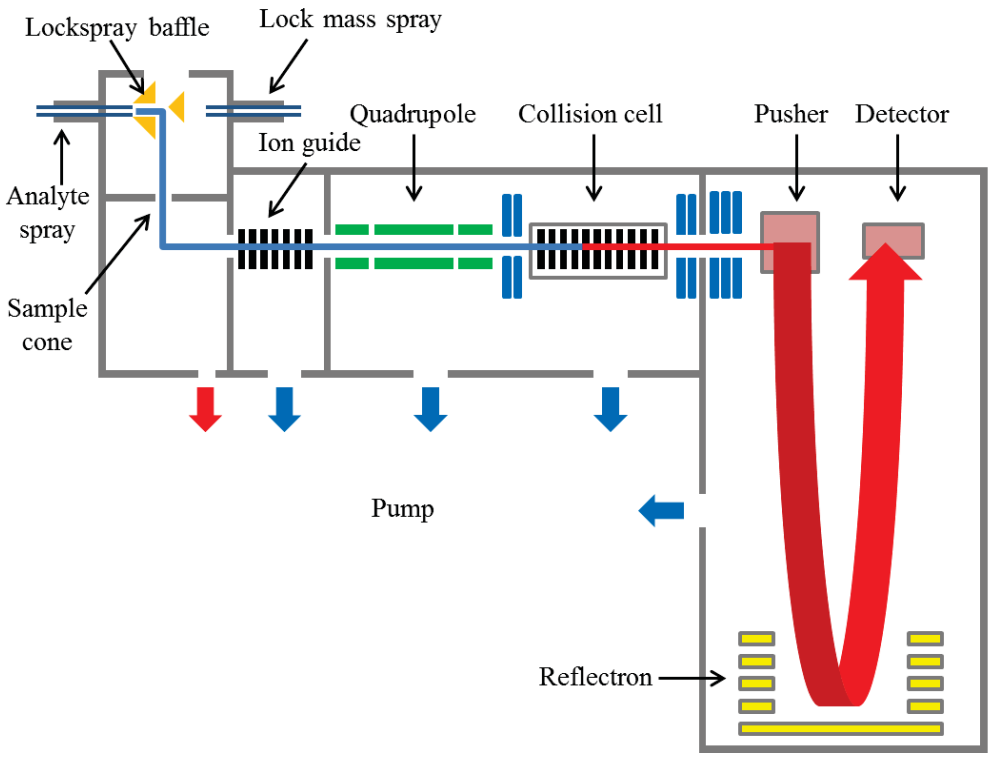


Figure 1-3. Schematic diagram of ESI Q-TOF mass spectrometry from Waters.

which is called data directed analysis (DDA).

1.1.4.3 Peptide ion fragmentation methods

The choice of peptide ion fragmentation method is essential in proteomics work. Currently there are various options available such as collision-induced dissociation (CID),[44] higher-energy collisional dissociation (HCD),[45] electron-capture dissociation (ECD)[46] and electron-transfer dissociation (ETD).[47]

CID is the fragmentation method used in this thesis. During this process, the selected precursor ions are accelerated by an electrical potential to high kinetic energy and then undergoes collision with neutral molecules (e.g., helium, nitrogen, and argon). In the collision, some of the kinetic energy is converted into internal energy of the ions and results in bond breakage and fragmentation of the ions into smaller fragment ions. Generally, b- and y-type fragment ions (shown in Figure 1-4) are most commonly generated in CID due to amide bond breakage. CID is widely used in QTOF, quadrupole ion trap, and fourier-transform ion cyclotron resonance (FTICR) for proteome analysis. HCD is normally used in conjunction with the orbitrap mass analyzer.

In ECD, the peptides are immersed in free electrons at near thermal

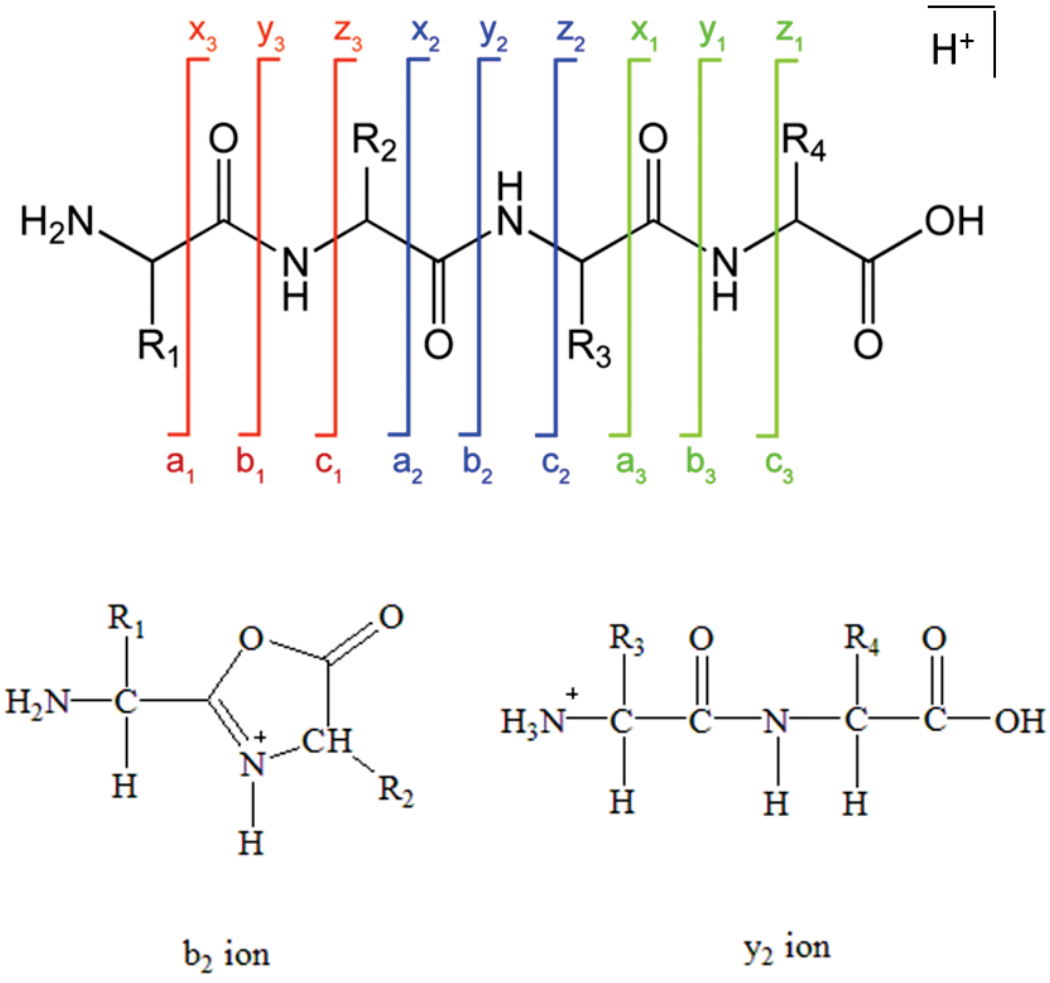


Figure 1-4. Fragmentation pattern of an ionized peptide.

energy. The peptide will convert into radical species if it captures an electron. The radical rearrangement could fragment the peptide into c- and z-type ions (shown in Figure 1-4). Similarly, ETD also cleaves randomly along the peptide backbone and generates c- and z-type ions, while the side chains and modifications are left intact. The electrons are transferred by collision between the analyte cations and reagent anions. Although ECD and ETD can be efficient in analyzing multiply-charged peptides and post-translational modifications, they are not suitable for analysis of small peptides with small charges. Therefore, they complement rather than replace the CID method.

1.1.4.4 Database search

Tandem MS analysis can be efficient in providing unambiguous peptide and protein identification because MS/MS spectrum contains the amino acid sequence information of the peptide rather than just the mass to charge ratio (m/z) from MS spectrum. After data collection, the MS/MS database search is performed for peptide and protein identification.

Many different database search algorithms have been developed such as Mascot (http://www.matrixscience.com/),[48] SEQUEST (http://fields.scripps.edu/sequest/)[49] and X!Tandem (http://www.thegpm.org/TANDEM/).[50] Among these, Mascot is the database

search engine used in this thesis. This search engine is based on the probability-based scoring algorithm MOWSE (Molecular Weight Search). When using Mascot，the mass values of the theoretical peptide sequences and fragment ions are calculated based on the digestion method used (e.g., trypsin) and the fragmentation method (e.g., CID) used. The experimental mass values are then compared with calculated theoretical peptide mass values and fragment ion mass values. Mascot calculates the probability (P) that the observed match between the experimental data set and each sequence database entry is a random event. The score reported by Mascot, $-10\mathrm{Log}_{10}(P)$, is calculated based on this random match probability. The probability based scoring algorithm has several advantages. The first one is that a simple rule can be used to judge whether a result is significant or not. Particularly, this is useful in guarding against false positives. The second advantage is that the scores reported can be compared with those generated from other types of searches. The third advantage is that the search parameters can be readily optimized by iteration. By using Mascot, any FASTA format sequence database can be searched, which makes it more suitable to variable research objectives.

1.2 Mass spectrometric protein quantification

1.2.1 Overview of mass spectrometric protein quantification methods

The aim of quantitative proteomics is to obtain quantitative information about all of the proteins in a given sample.[51] Rather than just identifying proteins in the biological samples, quantitative proteomics yields information about the differences between samples such as tissues from healthy and diseased people. Compared to other methods using dyes, fluorophores or radioactivity, mass spectrometric analysis is not limited to the high abundant proteins and can provide the protein identity together with the quantification analysis.[52] Therefore, mass spectrometric protein quantification methods have gained increasing popularity over the past decade. In the shotgun approach, the protein quantification relies on peptide identification and quantification. Many different methods have been developed for both relative and absolute quantification of proteins. In this chapter, only the most widely used relative quantification methods are introduced. These are label-free quantification, metabolic labeling, and chemical or enzymatic labeling methods.

1.2.2 Label-free quantification methods

Label-free quantification methods are rapid and low-cost compared to

other label-base protein quantification approaches. They are suitable for quantification analysis of large changes of protein expression, but less reliable for the measurement of small fold changes.[53] The most widely used label-free quantification methods are performed using ion peak intensity and spectral counting.

In the label-free quantification approach that measures and compares the ion peak intensity of peptide precursor ions, samples are collected, prepared under identical conditions, and then analyzed using the same instrument and parameters. The m/z values of the peptide precursor ions are detected and ion peak intensities are recorded. Since the signal intensity from ESI is highly correlated with the ion concentration, the relative peptide quantification between the different samples can be determined by comparison of the peptide peak intensities.[54] To improve the performance of this method, a mass spectrometer with high mass accuracy should be used to reduce the interference of signals with similar but distinct masses. Also, a balance between the MS scan and MS/MS scans needs to be found to efficiently identify and quantify the proteins in complex proteomic analysis.

Spectral counting is an alternative label-free protein quantification approach. This approach simply counts the number of spectra identified for a given peptide in different biological samples, which has been demonstrated to

correlate directly with the protein abundance.[54] The spectra of all detected peptides from one specific protein are subsequently integrated for protein quantification. Compared to quantification based on peptide ion intensities, spectral counting benefits from extensive MS/MS scans for both protein identification and quantification.

1.2.3 Metabolic labeling methods

The metabolic labeling methods are in vivo labeling approaches which introduce a stable isotopic signature into proteins during cell growth and division. The first metabolic labeling method was reported in 1999.[55] All amino acids in yeast were labeled with heavy nitrogen (^{15}N) by the use of a ^{15}N-enriched cell culture medium. In 2002, stable isotope labeling by amino acids in cell culture (SILAC) was developed for relative quantification of proteins in mammalian cell lines.[56] This has become the most commonly used approach for in vivo isotopic labeling.

In SILAC, rather than labeling all amino acids with heavy nitrogen (^{15}N), a cell culture medium containing $^{13}C_6$-arginine and $^{13}C_6$-lysine is used to label proteins in mammalian cell lines. As such, after tryptic digestion, all of the resultant peptides (except for the C-terminal peptides) carry at least one labeled amino acid that can be used for quantification. In this approach, protein

identification is based on MS/MS scans that provide the fragmentation spectra of the peptides. The relative quantification is based on the ratio of the peak intensities of heavy labeled and light labeled peptides from MS scans.

In metabolic labeling methods, the heavy labeled and light labeled proteins are combined before sample preparation process. Thus, less error or bias from sample preparation is introduced into the quantification result. These methods are particularly useful for detection of small fold changes of proteins between samples and post-translational modifications under different experimental conditions.

1.2.4 Protein and peptide labeling methods

Metabolic labeling methods are useful for relative protein quantification in cell lines. But for biological samples not amenable to metabolic labeling (e.g., tissue samples and biological fluids), in vitro protein and peptide labeling are necessary for relative quantitative proteomic analysis. These methods introduce isotope labels by chemical or enzymatic derivatization of proteins or peptides. Many protein and peptide labeling methods have been developed with different labeling sites, different labeling reactions, and different quantification mechanisms. In this chapter, the three most popular ones are described.

The isotope-coded affinity tag (ICAT) approach was developed in 1999.[57] The reagent used for protein labeling consists of three elements: an affinity tag (biotin), a heavy labeled or light labeled linker, and a thiol-specific reactive group that can attach to cysteine residues in the proteins. During the analysis, the protein samples to be quantified are heavy labeled and light labeled by the ICAT reagent, respectively. Subsequently, the protein mixtures are combined and digested into peptides. The cysteine-containing peptides with ICAT tag are isolated by avidin affinity chromatography. The final step is the MS and MS/MS analysis of the enriched peptides for protein identification and relative quantification. By using this method, the complexity of the sample is significantly reduced because only the cysteine-containing peptides are enriched for analysis. However, this decreases sequence coverage of the identified proteins and biases against proteins that lack cysteine residues.

In this thesis work (Chapter 3), the dimethylation after guanidination (2MEGA) labeling strategy is used. This method was developed in 2005 and unlike ICAT, the isotopic labeling tags are introduced on the level of digested peptides.[58] The reaction scheme for 2MEGA labeling is shown in Figure 1-5. First, the guanidination reaction selectively modifies the ε-amino group of the lysines to prevent incorporation of multiple labels. Then, the N-terminal of the peptides is isotopically labeled with either heavy labeled or light labeled formaldehyde. After

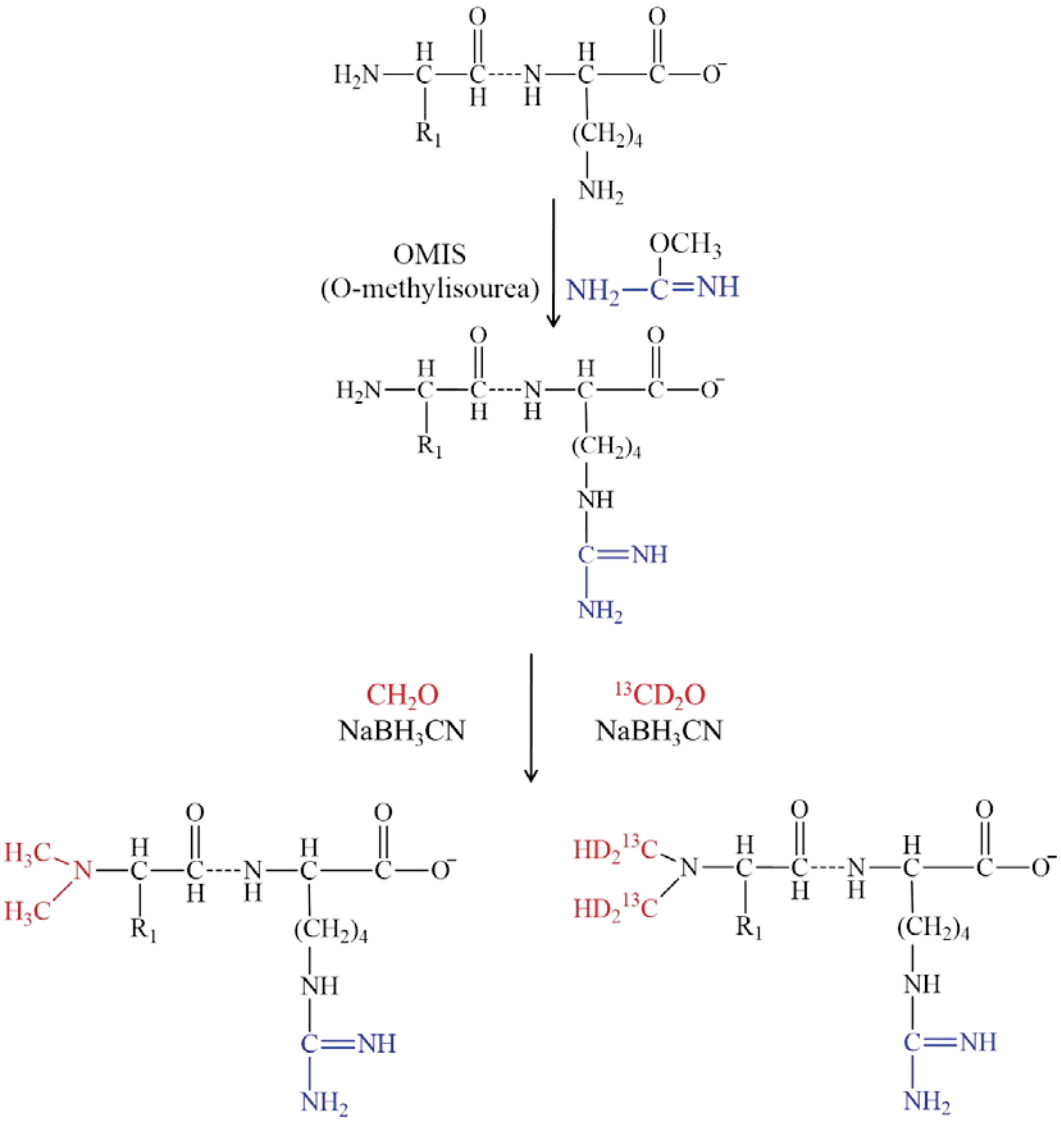


Figure 1-5. Reaction scheme for dimethylation after guanidination (2MEGA) labeling.

that, the labeled samples are combined and subjected to MS and MS/MS analysis. The relative quantification of the proteins is based on the ion peak intensities of the peptides from the MS scan. The MS/MS scan is employed for peptide and protein identification. There are several advantages of 2MEGA labeling for relative proteomic quantification. The first is that this method incorporates a single mass tag in every peptide that is not N-terminally blocked. This eliminates the bias of the analysis and increases the sequence coverage of the identified proteins compared to ICAT. The second advantage is that 2MEGA differentially labeling is inexpensive, which is essential for large scale proteomic quantification. Moreover, the labeling process is simple with high labeling efficiency under mild conditions.

The protein labeling method using a isobaric tag for relative and absolute quantification (iTRAQ) was introduced in 2007.[59] Similar to 2MEGA labeling, the iTRAQ reagents are also introduced on the level of digested peptides. The primary amino groups in the peptides are modified by linking a mass balance group and a reporter group via formation of an amide bond. In the MS scan, the differentially labeled peptides appear as a single peak because of the isobaric mass design of the iTRAQ reagent. The peptide identification and quantification are both performed in MS/MS scan. During the MS/MS analysis, the mass balance group in the tag is released as a neutral fragment. Then, the isotopic reporter ions

are separated with the peptides. The reporter ions are used for relative quantification and the fragment ions of peptides are used for identification. Since there are eight different iTRAQ reagents commercially available, multiplex quantification of protein samples can be performed using this strategy. The nature of this technique leads to the increase of precursor ion intensities and a reduction in the sample complexity relative to multiplexed, precursor-based quantitative methods.

1.3 Analysis of protein phosphorylation

1.3.1 Overview of protein phosphorylation

Protein phosphorylation is one of the most important post-translational modifications. It is reversible and can affect the structure of proteins. Moreover, protein phosphorylation in particular plays a significant role in a wide range of cellular processes such as cell division, cell proliferation, signal transduction, and enzymatic activity.[60] Within a protein, phosphorylation can occur on several different amino acids. Serine phosphorylation is the most common, followed by threonine and tyrosine. The chemical structures of proteins containing phosphoserine, phosphothreonine, and phosphotyrosine are shown in Figure 1-6. The ratio of the occurrence of phosphorylation on phosphoserine:phosphothreonine:phosphotyrosine is 1800:200:1 in vertebrates.[61] The phosphorylation of arginine, histidine, and lysine is also possible, but less common.

Besides mass spectrometric analysis of protein phosphorylation, several other strategies could be used to detect this post-translational modification, including radioactive labeling with ^{32}P-labeled ATP followed by autoradiography detection,[62] and western blot using antibodies for the detection of phosphoserine, phosphothreonine, or phosphotyrosine.[63] The drawback of the labeling method is

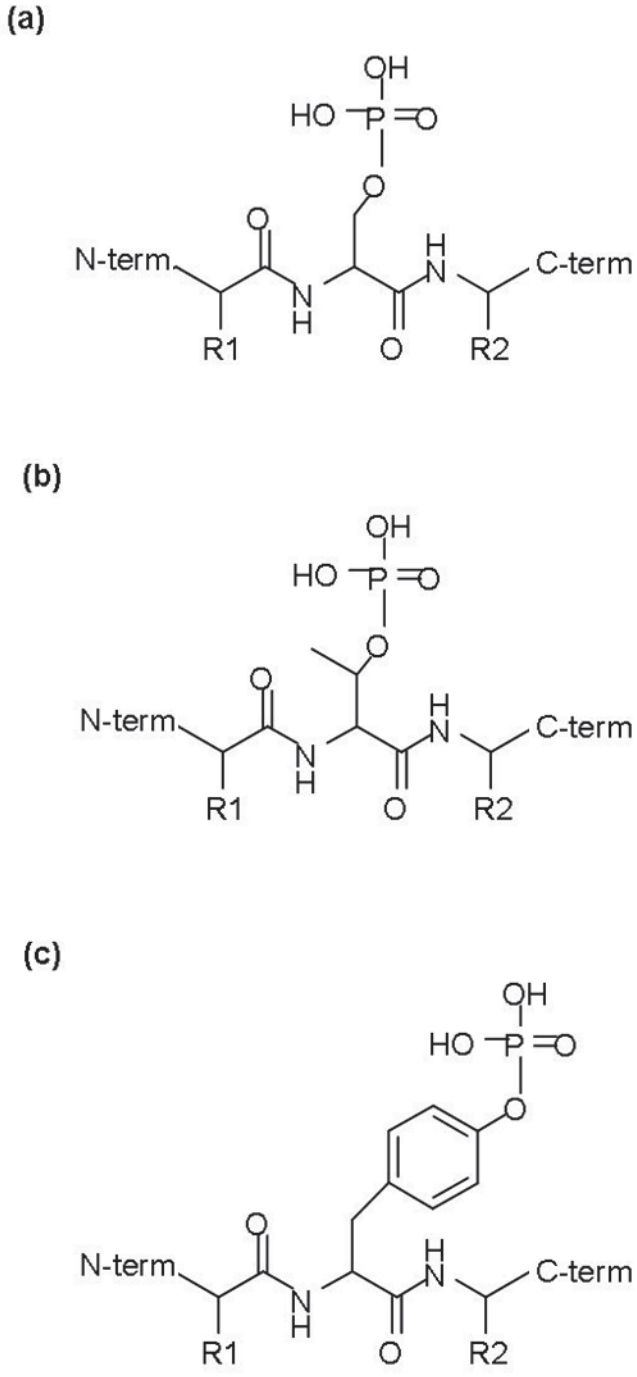


Figure 1-6. Chemical structure of the protein containing (a) phosphoserine; (b) phosphothreonine; (c) phosphotyrosine.

that it cannot provide information about the specific modified sites. For western blot, it relies on the prior knowledge of the type and position of modifications and the availability of antibodies. It has low throughput and not ideal for studying highly complicated samples. Since mass spectrometric analysis of protein phosphorylation can determine the specific site modified, it is more powerful and has been more and more widely used. During MS analysis, the enrichment of phosphopeptides is usually required in shotgun practice because the modified peptides are normally present at a lower level than their native counterparts. There are considerable diversity in the enrichment techniques available. Four of them are described in this chapter.

1.3.2 Phosphopeptide enrichment methods

1.3.2.1 Immunoaffinity enrichment

Phosphopeptides can be enriched by phosphoamino acid-selective antibodies, which is called immunoaffinity enrichment. This approach has been demonstrated to be suitable for identifying tyrosine-phosphorylated peptides.[64] This is why tyrosine phosphorylation sites are relatively well understood despite the lower level of occurrence. On the contrary, the phospho-selective antibodies against phosphoserine and phosphothreonine have not provided satisfactory results for phosphopeptide enrichment. The binding efficiency of the antibody to

the phosphoamino acid is not the same for all of the phosphoserine and phosphothreonine sites.

Since antibodies used for immunoaffinity purification are phosphoamino acid-specific, it is better to use this strategy to search for one type of phosphorylation such as tyrosine phosphorylation. For large scale proteome analysis, the enrichment of different types of phosphorylation needs to be performed in parallel, which makes the process tedious and expensive.

1.3.2.2 Immobilized metal affinity chromatography (IMAC)

Immobilized metal affinity chromatography (IMAC) is an alternative and the most widely used enrichment method for phosphopeptides.[65] In Chapter 5, the phosphopeptides generated from in-gel MAAH were enriched by IMAC. This technique employs a matrix composed of resins containing associated metal ions. These positively charged metal ions can be used to catch negatively charged phosphopeptides in the sample. There are two types of commonly used resins for IMAC, iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA). The chemical structures of IDA and NTA are shown in Figure 1-7. The metal ions could be Fe^{3+},[66] Ga^{3+},[67] Zr^{4+},[68] and others, all of which bind to phosphopeptides with high efficiency and specificity.

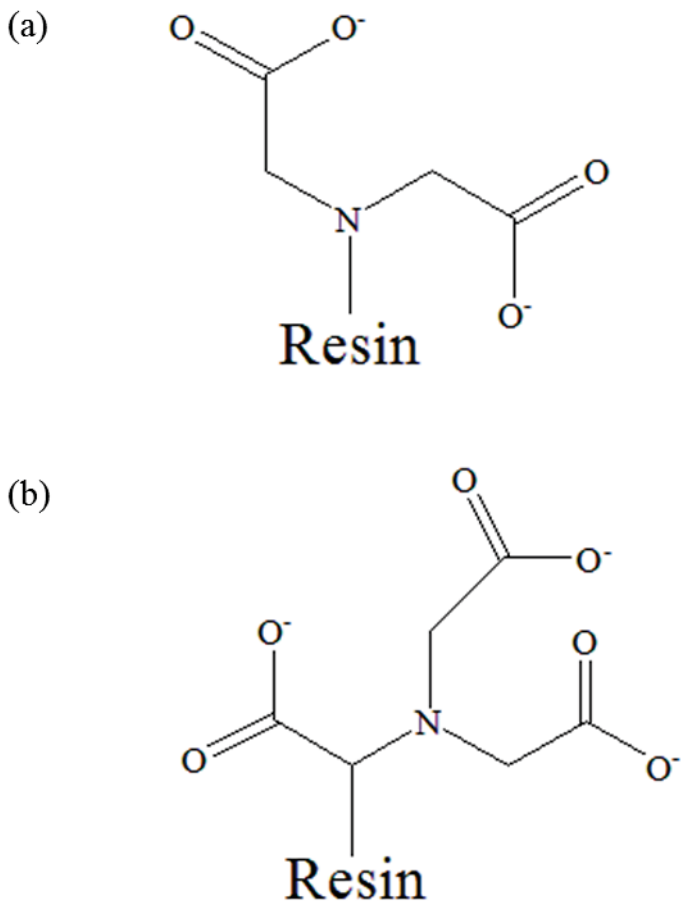


Figure 1-7. Chemical structures of (a) iminodiacetic acid (IDA) and (b) nitrilotriacetic acid (NTA).

The advantage of IMAC resides mainly in its long tradition and large amount of available data that make it possible to adjust the protocol to an optimal level for the desired sample. The main disadvantage of IMAC is the non-specific binding of acidic non-phosphopeptides to the resin. To overcome this drawback, blocking of acidic carboxyl groups by methylesterification can be used to improve the IMAC selectivity. Controlling the pH of the incubation buffer also improves the IMAC specificity. Another disadvantage of IMAC technique is that it is more specific to the multiply-phosphorylated peptides because the interaction between these and the resin is stronger. Therefore, in large scale phosphoproteome analysis, the combination of IMAC and other enrichment methods is applied to enhance the coverage.[69]

1.3.2.3 Metal oxide affinity chromatography (MOAC)

Metal oxide affinity chromatography (MOAC) is another robust phosphopeptide enrichment method. The matrix is composed of metal oxides or hydroxides such as TiO_2, ZrO_2 and $Al(OH)_3$. Currently, the most commonly used MOAC strategy is TiO_2-based.[70] The surface of TiO_2 is positively charged at acidic pH and interacts with negatively charged phosphopeptides. Similar to IMAC, non-specific peptide binding is an issue in MOAC. It could be reduced by methyl-esterification of the carboxyl groups prior to enrichment. Also, the pH of

the loading buffer is always controlled between 2.7 and 2.9 to protonate the acidic amino acid side chains and deprotonate the phosphate groups. In addition, it is reported that the amount of non-specifically binding peptides that were detected negatively correlated with the DHB concentration.[71] This could also improve the MOAC specificity.

1.3.2.4 Fractionation methods

The fractionation methods can be used for the separation of complex peptide mixtures into fractions with higher representation of the phosphopeptides. They are usually applied prior to IMAC or other enrichment methods to reduce the sample complexity and increase the efficiency of the following enrichment. SCX is one of the fractionation methods that is commonly used.[72] As mentioned before, SCX separation is based on the ionic interaction between positive charged peptides and a negative charged column stationary phase. Since most tryptic digested peptides possess at least one basic amino acid, the non-phosphopeptides would have a net charge of +2 in SCX. However, phosphorylated peptides usually have additional negative charges because of the phosphate groups. Thus, a peptide containing a single-phosphate moiety would have a net charge +1 instead of +2. Peptides containing two or three phosphate groups would have a net charge of 0 or -1, respectively, and are not captured by the SCX column. Therefore, the

phosphorylated peptides are present in the flow-through fraction and early elution fractions. Since this method cannot distinguish phosphopeptides from other peptides with the same net charge, a second purification step utilizing IMAC or MOAC is used to remove the remaining non-phosphorylated peptides prior to MS analysis.

1.4 Scope of this thesis

My thesis work focused on developing new sample preparation methods for proteome analysis and protein sequence mapping. In Chapter 2, a protein solubilization method using an acid–labile surfactant was reported. This method can generate better results in membrane proteome analysis compared to SDS-assisted solubilization. In Chapter 3, the acid-labile surfactant, RapiGest, was applied for plasma membrane proteome analysis. A high throughput plasma membrane protein purification, identification, and quantitation strategy was developed. In Chapter 4, an improved shotgun method for analyzing proteomic samples containing SDS was reported. SCX was used to remove SDS efficiently and, at the same time, minimize the sample loss. This improved method could provide even better result compared to the RapiGest-assisted method reported in Chapter 2. Chapter 5 described a newly developed in-gel MAAH MS method that combines the separation power of gel electrophoresis with the high sequence coverage of MAAH MS for improved protein sequence mapping. The applicability of this method was demonstrated by high sequence coverage of various human plasma proteins and successful characterization of the phosphoprotein isoforms of bovine alpha-S1-casein. In the last chapter, future work related to my research was discussed.

1.5 Literature cited

1. Kettman, J. R.; Frey, J. R.; Lefkovits, I., Proteome, transcriptome and genome: top down or bottom up analysis? *Biomol. Eng.* **2001,** *18* (5), 207-12.

2. Bogdanov, B.; Smith, R., Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.* **2005,** *24* (2), 168-200.

3. Ge, Y.; Lawhorn, B.; ElNaggar, M.; Strauss, E.; Park, J.; Begley, T.; McLafferty, F., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **2002,** *124* (4), 672-678.

4. Patrie, S.; Ferguson, J.; Robinson, D.; Whipple, D.; Rother, M.; Metcalf, W.; Kelleher, N., Top down mass spectrometry of < 60-kDa proteins from Methanosarcina acetivorans using quadrupole FTMS with automated octopole collisionally activated dissociation. *Mol. Cell. Proteomics* **2006,** *5* (1), 14-25.

5. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W., Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **1999,** *121* (4), 806-812.

6. Cooper, H.; Hakansson, K.; Marshall, A., The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom. Rev.* **2005,** *24* (2), 201-222.

7. Parks, B.; Jiang, L.; Thomas, P.; Wenger, C.; Roth, M.; Boyne, M.; Burke, P.; Kwast, K.; Kelleher, N., Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers. *Anal. Chem.* **2007,** *79* (21), 7984-7991.

8. Yates, J. R.; Osterman, A., Introduction: Advances in genomics and proteomics. *Chem. Rev.* **2007,** *107* (8), 3363-3366.

9. Ahmed, F., Mining the oncoproteome and studying molecular interactions for biomarker development by 2DE, ChIP and SPR technologies. *Expert Rev. Proteomics* **2008,** *5* (3), 469-496.

10. Tomas, R.; Kleparnik, K.; Foret, F., Multidimensional liquid phase separations for mass spectrometry. *J. Sep. Sci.* **2008,** *31* (11), 1964-1979.

11. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970,** *227* (5259), 680-5.

12. Adams, L. D.; Gallagher, S. R., Two-dimensional gel electrophoresis. *Curr Protoc Immunol* **2005,** *Chapter 8*, Unit 8 5.

13. Margoliash, E., Studies on the growth-promoting factor of adult tissue extracts; the effect of acetone precipitation. *Growth* **1950,** *14* (1), 19-30.

14. Lu, B. W.; McClatchy, D. B.; Kim, J. Y.; Yates, J. R., Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry. *Proteomics* **2008,** *8* (19), 3947-3955.

15. Bosserhoff, A.; Wallach, J.; Frank, R. W., Micropreparative separation of peptides derived from sodium dodecyl sulfate-solubilized proteins. *J. Chromatogr.* **1989,** *473* (1), 71-77.

16. Kawasaki, H.; Suzuki, K., Separation of peptides dissolved in a dodium dodecyl-sulfate solution by reversed-phase liquid-chromatography - removal of sodium dodecyl-sulfate from peptides using an ion-exchange precolumn. *Anal. Biochem.* **1990,** *186* (2), 264-268.

17. Loo, R. R.; Dales, N.; Andrews, P. C., Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Sci.* **1994,** *3* (11), 1975-83.

18. Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A., Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J. Proteome Res.* **2010,** *9* (6), 2863-2870.

19. Ross, A. R. S.; Lee, P. J.; Smith, D. L.; Langridge, J. I.; Whetton, A. D.; Gaskell, S. J., Identification of proteins from two-dimensional polyacrylamide gels using a novel acid-labile surfactant. *Proteomics* **2002,** *2* (7), 928-936.

20. Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C., Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **2003,** *75* (21), 6023-6028.

21. Arnold, R. J.; Hrncirova, P.; Annaiah, K.; Novotny, M. V., Fast proteolytic digestion coupled with organelle enrichment for proteomic analysis of rat liver. *J. Proteome Res.* **2004,** *3* (3), 653-657.

22. Ruth, M. C.; Old, W. M.; Emrick, M. A.; Meyer-Arendt, K.; Aveline-Wolf, L. D.; Pierce, K. G.; Mendoza, A. M.; Sevinsky, J. R.; Hamady, M.; Knight, R. D.; Resing, K. A.; Ahn, N. G., Analysis of membrane proteins from human chronic myelogenous leukemia cells: Comparison of extraction methods for multidimensional LC-MS/MS. *J. Proteome Res.* **2006,** *5* (3), 709-719.

23. Chen, E. I.; McClatchy, D.; Park, S. K.; Yates, J. R., Comparisons of mass spectrometry compatible surfactants for global analysis of the mammalian brain proteome. *Anal. Chem.* **2008,** *80* (22), 8694-8701.

24. Mbeunkui, F.; Goshe, M. B., Investigation of solubilization and digestion

methods for microsomal membrane proteome analysis using data-independent LC-MS(E). *Proteomics* **2011,** *11* (5), 898-911.

25. Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **2007,** *6* (7), 2529-2538.

26. Norrgran, J.; Williams, T. L.; Woolfitt, A. R.; Solano, M. I.; Pirkle, J. L.; Barr, J. R., Optimization of digestion parameters for protein quantification. *Anal. Biochem.* **2009,** *393* (1), 48-55.

27. Smith, J. A., Electroelution of proteins from stained gels. *Current Protocols in Immunology / edited by John E. Coligan* **2002,** *Chapter 8*, Unit 8.8.

28. Fournier, M.; Gilmore, J.; Martin-Brown, S.; Washburn, M., Multidimensional separations-based shotgun proteomics. *Chem. Rev.* **2007,** *107* (8), 3654-3686.

29. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001,** *19* (3), 242-7.

30. Zhong, H.; Zhang, Y.; Wen, Z.; Li, L., Protein sequencing by mass analysis of polypeptide ladders after controlled protein hydrolysis. *Nat. Biotechnol.* **2004,** *22* (10), 1291-6.

31. 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), 471-81.

32. Reiz, B.; Li, L., Microwave-assisted acid and base hydrolysis of intact

proteins containing disulfide bonds for protein sequence analysis by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2011,** *21* (9), 1596-1605.

33. 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.* **2010,** *21* (9), 1573-87.

34. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **1996,** *68* (5), 850-858.

35. Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006,** *1* (6), 2856-2860.

36. Motoyama, A.; Yates, J. R., 3rd, Multidimensional LC separations in shotgun proteomics. *Anal. Chem.* **2008,** *80* (19), 7187-93.

37. Wang, N.; Xie, C.; Young, J. B.; Li, L., Off-line two-dimensional liquid chromatography with maximized sample loading to reversed-phase liquid chromatography-electrospray ionization tandem mass spectrometry for shotgun proteome analysis. *Anal. Chem.* **2009,** *81* (3), 1049-60.

38. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.

39. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988,** *60* (20), 2299-301.

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

41. Beavis, R. C.; Chait, B. T.; Fales, H., Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.* **1989,** *3* (12), 432-435.

42. Beavis, R.; Chaudhary, T.; Chait, B., α-Cyano-4-hydroxycinnamic acid as a matrix for matrixassisted laser desorption mass spectromtry. *Org. Mass Spectrom.* **1992,** *27* (2), 156-158.

43. Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H., High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **1996,** *10* (8), 889-896.

44. Mitchell Wells, J.; McLuckey, S. A., Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol.* **2005,** *402*, 148-185.

45. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007,** *4* (9), 709-712.

46. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998,** *120* (13), 3265-3266.

47. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass

spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (26), 9528-9533.

48. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** *20* (18), 3551-67.

49. MacCoss, M. J.; Wu, C. C.; Yates, J. R., 3rd, Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **2002,** *74* (21), 5593-9.

50. Craig, R.; Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004,** *20* (9), 1466-7.

51. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005,** *1* (5), 252-62.

52. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **2007,** *389* (4), 1017-31.

53. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G., Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005,** *4* (10), 1487-502.

54. Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004,** *76* (14), 4193-201.

55. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl.*

Acad. Sci. U. S. A. **1999,** *96* (12), 6591-6.

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

57. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999,** *17* (10), 994-999.

58. Ji, C. J.; Guo, N.; Li, L., Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. *J. Proteome Res.* **2005,** *4* (6), 2099-2108.

59. Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B., Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007,** *7* (3), 340-350.

60. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science* **2002,** *298* (5600), 1912-34.

61. Hunter, T., The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1998,** *353* (1368), 583-605.

62. Eymann, C.; Becher, D.; Bernhardt, J.; Gronau, K.; Klutzny, A.; Hecker, M., Dynamics of protein phosphorylation on Ser/Thr/Tyr in Bacillus subtilis. *Proteomics* **2007,** *7* (19), 3509-26.

63. Kaufmann, H.; Bailey, J. E.; Fussenegger, M., Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics* **2001,** *1* (2), 194-199.

64. Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X. M.; Polakiewicz, R. D.; Comb, M. J., Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **2005,** *23* (1), 94-101.

65. Andersson, L.; Porath, J., Isolation of phosphoproteins by immobilized metal (Fe^{3+}) affinity chromatography. *Anal. Biochem.* **1986,** *154* (1), 250-4.

66. Neville, D. C. A.; Rozanas, C. R.; Price, E. M.; Gruis, D. B.; Verkman, A. S.; Townsend, R. R., Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. *Protein Sci.* **1997,** *6* (11), 2436-2445.

67. Posewitz, M. C.; Tempst, P., Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **1999,** *71* (14), 2883-2892.

68. Feng, S.; Ye, M. L.; Zhou, H. J.; Jiang, X. G.; Jiang, X. N.; Zou, H. F.; Gong, B. L., Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis. *Mol. Cell. Proteomics* **2007,** *6* (9), 1656-1665.

69. Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J.; Zou, H.; Mohammed, S., Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **2013,** *8* (3), 461-80.

70. Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R., Selective isolation at the femtomole level of phosphopeptides from proteolytic

digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **2004,** *76* (14), 3935-3943.

71. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. D., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **2005,** *4* (7), 873-886.

72. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P., Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (33), 12130-12135.

Chapter 2 Comparison of Surfactant-Assisted Shotgun Methods Using Acid-Labile Surfactants and Sodium Dodecyl Sulfate for Membrane Proteome Analysis

This work was done under the supervision of Dr. Fang Wu. Difei Sun contributed partially to the sample preparation, data acquisition and data analysis. A version of this chapter was published as Wu, F.; Sun, D. F.; Wang, N.; Gong, Y.; Li, L., Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis. *Anal. Chim. Acta* 2011, *698* (1-2), 36-43.

2.1 Introduction

Due to the high hydrophobicity of membrane proteins, characterization of the membrane proteome remains a challenging task. Membrane proteins can be defined as all of the proteins attached to or associated with the membrane of a cell or organelle.[1-5] Polyacrylamide gel electrophoresis (PAGE) separation of proteins followed by mass spectrometric identification has been widely used for proteome analysis, including for membrane protein identification.[6-10] In the past several years, shotgun proteomics has been rapidly developed as a complementary method for membrane protein identification.[11,12] Shotgun methods usually require the digestion of the solubilized proteins into complex peptide mixtures which are then analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS).[13-15] The overall performance of a shotgun proteome analysis method is very much dependent on the efficiencies of the protein solubilization and digestion. In analyzing membrane proteins, a challenging issue is related to the difficulty in maintaining high solubility of the hydrophobic proteins throughout the entire isolation and separation process while avoiding reagents which may interfere with LC-MS/MS. Considerable efforts have been devoted to the development of improved sample handling techniques tailored to membrane proteome analysis. Of particular interest, surfactants have been employed to mimic a lipid-like environment on the inside of the micelles and achieve varying

degrees of solubilization of membrane proteins, allowing subsequent protein digestion in a lower concentration of detergent.[1,16] There are different types of surfactants, including ionic surfactants (e.g., sodium dodecyl sulfate or SDS), non-ionic surfactants (e.g., Triton X-100) and zwitterionic surfactants (e.g., CHAPS).[17] Among them, SDS is one of the most commonly used surfactants due to its ability to assist in the solubilization of membrane proteins.[17,18] Trypsin digestion is not significantly affected by the presence of SDS as long as the SDS concentration is low (<0.1%), which is usually achieved by diluting the SDS-solubilized protein sample prior to the addition of trypsin.[19-21] However, SDS, even at low concentrations, is notoriously incompatible with mass spectrometric peptide detection, particularly with electrospray ionization (ESI). SDS also reduces the separation power of RPLC. Therefore, removal of SDS is required prior to LC-MS/MS. Dialysis,[22-25] ultrafiltration device[26-28] and strong cation exchange (SCX)[19, 29] have been used to remove SDS, although the SDS removal efficiency with these methods for peptide samples has not been extensively investigated. More recently, a spin column has been introduced for removing detergents including SDS.[30] According to the manufacturer, 96% of SDS could be removed from a sample containing a small protein (i.e., insulin, ~5.7 kDa) in SDS and the sample recovery rate was found to be 84%. However,

the SDS removal efficiency and the sample recovery rate for peptides (<3 kDa) have not been reported.

As an alternative to SDS, several MS-compatible strong surfactants have been developed and some of them are commercially available. For example, an acid labile surfactant, RapiGest from Waters, is designed to structurally resemble SDS; it consists of an ionic moiety (sulfonate) and a hydrophobic alkyl chain (undecyl).[31-33] Other acid labile surfactants include 3-[3-(1,1-bisalkyloxyethyl)pyridin-1-yl]propane-1-sulfonate (PPS) from Protein Discovery and Invitrosol from Invitrogen. It was reported that trypsin enzyme activity was not significantly decreased in 0.1-1% RapiGest solution.[34-36] RapiGest rapidly decomposes into two products by adding trifluoroacetic acid (TFA) to the digested protein samples (pH<2) and incubation at 37 °C. One product is water immiscible and can be removed by centrifugation. The other is a water-soluble product that can be used for LC-MS/MS. Several studies have shown that RapiGest enables the proteome analysis using the gel-based method as well as the shotgun method.[37-44] For the shotgun method, the relative merits of RapiGest *vs*. several commonly used reagents, such as methanol and urea, have been investigated.[32,40,44] RapiGest was found to be superior for whole cell extract analysis. For membrane proteome analysis, the applicability of RapiGest and PPS has been reported.[45,46] However, there is no report of a direct comparison of

acid-labile surfactants such as RapiGest and PPS with SDS for shotgun membrane proteome analysis.

In this work, we report an investigation on the performance of RapiGest-PPS- and SDS-based sample preparation methods for shotgun membrane proteome analysis. The membrane fractions of the *E. coli* cells and the MCF7 cells were used as model systems for performance comparison. After determining that the RapiGest method could identify significantly more peptides and proteins in LC-ESI MS/MS, we applied the RapiGest method, in combination with two-dimensional LC-MS/MS, to generate a proteome map from a membrane-enriched fraction of *E. coli* cell extracts. It was found that the proteome consisted of a large number of membrane proteins including many transmembrane proteins with multi-transmembrane domains, illustrating that the RapiGest method was applicable to handle the membrane proteome with high performance.

2.2 Experimental section

2.2.1 Chemicals and reagents

Dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), potassium chloride (KCl), potassium phosphate monobasic (KH_2PO_4), sodium dodecyl sulfate (SDS) and bovine serum albumin (BSA) were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). Sequencing grade modified trypsin, LC/MS-grade formic acid, acetonitrile (ACN), and Optimal LC/MS-water were from Fisher Scientific Canada (Edmonton, Canada). RapiGest^{TM} SF was purchased from Waters (Milford, MA). PPS was purchased from Protein Discovery (Knoxville, TN).

2.2.2 Instrumentation

French press (Aminco Rochester, NY) was used to prepare the cell lysate. Strong-cation exchange (SCX) chromatography was done using an Agilent 1100 HPLC system with a 2.1 × 250 mm polySULFOETHYL^{TM} A column (Part #: 252SE0503, PolyLC, Columbia, MD). An Agilent 1100 HPLC system using a reversed-phase (RP) 4.6 × 50 mm C18 column (Part #: 2001-050×046, Varian, Ontario, Canada) equipped with a diode-array UV detector was employed to desalt and quantify the peptide mixture. A quadrupole time-of-flight (QTOF)

Premier mass spectrometer (Waters, Milford, MA) equipped with a nanoACQUITY Ultra Performance LC system (Waters) was used to analyze the peptide samples.

2.2.3 *E. coli* membrane fraction preparation

A single colony of *Escherichia coli* K-12 (ATCC 47076) was taken to inoculate 50 mL LB medium in a 250 mL flask. The culture was grown at 35 °C overnight. About 12.5 mL of the saturated culture were put into a 500 mL LB media in baffled 2 L Erlenmeyer flask and grown at 37 °C for 3 hours with shaking at 4,400g. Cells (A_{600}=3.0) were harvested by centrifugation at 3,200g for 15 min at 4 °C. After the cells were washed with 150 mL phosphate-buffered saline (PBS) (1.4 mM NaCl, 0.27 mM KCl, 1 mM Na_2HPO_4, 0.18 mM KH_2PO_4, pH 7.4), they were centrifuged at 3,200g for 15 min at 4 °C. The cells were resuspended in 30 mL water and 3 tablets of Roche mini protease inhibitor cocktail were added. The suspension was passed through a French press (Aminco Rochester, NY) at 20,000 psi twice. The lysate was centrifuged at 2,300g for 15 min to remove the unbroken cells. The supernatant was collected and centrifuged at 118,000g for 55 min. The pellet was then suspended in 20 mL of 50 mM ammonium bicarbonate and centrifuged at 166,811g for 40 min. The supernatant

was discarded and the pellet contained membrane proteins. The pellet was stored at -20 °C.

2.2.4 MCF7 membrane protein fraction preparation

MCF7 cell line was maintained in Dulbecco's Modified Eagle's Medium (Sigma, Ontario, Canada), supplemented with 10% heat-inactivated fetal bovine serum (Gibco, Grand Island, NY), antibiotics (10 mg/mL streptomycin, and 10,000 U/mL penicillin). Cells were grown to 90% confluency and washed twice with ice cold phosphate buffered saline (PBS). Cell lysates were collected using Triton X-114 lysis buffer (50 mM Tris HCl, 300 mM NaCl, 5 mM EDTA, 2% Triton X-114 and protease inhibitor cocktail, pH 7.5). The lysates were then centrifuged at 20,000g for 30 min at 4°C. The supernatant was transferred to a new tube and incubate at 37 °C for 30 min, followed by centrifugation at room temperature at 20,400g for 30 min to separate the phases. Four mL of the detergent phase was then collected and 25 mL of ice cold PBS was added to the detergent phase. The mixture was then stirred for 1 h at 4 °C, followed by centrifugation at 20,400g for 30 min at room temperature. This step was repeated once. Fifty mL of PBS was then added to the detergent phase and incubated at 37 °C for 30 min, followed by centrifugation at 13,300g for 20 min. Fifteen mL of the detergent phase was collected. About 140 mL of acetone was added to the

detergent phase and incubated at -20 °C. After overnight incubation, the sample was centrifuged at 20,400g for 30 min. The pellet enriched with the integral membrane proteins was then stored at -80 °C for further use.

2.2.5 Protein solubilization and digestion

The *E. coli* membrane protein pellets were divided into two equal parts and then solubilized individually in 1% RapiGest, 1% PPS or 1% SDS. Standard reduction and alkylation were carried out on the proteins (~300 μg). Reduction for 1 h at 37 °C was done by the addition of 6 μL 900 mM DTT, followed by alkylation for 1 h at room temperature in the dark by the addition of 12 μL 900 mM IAA. RapiGest-assisted and SDS-assisted solubilization and digestion were carried out as previously described with some modifications.[19, 34] In brief, 1% (w/v) RapiGest was used to re-suspend the *E. coli* membrane proteins. Following the manufacturer's instructions, the proteins and RapiGest mixture were boiled at ~100 °C for 5 min to enhance the protein solubilization. The sample was then cooled down before adding trypsin solution to the protein solution at a ratio of 1 to 50 after 10-fold dilution. The mixture was then incubated at 37 °C overnight. 10% TFA was added to the peptide mixture to make the final TFA concentration approximately 0.5% (pH<2). The mixture was incubated at 37 °C for 45 min and then centrifuged at 18,000 g for 10 min. The supernatant was carefully transferred

into another vial and was store in -80 °C for future use. For the PPS-assisted solubilization and proteolysis experiment, the manufacturer's instructions, similar to the RapiGest protocol described above except that the digest was incubated at room temperature for 1 h after acidification, were followed. For the SDS-assisted solubilization and proteolysis experiment, 1% SDS (w/v) was used to solubilize the membrane pellet. Trypsin solution was added into the protein solution at a ratio of 1 to 50 after 10-fold dilution and the mixture was incubated at 37 °C overnight. 10% TFA was added to acidify the peptide mixture. The peptide mixture was store in -80 °C for future use. The same trypsin digestion procedure was applied to BSA. The same protocols used for processing the *E. coli* sample were applied to the MCF7 membrane protein sample.

2.2.6 SCX in SDS removal and desalting

The digest was subjected to SDS removal by SCX chromatography. Buffer A in SCX was 10 mM KH_2PO_4 and buffer B was 500 mM KCl and 10 mM KH_2PO_4. After sample loading onto the column, the column was flushed with Buffer A for 5 min at a flow rate of 0.25 mL/min to remove the SDS. The peptides were eluted by using 100% buffer B and collected based on the UV signal at 214 nm without fractionation. All the samples collected from SCX were concentrated to ~100 μL using a SpeedVac. An Agilent 1100 HPLC system using

a reversed-phase (RP) 4.6 × 50 mm C18 column (Part #: 2001-050×046, Varian, Ontario, Canada) was employed to desalt the samples collected from the SCX and the peptide amount was measured based on the UV absorbance.

2.2.7 SCX separation in 2D-LC MS/MS

In 2D-LC MS/MS analysis of the *E. coli* sample, the RapiGest digest was first separated by SCX on the Agilent system. Gradient elution was performed with mobile phases A (10 mM KH_2PO_4, pH 2.76) and B (10 mM KH_2PO_4, pH 2.76, 500 mM KCl) at a flow rate of 0.2 mL/min. The gradient profile was as follows: 0 min: 0% B, 2 min: 4% B, 40 min: 60% B, and 47 min: 100% B. A total of 28 fractions were collected and directly desalted and quantified by the HPLC-UV desalting setup described above.

2.2.8 LC-ESI QTOF MS and MS/MS analysis

This step was performed as previously described with minor changes.[18] In brief, about 1.5 µg of peptides of each desalted sample were analyzed using a QTOF Premier mass spectrometer equipped with a nanoACQUITY Ultra Performance LC system. The RPLC separation was done using a 250 min gradient elution between Solvent A and Solvent B. Solvent A was 0.1% formic acid in water and Solvent B was 0.1 % formic acid in ACN. The gradient was as follows:

10-20 % B for 138 min, 20-30 % B for 70 min, 30-45 % B for 15 min, 45-85 % B for 15 min, and 85-95 % B for 8 min. All samples were sprayed into the mass spectrometer at a flow rate of 350 nL/min. MS data were recorded with an *m/z* window of 350-1600 for 0.8 s, followed by 4 data dependent MS/MS scans of the four most intense ions with an m/z window of 50-1990 for 0.8 s. Various collision energies were used based on the mass and charge state of the peptide. Mass scans of a mixture of mass calibrants consisting of leucine enkephalin and (Glu1)-Fibropeptide B was acquired for 1 scan/min throughout the run.

2.2.9 Protein database search

Database searches were performed as previously described with minor changes.[18] Raw search data were lock-mass-corrected, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.2.5 (Waters). Peptide sequences were identified by automated database searching of the peak list files using the Mascot search program. Database searching was restricted to *E. coli* or *Homo sapiens* (human) in the Swissprot or BSA. The following search parameters were selected for all database searching: enzyme, trypsin; missed cleavages, 1; peptide tolerance, 30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, 1+, 2+, and 3+; fixed modification, carbamidomethyl (C); variable modifications, oxidation (M). The search results, including protein names, access IDs, molecular mass, unique

peptide sequences, ion score, MASCOT threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses were extracted to Excel files using in-house software. All of the identified peptides with scores lower than the Mascot threshold score for identity at the confidence level of 95% were then removed from the protein list. The redundant peptides for different protein identities were deleted, and the redundant proteins identified under the same gene name but different access ID numbers were also removed from the list. Specifically, the final unique protein or peptide list was generated by merging all of the protein or peptide lists from the individual runs according to the following roles: only unique proteins (under unique gene names) and peptides with the highest scores were kept; each peptide was only associated to one unique protein; only the first hit within each identified protein group was kept in the list as a representative protein. Redundant peptides with lower identification scores were removed. And redundant proteins with either lower scores or lower number of peptides were also removed.

To gauge the false positive peptide matching rate in our analysis, we applied the target-decoy search strategy by searching the MS/MS data against the forward and reversed proteome sequences.[47,48] The false positive matching rate was calculated by using the equation: $2\times n(rev)/[n(forward)+n(rev)]$, where n(rev)

and n(forward) are the number of matches from the reversed (decoy) and forward (correct) sequence, respectively.

All peptides and proteins identified were examined using the ProtParam program available at the EXPASY web site (http:// us.expasy.org/tools/protparam.html), which allows calculation of the grand average of hydropathy (GRAVY). The pI calculation tool from the EXPASY web site (http://ca.expasy.org/tools/pi_tool.html) was used to calculate the isoelectric points of the peptides (pI values).

2.3 Results and discussion

In this study, we investigated the performance of two MS-compatible surfactants (RapiGest and PPS) and SDS-based sample preparation methods for shotgun membrane proteome analysis. We used the *E. coli* membrane fraction as the first model system. Table 2-1 summarizes the number of peptides and proteins identified from two replicate sample preparations of the *E. coli* membrane fraction with SDS, PPS or RapiGest. In each sample preparation, triplicate LC-ESI MS/MS runs were carried out and, thus, each peptide or protein number shown in Table 2-1 represents the combined results of three LC MS/MS runs. An average of 1529±56 (n=6) unique peptides and 421±14 (n=6) unique proteins were identified from one run of the RapiGest sample, compared to 922±27 peptides and 307±7 proteins identified from the PPS sample, and 960±112 peptides and 341±26 proteins identified from the SDS sample. The reproducibility of the triplicate LC MS/MS runs averaged 86.5% between two runs in terms of the common proteins identified. After merging the protein identification results shown in Table 2-1, 554 unique proteins were identified from the RapiGest sample, 419 proteins were identified from the PPS sample and 477 proteins were found in the SDS sample. Figure 2-1 (a) shows the protein number comparison between the SDS, PPS and RapiGest samples. In total, 281 common proteins were

Table 2-1. Summary of the numbers of peptides and proteins identified from two replicate sample preparations of the E. coli membrane proteome fraction using RapiGest, PPS or SDS. Each sample was analyzed by triplicate LC MS/MS runs.

Sample preparation method	# of peptides identified				# of proteins identified			
	Run#1	Run#2	Run#3	Average	Run#1	Run#2	Run#3	Average
RapiGest Prep#1	1491	1506	1464	1487±17	411	410	404	408±3
RapiGest Prep#2	1626	1541	1550	1572±38	438	435	429	434±4
PPS Prep#1	901	904	891	898±7	300	303	300	302±2
PPS Prep#2	944	937	956	945±10	316	314	313	314±2
SDS Prep#1	889	839	851	859±21	324	321	307	317±7
SDS Prep#2	1081	1062	1039	1060±17	363	367	364	365±2

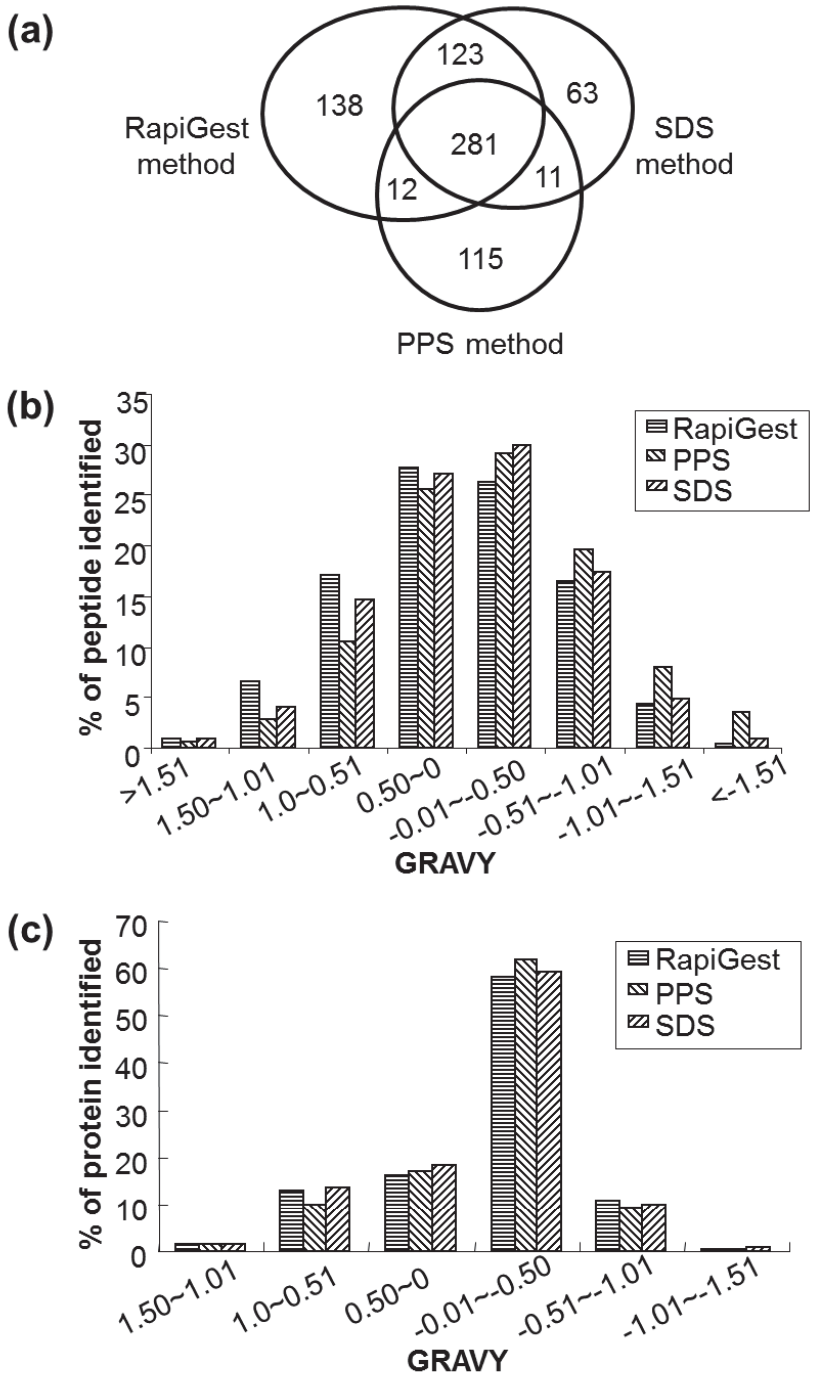


Figure 2-1. (a) Distribution of numbers of proteins identified by the RapiGest, PPS and SDS methods. Distribution of the hydrophobicity of (b) peptides and (c) proteins identified from the E. coli samples prepared by the three methods.

identified from the three methods, while 138, 115 and 63 proteins were uniquely identified from the RapiGest, PPS and SDS samples, respectively. Combining the results from the three methods, 739 different proteins were identified.

To gauge the effect of the peptide and protein properties on the detectability of the three methods, Figure 2-1 (b) shows the distributions of the peptides identified in the RapiGest, PPS and SDS samples plotted as a function of GRAVY indexes. For the RapiGest sample, a slightly higher proportion of very hydrophobic peptides (index > +0.51) were identified, compared to the SDS sample, while a slightly higher proportion of very hydrophilic peptides (index < -0.51) and a slightly lower proportion of very hydrophobic peptide (index > +0.51) were detected from the PPS sample. Similar proportions of the mildly hydrophobic peptides (Index = -0.5 to +0.5) were detected in the three samples. At the protein level, Figure 2-1 (c) shows the distribution of the percentage of proteins as a function of their calculated GRAVY values. Out of 554 proteins identified from the RapiGest sample, 171 proteins (30%) are considered to be hydrophobic with positive GRAVY values, while 153 proteins (32%) out of 477 proteins identified from the SDS sample and 119 proteins (29%) out of 419 proteins identified from the PPS sample are hydrophobic with positive GRAVY. Although the total number of proteins identified was different (554 for RapiGest, 477 for SDS and 419 for PPS), the normalized GRAVY distribution of the

proteins found in RapiGest is almost the same as that in SDS, indicating that there is no bias for detection at the protein level by the two methods. However, a slightly lower number of hydrophobic proteins were detected from the PPS sample.

Of particular interest, information on the percentage of membrane proteins identified in the samples can be deduced. Figure 2-2 shows the protein number distributions as a function of protein locations in the cells. About 65 to 80% of the identified proteins have known cellular locations. Among them, for both the SDS and RapiGest samples, the proportion of membrane proteins identified was quite high (84.8% in Figure 2-2 (a) for the SDS sample and 85.3% in Figure 2-2 (b) for the RapiGest sample). For the PPS sample, the proportion of membrane proteins identified was 67.7% (Figure 2-2 (c)), but many unclassified proteins could be membrane proteins. For the combined results, 78.3% of the identified proteins with known cellular locations are membrane proteins (see Figure 2-2 (d)).

The above results indicate that there is no apparent bias for the type of peptides and proteins identified by the SDS and RapiGest methods and there is a small bias towards the detection of more hydrophilic peptides and proteins by the PPS method. The number of peptides and proteins identified is similar for the

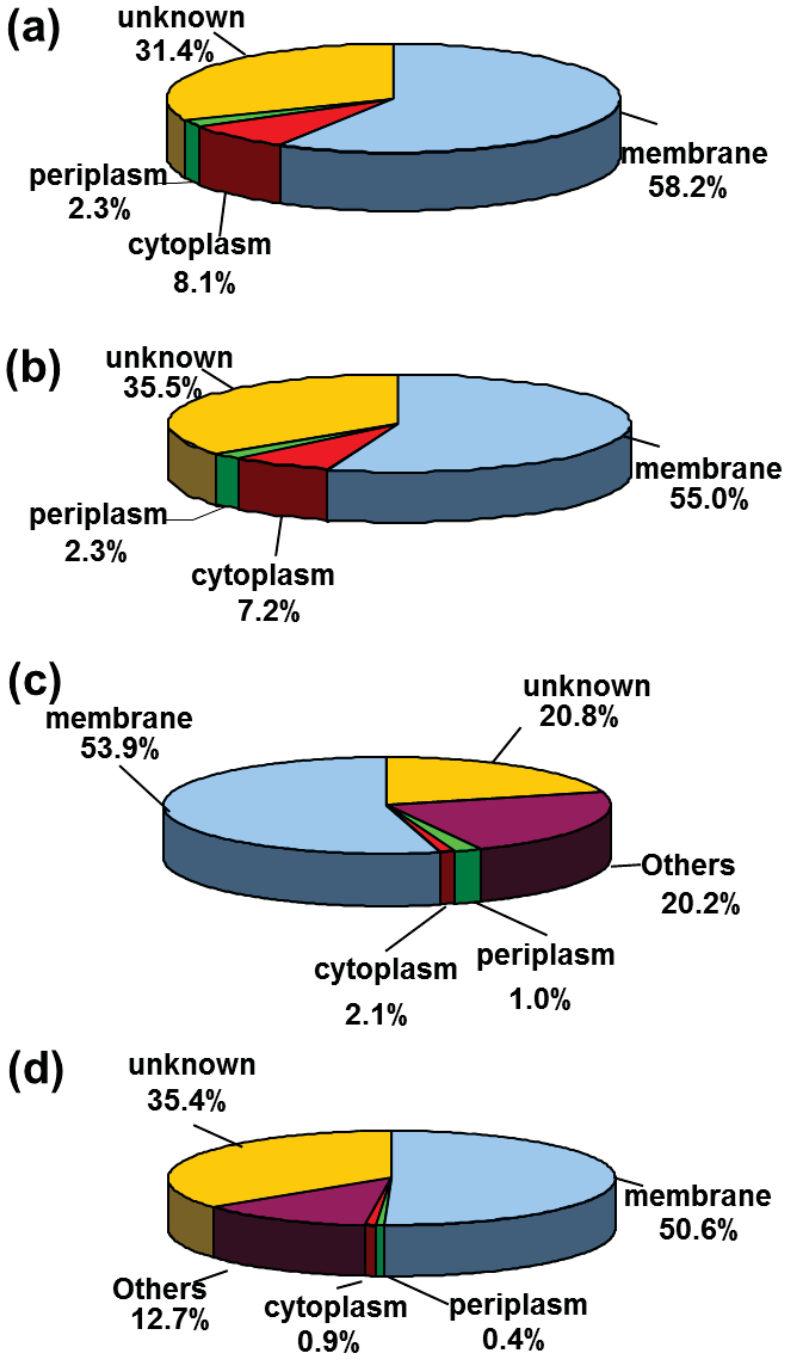


Figure 2-2. Distribution of the cellular locations of proteins identified from (a) the SDS sample, (b) the RapiGest sample, (c) the PPS sample, and (d) three samples combined.

SDS and PPS methods. However, the use of RapiGest allows identification of more peptides and proteins than PPS or SDS. Note that the same amount of digest (1.5 μg) was injected into LC MS/MS for peptide identification. Thus, the differences seen are due to the peptide composition difference between the two samples. Comparing to the RapiGest method, a lower number of peptides and proteins were identified from the PPS method, which is consistent with other reported results of membrane proteome analysis.[45,46] The two surfactants have different proteolysis efficiencies; the RapiGest surfactant has better denaturing property than PPS for membrane proteins.[46]

As shown in Table 2-1, the number of peptides and proteins identified from the SDS sample is similar to the PPS method, but significantly lower than the RapiGest method. SDS dissolves proteins very efficiently and trypsin digestion can be done with high efficiency as well, as long as the SDS content is less than 0.1%.[20] Thus, the difference between the SDS and RapiGest methods is mostly likely related to sample handling. For the results shown in Table 2-1, the SDS sample was treated by passing through the SCX column to remove the SDS while the RapiGest sample was not subjected to the SCX column cleaning. It is clear that, after passing through the SCX column, the peptide composition of the SDS sample was altered. Realizing that sample loss must have occurred during the SCX cleaning process, another set of experiments were carried out to gauge

the extent of sample loss. An aliquot of the desalted peptide sample from the SDS preparation was re-injected into the SCX column, followed by applying the same protocol of cleaning and peptide elution as that used in the initial SCX SDS removal step. By comparing the peptide amount injected with that recovered from the SCX column, it was found that only 38% of the sample amount was recovered. Similarly, for the RapiGest sample, after the SCX column cleaning, 41% the sample amount was recovered. For the SDS sample after the SCX cleaning, we profiled the peptides in the sample by LC-ESI MS/MS operated at the same condition as that used for analyzing the initial SDS sample with one passage of SCX cleaning. In this sample, 664 unique peptides and 297 unique proteins were identified, compared to an average of 1060 peptides and 365 proteins identified in the initial sample. This result further indicates that sample loss during the SCX cleaning step resulted in smaller numbers of peptides and proteins identified.

To further study the effect of SCX cleaning on peptide composition and number of peptides identified, another set of LC MS/MS runs were performed to directly compare the peptide identification results generated from the RapiGest sample before and after the SCX cleaning. A total of 1471 peptides and 419 proteins were identified from the RapiGest sample before the SCX cleaning, compared to 543 peptides and 264 proteins from the RapiGest sample after the SCX cleaning. Figure 2-3 (a) shows the GRAVY distributions of the peptides

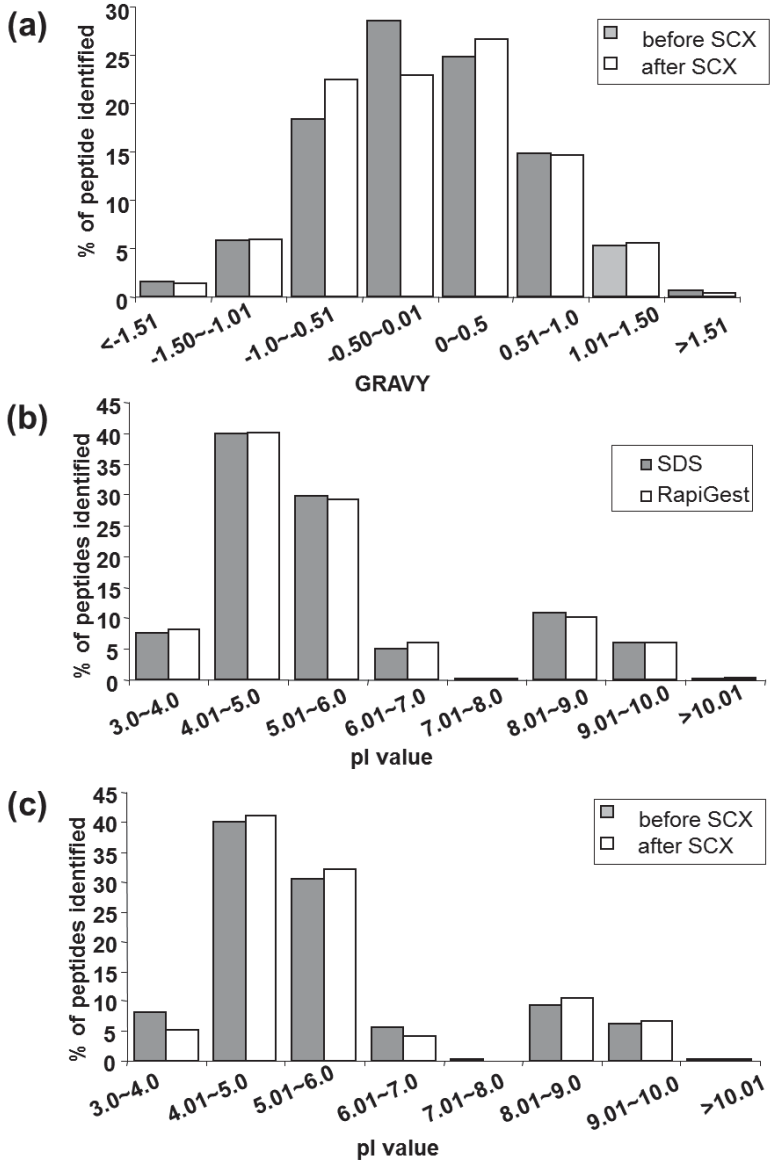


Figure 2-3. (a) Distributions of the GRAVY of peptides identified from the RapiGest sample before and after SCX. Distributions of the calculated pI of peptides identified from (b) the SDS and RapiGest samples and (c) the RapiGest samples before and after SCX.

identified from the RapiGest sample before and after the SCX cleaning. In this case, the proportions of peptides detected with GRAVY indexes in the range of -1.0 to +0.5 are different from the two samples. However, there is no clear trend of any preferential loss of peptides based on their hydrophobicity. This result indicates that the type of peptides lost during the SCX cleaning is not strongly correlated to the hydrophobicity of the peptides. It is likely that the ability of the peptides to be charged at a given pH of the solution for SCX plays a more important role. We have examined the distributions of the peptides detected from the SDS and RapiGest samples, as well as from the RapiGest samples before and after SCX cleaning, as a function of their calculated isoelectric points (pI values) (see Figure 2-3 (b,c)). As in the case of GRAVY index comparison shown in Figure 2-3 (a), there is no clear trend of preferential loss of peptides based on their pI values. Thus, it appears that what types of peptides are preferentially lost is difficult to predict during the SCX cleaning of the digests from the membrane protein fraction.

We have carried out a set of experiments using a simple non-membrane sample (i.e., BSA digest) to further examine the effects of SDS and RapiGest on the sample loss during the SCX cleaning step. The BSA digests were mixed with different amounts of RapiGest or SDS, followed by passing through the SCX column. During the washing step, the un-retained peptides were collected for the

RapiGest and non-surfactant samples. The retained peptides after elution using salts were collected. The collected samples were subjected to de-salting and quantification using HPLC-UV. For the BSA digest without containing any surfactant, an average of 64.6%±1.5% (from triplicates, n=3) of the peptide amount was recovered from the elutes and an additional 10.7%±0.4% (n=3) was recovered from the un-retained fraction. Thus, a total of 75.3% of the sample was recovered from the SCX sample cleanup. About 24.7% of the sample was lost. Sample loss took place likely in several steps, including sample injection, fraction collection, Speed-Vac concentration of the collected fractions and the de-salting process. Sample loss might also occur due to the strong retention of some peptides on the SCX column; we usually washed and re-equilibrated the column between sample runs. However, sample loss due to their reaction with the stationary phase in the SCX column is not likely, as we found that an SCX column could be used without performance degradation even after more than 500 injections.

Similar results were obtained for the BSA digest containing 0.1% RapiGest. In this case, 10.4%±0.4% were recovered from the un-retained fraction and 64.0%±0.7% were found in the elutes, with an average total recovery rate of 74.4%. However, for the BSA digest containing 0.025% SDS, only 30.7%±0.2% of the sample were recovered from the elutes, compared to 64.6% for the BSA digest with no surfactant and 64.0% for the BSA digest containing 0.1% RapiGest.

Furthermore, we found that the recovery rate increased as the SDS concentration was reduced (see Figure 2-4 (a)). For the digest containing 0.005% or 0.01%, the average recovery rate was 55.5% or 48.6%, respectively. Peptides bound with SDS are likely not retained on the column as they would be neutral or negatively charged. Unfortunately, the reversed-phase HPLC-UV device could not be used to recover and quantify the un-retained peptides from the SDS samples, due to strong interference of SDS. We note that the sample recovery rate of 64.0% from the BSA digest in 0.1% RapiGest is different from the 41% recovery found in the SCX cleaning of the *E. coli* digest containing 0.1% RapiGest. This difference is likely due to the difference in sample complexity. In the *E. coli* digest sample where many more different peptides are present compared to the BSA sample, it is possible that the RapiGest molecules strongly bound to some peptides might not be dissociated completely during the acidification step and thus the RapiGest-bound peptides would remain neutral or negatively charged, as in the case of SDS, resulting in the loss of these peptides during the SCX cleaning step.

Figure 2-4 (a) also shows the sequence coverage of BSA determined by LC MS/MS analysis of the elutes collected after the SCX cleaning of the digests

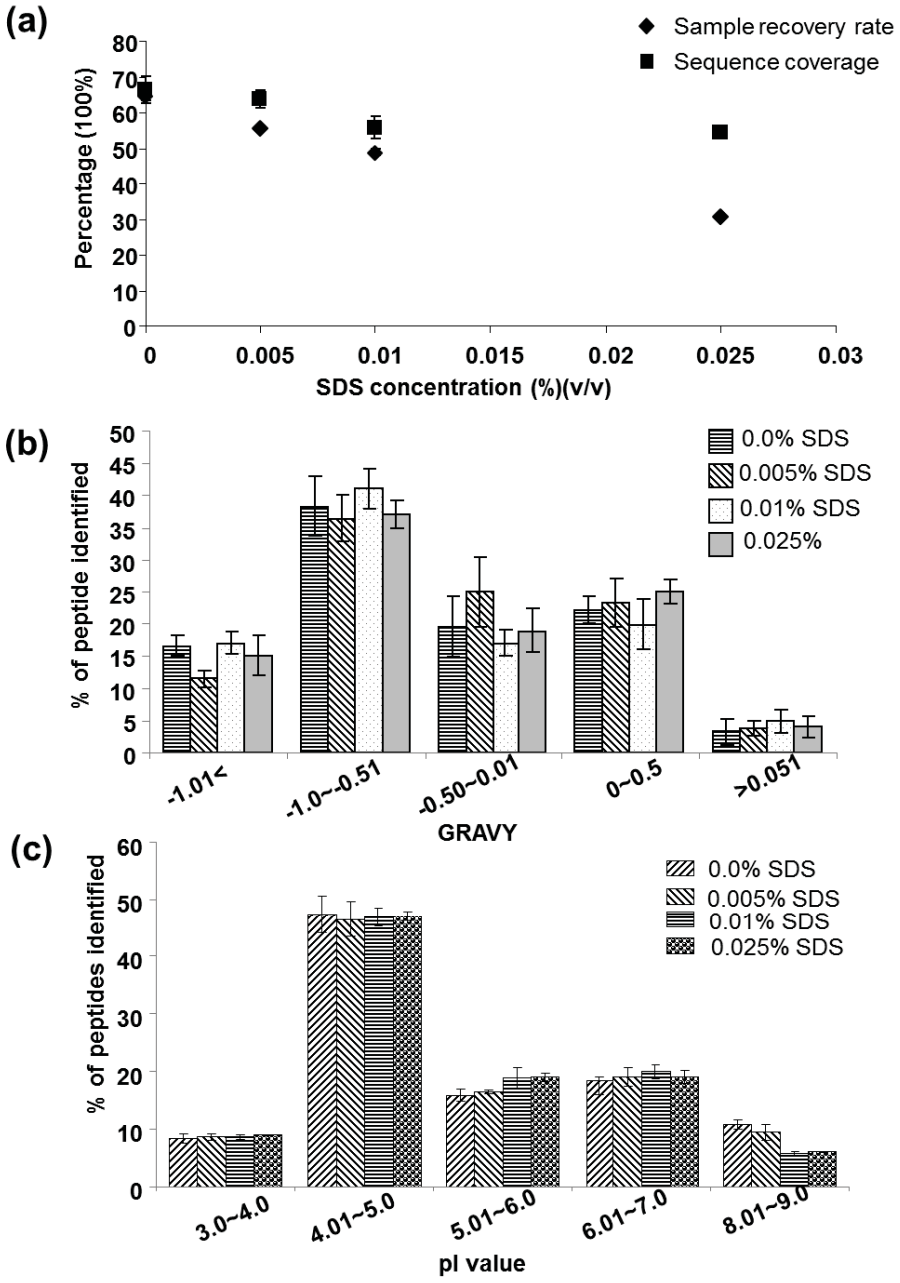


Figure 2-4. (a) Comparison of sample recovery rate and sequence coverage from the BSA digest samples containing different concentrations of SDS (n=3). Distributions of (b) the GRAVY and (c) the calculated pI of peptides identified from the BSA digests (n=3). Error bars represent one standard deviation.

containing different concentrations of SDS. The sequence coverage decreases from 66.3% ± 4.0% (42 ± 2 peptides) to 54.3% ± 1.5% (33 ± 1 peptides) when the SDS concentration increases from 0% to 0.025%. For the RapiGest sample, the sequence coverage is 66.0% ± 0.0% (41 ± 1) which is almost the same as that found in the BSA digest with no surfactant. To gauge what types of peptides are preferentially lost during the SCX cleaning of the SDS samples, we analyzed the GRAVY distribution (see Figure 2-4 (b)) and pI values (Figure 2-4 (c)) of the identified peptides. There are no significant changes in the GRAVY and pI distribution of the identified peptides. Like in the case of the *E. coli* digest, it is difficult to predict what peptides are preferentially lost during the SCX cleaning even for this simple digest from a non-membrane protein.

To investigate further the performance differences of the three sample preparation methods, we applied the RapiGest, PPS and SDS methods to analyze the proteome of the MCF7 membrane fraction. Table 2-2 summarizes the numbers of peptides and proteins identified from two replicate sample preparations of the MCF7 sample with RapiGest, PPS, or SDS. In each sample preparation, triplicate LC-ESI MS/MS runs were performed. An average of 1076±46 unique peptides and 399±16 unique proteins were identified from the MCF7 sample prepared with RapiGest, compared to 985±24 peptides and 368±6 proteins identified with PPS, and 864±14 peptides and 381±7 proteins with SDS. As in the case of the *E. coli*

Table 2-2. Summary of the numbers of peptides and proteins identified from two replicate sample preparations of the MCF7 proteome fraction using RapiGest, PPS or SDS.

Sample preparation method	# of peptides identified				# of proteins identified			
	Run# 1	Run# 2	Run# 3	Average	Run# 1	Run# 2	Run# 3	Average
RapiGest Prep#1	1073	1018	1018	1036±32	406	382	379	389±15
RapiGest Prep#2	1116	1110	1114	1113±3	412	396	422	410±13
PPS Prep#1	1015	985	951	983±32	371	360	363	364±6
PPS Prep#2	1003	997	961	987±23	368	372	377	373±5
SDS Prep#1	875	863	881	873±9	374	385	389	382±8
SDS Prep#2	865	863	838	855±15	385	385	373	381±7

sample analysis, the number of peptides and proteins identified from the RapiGest is higher than that obtained by the PPS or SDS method. In this case, the number of proteins identified from the SDS is slightly higher than the PPS method. After merging the protein identification results shown in Table 2-2, 2564 unique peptides and 827 unique proteins were identified from the MCF7 sample.

In summary, more peptides and proteins were identified from the RapiGest method, compared to the SDS or PPS method. The major cause of lower detectability in the SDS method appears to be related to sample loss during the SDS removal step by SCX; for the RapiGest sample, SCX-based surfactant removal is not needed. Moreover, the extent of sample loss was not uniform for all of the peptides; some peptides in the sample would be preferentially lost over the others, resulting in the change of the sample composition, not merely overall concentration reduction. Thus, increasing the starting materials to compensate for the sample loss during the SCX cleaning step will not result in more peptides and proteins identified. It was also found that there were no correlations between the peptide hydrophobicity gauged by GRAVY index or chargeability gauged by pI and the extent of peptide loss. Overall, the RapiGest method handles the membrane proteome sample better than the SDS method. However, RapiGest is a much more expensive reagent than SDS.

Finally, since the membrane proteome of *E. coli* is less well defined than the soluble proteins from cell extracts, we have applied the RapiGest method to generate a more comprehensive proteome map by using two-dimensional (2D) HPLC QTOF MS/MS. The membrane proteins from the *E. coli* cell extract were enriched as described in the Experimental section and then dissolved in RapiGest. After trypsin digestion, the digest was subjected to SCX fractionation, followed by reversed-phase LC-ESI MS/MS analysis. Two replicate experiments (i.e., the cells were divided into two samples for independent cell lysis, protein extraction, and other downstream processes) were carried out to gauge the reproducibility of the technique. As illustrated in Figure 2-5 (a), a total of 1345 unique proteins (4514 unique peptides) were identified from experiment #1, while a total of 1394 unique proteins (4948 unique peptides) were identified from experiment #2. The false discovery rate of the identified peptides was found to be 2.4%. There are 1112 common proteins detected, representing an average of 86.5% overlap between the numbers of proteins identified from the two experiments. After merging the protein identification results from the replicates, 5799 unique peptide corresponding to 1626 unique proteins were identified. About 67.5% of the 1100 identified proteins with known cellular locations are membrane proteins. Among them, about 75% are integral membrane proteins. We have analyzed these integral membrane proteins with respect to the number of transmembrane helices.

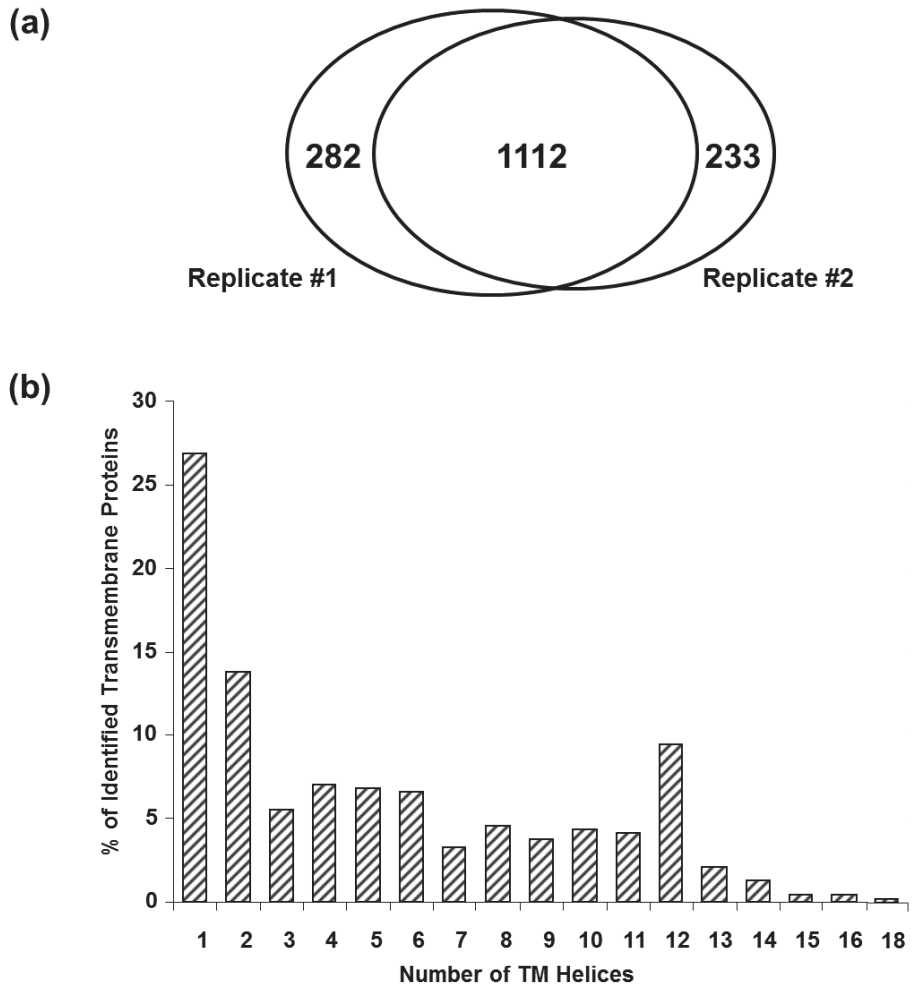


Figure 2-5. (a) Distribution of the numbers of proteins identified by 2D-LC MS/MS in two replicate experiments. (b) Distribution of the number of transmembrane domains of the known integral membrane proteins identified from 2D-LC MS/MS.

Figure 2-5 (b) shows a wide distribution of proteins with varying numbers of transmembrane domains. It appears that this method can be applied for analyzing a great variety of integral membrane proteins. To our knowledge, the number of proteins and membrane proteins identified in this work represents one of the most comprehensive profiles of the *E. coli* proteome.[49-52]

2.4 Conclusions

In this study, we have compared the applicability of SDS-, PPS- and RapiGest -assisted shotgun methods for membrane proteome analysis. Using the membrane fractions of *E. coli* and MCF7 cell extracts as models, we have demonstrated that the use of RapiGest allows identification of more peptides and proteins than PPS or SDS. Comparing the SDS and RapiGest methods, the RapiGest method does not require an additional step for processing the digest, while the SDS method requires the removal of SDS. We have shown that the SCX cleaning step during the sample preparation process resulted in peptide sample loss. For the *E. coli* membrane protein fraction, sample loss of as high as 62% was observed. Moreover, sample loss was not uniform for all peptides in the sample – some peptides were more readily lost than others. However, there was no correlation between the type of peptides preferentially lost and their hydrophobicity or calculated isoelectric point (pI). Thus, sample loss cannot be compensated by using more starting materials or selecting a certain type of peptides or proteins based on hydrophobicity or pI values for analysis. Overall, the RapiGest method is superior to the SDS and PPS methods in handling membrane proteome samples, while the performance of the SDS and PPS methods is similar in terms of the numbers of peptides and proteins identified. To illustrate the applicability of the RapiGest method for comprehensive

membrane proteome analysis, we have analyzed the *E. coli* membrane proteome using 2D-LC MS/MS. In total, 5799 unique peptides corresponding to 1626 unique proteins were identified with a false discovery rate of 2.4%.

2.5 Literature cited

1. Speers, A.; Wu, C., Proteomics of integral membrane proteins-theory and application. *Chem. Rev.* **2007,** *107* (8), 3687-3714.

2. Klein, C.; Garcia-Rizo, C.; Bisle, B.; Scheffer, B.; Zischka, H.; Pfeiffer, F.; Siedler, F.; Oesterhelt, D., The membrane proteome of Halobacterium salinarum. *Proteomics* **2005,** *5* (1), 180-197.

3. Santoni, V.; Molloy, M.; Rabilloud, T., Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **2000,** *21* (6), 1054-1070.

4. Wu, C. C.; Yates, J. R., The application of mass spectrometry to membrane proteomics. *Nat. Biotechnol.* **2003,** *21* (3), 262-267.

5. Adele R. Blackler, M. J. M., Christine C. Wu, Comparison of acid-labile and traditional detergents for membrane protein solubilization and shotgun proteomic analysis. *Proceedings of the 55th ASMS Conference on Mass Spectrometry* **2007**.

6. Ruan, Y.; Wan, M., An optimized procedure for solubilization, reduction, and transfer of human breast cancer membrane-enriched fraction by 2-DE. *Electrophoresis* **2007,** *28* (18), 3333-3340.

7. Aivaliotis, M.; Karas, M.; Tsiotis, G., An alternative strategy for the membrane proteome analysis of the green sulfur bacterium Chlorobium tepidum using blue native PAGE and 2-D PAGE on purified membranes. *J. Proteome Res.* **2007,** *6* (3), 1048-1058.

8. Zuobi-Hasona, K.; Crowley, P.; Hasona, A.; Bleiweis, A.; Brady, L., Solubilization of cellular membrane proteins from Streptococcus mutans for

two-dimensional gel electrophoresis. *Electrophoresis* **2005,** *26* (6), 1200-1205.

9. Zischka, H.; Gloeckner, C.; Klein, C.; Willmann, S.; Lange, M.; Ueffing, M., Improved mass spectrometric identification of gel-separated hydrophobic membrane proteins after sodium dodecyl sulfate removal by ion-pair extraction. *Proteomics* **2004,** *4* (12), 3776-3782.

10. Rabilloud, T.; Chevallet, M.; Luche, S.; Lelong, C., Fully denaturing two-dimensional electrophoresis of membrane proteins: A critical update. *Proteomics* **2008,** *8* (19), 3965-3973.

11. Poetsch, A.; Wdlters, D., Bacterial membrane proteomics. *Proteomics* **2008,** *8* (19), 4100-4122.

12. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009,** *11*, 49-79.

13. Chen, E. I.; Hewel, J.; Felding-Habermann, B.; Yates, J. R., Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). *Mol. Cell. Proteomics* **2006,** *5* (1), 53-56.

14. Villen, J.; Beausoleil, S. A.; Gerber, S. A.; Gygi, S. P., Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (5), 1488-1493.

15. Wu, C. C.; MacCoss, M. J., Shotgun proteomics: Tools for the analysis of complex biological systems. *Curr. Opin. Mol. Ther.* **2002,** *4* (3), 242-250.

16. Lu, X. N.; Zhu, H. N., Tube-gel digestion - A novel proteomic approach for

high throughput analysis of membrane proteins. *Mol. Cell. Proteomics* **2005,** *4* (12), 1948-1958.

17. Linke, D., Detergents: an overview. *Methods Enzymol.* **2009,** *463*, 603-17.

18. Wang, N.; MacKenzie, L.; De Souza, A. G.; Zhong, H. Y.; Goss, G.; Li, L., Proteome profile of cytosolic component of zebrafish liver generated by LC-ESI MS/MS combined with trypsin digestion and microwave-assisted acid hydrolysis. *J. Proteome Res.* **2007,** *6* (1), 263-272.

19. Zhang, N.; Chen, R.; Young, N.; Wishart, D.; Winter, P.; Weiner, J. H.; Li, L., Comparison of SDS- and methanol-assisted protein solubilization and digestion methods for Escherichia coli membrane proteome analysis by 2-D LC-MS/MS. *Proteomics* **2007,** *7* (4), 484-493.

20. Zhang, N.; Li, L., Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid Commun. Mass Spectrom.* **2004,** *18* (8), 889-896.

21. Zhang, N.; Shaw, A. R. E.; Li, N.; Chen, R.; Mak, A.; Hu, X. Y.; Young, N.; Wishart, D.; Li, L., Liquid chromatography electrospray ionization and matrix-assisted laser desorption ionization tandem mass spectrometry for the analysis of lipid raft proteome of monocytes. *Anal. Chim. Acta* **2008,** *627* (1), 82-90.

22. Hixson, K. K.; Rodriguez, N.; Camp, D. G.; Strittmatter, E. F.; Lipton, M. S.; Smith, R. D., Evaluation of enzymatic digestion and liquid chromatography-mass spectrometry peptide mapping of the integral membrane protein bacteriorhodopsin.

Electrophoresis **2002,** *23* (18), 3224-3232.

23. Weber, K.; Kuter, D. J., Reversible denaturation of enzymes by sodium dodecyl sulfate. *J. Biol. Chem.* **1971,** *246* (14), 4504-9.

24. Allen, T. M.; Romans, A. Y.; Kercret, H.; Segrest, J. P., Detergent removal during membrane reconstitution. *Biochim. Biophys. Acta* **1980,** *601* (2), 328-42.

25. Nagaraj, N.; Lu, A. P.; Mann, M.; Wisniewski, J. R., Detergent-based but gel-free method allows identification of several hundred membrane proteins in single LC-MS runs. *J. Proteome Res.* **2008,** *7* (11), 5028-5032.

26. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* **2009,** *6* (5), 359-62.

27. Manza, L. L.; Stamer, S. L.; Ham, A. J.; Codreanu, S. G.; Liebler, D. C., Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **2005,** *5* (7), 1742-5.

28. Wisniewski, J. R.; Mann, M., Spin filter-based sample preparation for shotgun proteomics Reply. *Nat. Methods* **2009,** *6* (11), 785-786.

29. Timperman, A. T.; Aebersold, R., Peptide electroextraction for direct coupling of in-gel digests with capillary LC-MS/MS for protein identification and sequencing. *Anal. Chem.* **2000,** *72* (17), 4115-4121.

30. Antharavally, B. S.; Salunkhe, A. M.; Mallia, K. A.; Rosenblatt, M. M.; Rogers, J. C.; Haghdoost, N. R.; Haney, P. J.; Benton, B.; Bell, P. A., A rapid method for removal of detergents from proteins and protein digests in a spin column format. *FASEB J.* **2010,** *24* (1), 519.1.

31. Imre, T.; Schlosser, G.; Pocsfalvi, G.; Siciliano, R.; Molnar-Szollosi, E.; Kremmer, T.; Malorni, A.; Vekey, K., Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography-electrospray mass spectrometry. *J. Mass Spectrom.* **2005,** *40* (11), 1472-1483.

32. Yu, Y. Q.; Gilar, M.; Gebler, J. C., A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin. *Rapid Commun. Mass Spectrom.* **2004,** *18* (6), 711-715.

33. Stover, T.; Amari, J. V.; Mazsaroff, I.; Gilar, M.; Yu, Y. Q.; Gebler, J. C., Novel characterization tool for Mab digestion-Technical note: RapiGest SF denaturant tool for improved trypsin digestion of monoclonal antibodies. *Genetic Engineering News* **2003,** *23* (17), 48-49.

34. Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C., Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **2003,** *75* (21), 6023-6028.

35. Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., III, Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **2007,** *6* (7), 2529-2538.

36. Arnold, R. J.; Hrncirova, P.; Annaiah, K.; Novotny, M. V., Fast proteolytic digestion coupled with organelle enrichment for proteomic analysis of rat liver. *J. Proteome Res.* **2004,** *3* (3), 653-657.

37. Nomura, E.; Katsuta, K.; Ueda, T.; Toriyama, M.; Mori, T.; Inagaki, N., Acid-labile surfactant improves in-sodium dodecyl sulfate polyacrylamide gel protein digestion for matrix-assisted laser desorption/ionization mass

spectrometric peptide mapping. *J. Mass Spectrom.* **2004,** *39* (2), 202-207.

38. Konig, S.; Schmidt, O.; Rose, K.; Thanos, S.; Besselmann, M.; Zeller, M., Sodium dodecyl sulfate versus acid-labile surfactant gel electrophoresis: Comparative proteomic studies on rat retina and mouse brain. *Electrophoresis* **2003,** *24* (4), 751-756.

39. Meng, F.; Cargile, B.; Patrie, S.; Johnson, J.; McLoughlin, S.; Kelleher, N., Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal. Chem.* **2002,** *74* (13), 2923-2929.

40. Ruth, M.; Old, W.; Emrick, M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K.; Mendoza, A.; Sevinsky, J.; Hamady, M.; Knight, R.; Resing, K.; Ahn, N., Analysis of membrane proteins from human chronic myelogenous leukemia cells: Comparison of extraction methods for multidimensional LC-MS/MS. *J. Proteome Res.* **2006,** *5* (3), 709-719.

41. Norris, J.; Porter, N.; Caprioli, R., Mass spectrometry of intracellular and membrane proteins using cleavable detergents. *Anal. Chem.* **2003,** *75* (23), 6642-6647.

42. Ross, A.; Lee, P.; Smith, D.; Langridge, J.; Whetton, A.; Gaskell, S., Identification of proteins from two-dimensional polyacrylamide gels using a novel acid-labile surfactant. *Proteomics* **2002,** *2* (7), 928-936.

43. Suder, P.; Bierczynska, A.; Konig, S.; Silberring, J., Acid-labile surfactant assists insolution digestion of proteins resistant to enzymatic attack. *Rapid Commun. Mass Spectrom.* **2004,** *18* (7), 822-824.

44. Klammer, A.; MacCoss, M., Effects of modified digestion schemes on the

identification of proteins from complex mixtures. *J. Proteome Res.* **2006,** *5* (3), 695-700.

45. Chen, E. I.; McClatchy, D.; Park, S. K.; Yates, J. R., Comparisons of mass spectrometry compatible surfactants for global analysis of the mammalian brain proteome. *Anal. Chem.* **2008,** *80* (22), 8694-8701.

46. Mbeunkui, F.; Goshe, M. B., Investigation of solubilization and digestion methods for microsomal membrane proteome analysis using data-independent LC-MSE. *Proteomics* **2011,** *11* (5), 898-911.

47. Elias, J.; Haas, W.; Faherty, B.; Gygi, S., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2005,** *2* (9), 667-675.

48. Gong, Y.; Wang, N.; Wu, F.; Cass, C.; Damaraju, S.; Mackey, J.; Li, L., Proteome profile of human breast cancer tissue generated by LC-ESI-MS/MS combined with sequential protein precipitation and solubilization. *J. Proteome Res.* **2008,** *7* (8), 3583-3590.

49. Masuda, T.; Saito, N.; Tomita, M.; Ishihama, Y., Unbiased quantitation of Escherichia coli membrane proteome using phase transfer surfactants. *Mol. Cell. Proteomics* **2009,** *8* (12), 2770-2777.

50. Weiner, J. H.; Li, L., Proteome of the Escherichia coli envelope and technological challenges in membrane proteome analysis. *Biochimica Et Biophysica Acta-Biomembranes* **2008,** *1778* (9), 1698-1713.

51. Iwasaki, M.; Miwa, S.; Ikegami, T.; Tomita, M.; Tanaka, N.; Ishihama, Y., One-dimensional capillary liquid chromatographic separation coupled with

tandem mass spectrometry unveils the Escherichia coil proteome on a microarray scale. *Anal. Chem.* **2010,** *82* (7), 2616-2620.

52. Diaz-Mejia, J. J.; Babu, M.; Emili, A., Computational and experimental approaches to chart the Escherichia coli cell-envelope-associated proteome and interactome. *FEMS Microbiol. Rev.* **2009,** *33* (1), 66-97.

Chapter 3 Characterization of Differentially Expressed Plasma Membrane Proteins Induced by NPM-ALK for Investigating Aberrant Signaling Pathways in ALK+ Anaplastic Large Cell Lymphoma

3.1. Introduction

Anaplastic lymphoma kinase-positive anaplastic large cell lymphoma (ALK+ ALCL) is a distinct subtype of non-Hodgkin's lymphoma of T/null-cell lineage characterized by the consistent expression of CD30 and anaplastic cytologic features.[1] Approximately 80% of ALK+ ALCL tumors possess the t(2;5)(p23:q35) translocation, which places the ALK gene under the regulation of the nucleophosmin (NPM) gene promoter. The resultant tyrosine kinase oncoprotein, NPM-ALK, has been reported to dramatically change the expression level of multiple cell membrane marks and induce the development of ALK+ Anaplastic Large Cell Lymphoma (ALCL).[2-4]

Plasma membrane proteins play important roles in regulating various cellular functions such as signal transduction, material and energy exchanges between cells and the environment.[5] Previous studies have shown that, by regulating the expression level of these plasma membrane proteins, oncoproteins are able to regulate crucial cellular signal pathways which contribute to the malignant transformation.[6] Since the plasma membrane proteins form key nodes in the communication between the extracellular and intracellular space, they are, in general, very popular drug targets.[7]

For the analysis of plasma membrane proteins, the most challenging issue is how to maintain the solubility of hydrophobic proteins during the whole sample preparation process including protein isolation and digestion. Several conventional solvents, including surfactants, chaotropic agents (e.g., urea), aqueous-organic solvents (e.g., methanol-water), and organic acids (e.g., formic acid) have been demonstrated to be helpful in membrane protein handling. Surfactants aid in dissolving proteins, and among them, sodium dodecyl sulfate (SDS), an ionic detergent with strong solubilization capability, can be used to dissolve and denature a wide range of proteins, including membrane proteins.[8] However, the resultant peptides containing SDS cannot be analyzed directly by reversed-phase (RP) LC-MS, as SDS often causes interference with RPLC separation[9,10] and MS analysis.[11,12] Therefore, efficient removal of SDS is required before RPLC tandem MS (MS/MS) analysis. This step could cause huge sample loss and compromise the final result. Alternatively, SDS-mimic surfactants, such as acid labile surfactants (e.g., RapiGest from Waters, PPS from Protein Discovery), may be used. These reagents rapidly decompose into two easily removed products upon addition of an acid, such as trifluoroacetic acid (TFA) to the digested protein sample. Therefore it will not cause interference in LC-MS/MS analysis. The use of RapiGest allowed identification of more peptides and proteins than the use of SDS.[13]

In this work, we developed a high throughput plasma membrane protein purification, identification, and quantitation strategy. RapiGest was applied for plasma membrane solubilization. The plasma membrane proteins of two NPM-ALK-expressing cell lines, Karpas 299 and SUPM2, were identified using 2D LC-MS/MS analysis. In addition, 2-MEGA labeling method combined with 2D LC-MS/MS was used for plasma membrane protein quantitation analysis in NPM-ALK expressing HEK 293 cells and NPM-ALK absent HEK 293 cells (control). The identified NPM-ALK-induced differentially expressed plasma membrane proteins should assist in furthering our understanding of NPM-ALK-induced tumorigenesis.

3.2. Experimental section

3.2.1. Chemicals and reagents

Dithiothreitol (DTT) was from Bio-Rad Laboratories (Mississauga, ON, Canada). Sequencing grade modified trypsin, LC-MS grade water, acetone, and HPLC grade acetonitrile (ACN) were from Fisher Scientific Canada (Edmonton, AB, Canada). Phosphoric acid (H_3PO_4) and potassium chloride (KCl) were purchased from Anachemia (Edmonton, AB, Canada). RapiGest^{TM} SF was purchased from Waters (Milford, MA). The BCA assay kit was from Pierce (Rockford, IL). The plasma membrane protein extraction kit was from BioVision (Milpitas, California). The isotope reagent, d(6), ^{13}CD-formaldehyde (20%, w/w in deuterated H_2O), was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). All the other chemicals were purchased from Sigma-Aldrich Canada (Markham, ON, Canada).

3.2.2. Plasma membrane protein preparation

All the cell lines were kindly provided by Dr. Raymond Lai's lab from University of Alberta. Plasma membrane protein extraction kit (Catalog #K268-50) was used for plasma membrane protein preparation from Karpas 299 and SUPM2 cells. The same method was applied to NPM-ALK expressing HEK 293 cells and

NPM-ALK absent HEK 293 cells. Cells were collected by centrifugation at 600 g for 5 min at 4°C and washed once with ice cold PBS. Cells were resuspended in the Homogenize Buffer Mix in an ice-cold Dounce homogenizer and homogenized on ice for 30-50 times. The homogenate was transferred to a 1.5 mL microcentrifuge tube and centrifuged in 700 g for 10 min at 4°C. The supernatant was collected. The supernatant was transferred to a new vial and centrifuged at 10,000 g for 30 min at 4°C. The resultant pellet was the total cellular membrane protein. The total membrane proteins were resuspended in 200 μL of the Upper Phase Solution. The same volume of the Lower Phase Solution was added. The solution was mixed well and incubated on ice for 5 min. The tube was centrifuged in a microcentrifuge at 3500 rpm (1000 g) for 5 min. The upper phase was carefully collected and diluted in five volume of water. The solution was kept on ice for 5 min. Then the solution was spun at top speed at a microcentrifuge tube for 10 min at 4°C. The plasma membrane protein pellet was collected and stored at -80°C.

3.2.3. Protein solubilization and digestion

The plasma membrane protein pellet from each cell line was solubilized individually in 1% RapiGest solution. RapiGest-assisted solubilization and digestion were carried out as previously described with some modifications.[13] In

brief, 1% (w/v) RapiGest was used to re-suspend plasma membrane proteins. Standard reduction and alkylation were carried out on the proteins (300 μg). Reduction for 1 h at 37 °C was done by the addition of 6 μL 900 mM DTT, followed by alkylation for 1 h at room temperature in the dark by the addition of 12 μL 900 mM IAA. Trypsin solution was then added to the protein solution at a ratio of 1 to 50 after 10-fold dilution. The mixture was then incubated at 37 °C overnight. 10% TFA was added to the peptide mixture to make the final TFA concentration approximately 0.5% (pH<2). The mixture was incubated at 37 °C for 45 min and then centrifuged at 18,000 g for 10 min. The supernatant was carefully transferred into another vial and was stored in -80 °C for future use.

3.2.4. 2-MEGA isotopic labeling

Peptides were isotopic labeled with the 2-MEGA method as described previously[14,15] with some modifications. In brief, the protein digest samples were adjusted to pH 11 using 2 M NaOH and 2 M *O*-methylisourea was added. Samples were incubated at 60 °C for 20 min with intermittent shaking to guanidinate the ε-amino groups of the lysines. The pH of the solution was adjusted with 50% TFA to approximately pH 7. Acetate buffer was added to adjust the protein digest solution to pH 4. Formaldehyde (4%, v/v; $^{12}CH_2O$ for light chain labeling and $^{13}CD_2O$ for heavy chain labeling) and sodium

cyanoborohydride (1 M) were added to dimethylate the N-termini of the peptides.

3.2.5. Desalting

The sample desalting step was performed by RPLC with an Agilent 1100 HPLC system fitted with a 4.6 × 50 mm C18 column (Part #: 2001-050×046, Varian, Ontario, Canada). After sample loading, the column was first flushed with 97.5% mobile phase A (0.1% TFA/H_2O) and 2.5% mobile phase B (0.1% TFA/ACN) for 5 min to remove the salts. Then, 85% mobile phase B was used to completely elute the peptides. The collected samples were dried down and reconstituted in 0.1% FA/H_2O. The amount of peptides was determined by the UV absorbance at 214 nm.[16]

3.2.6. SCX in 2D-LC MS/MS

In 2D-LC MS/MS, the protein digest was first separated by SCX on the Agilent 1100 HPLC system. The protein digest was reconstituted in 0.2% H_3PO_4/H_2O (pH 2). Gradient elution was performed by using mobile phase A (10 mM KH_2PO_4 with 20% ACN, pH 2.67) and mobile phase B (500 mM KCl and 10 mM KH_2PO_4 with 20% ACN, pH 2.67) at a flow rate of 0.2 mL/min. The gradient profile was as follows: 0 min: 0% B, 5 min: 0% B, 5.1 min: 4% B, 50 min: 70% B, 52 min: 100% B, 57 min: 100% B, 58 min: 0% B and 70 min: 0% B. In total, 34

fractions were collected and directly desalted and quantified by RPLC-UV, as described above. The adjacent low concentration fractions were combined to produce a sufficient amount of peptide (i.e., >1.5 μg); 20 fractions were generated for LC-MS.

3.2.7. LC-MS Analysis

The desalted peptide mixtures were analyzed by using a quadrupole time-of-flight (QTOF) Premier mass spectrometer (Waters, Milford, MA) equipped with a nanoACQUITY Ultra Performance LC system (Waters). In brief, a 1.5 μg sample was injected each time onto a 75 μm × 150 mm Atlantis C18 column (Part #: 186003500, Waters, Milford, MA). Mobile phase A consisted of 0.1% FA/H_2O and mobile phase B consisted of 0.1% FA/ACN. The peptide sample was separated using a 130-min gradient with the following profile: 0 min: 2% B, 2 min: 7% B, 85 min: 20% B, 110 min: 30% B, 115 min: 45% B, 120 min: 90% B, 125 min:90% B, and 130 min: 2% B. All samples were electrosprayed into the mass spectrometer at a flow rate of 350 nL/min. MS data were recorded within an m/z window of 350-1600 for 0.8 s, followed by 4 data-dependent MS/MS scans of the four most intense ions within an m/z window of 50-1990 for 0.8 s. For dynamic mass exclusion, a time window of 180 s and a mass tolerance window of 50 mDa were applied. Various collision energies were used based on

the mass and charge state of the peptide. Mass scans of a mixture of mass calibrants consisting of leucine enkephalin and (Glu1)-fibrinopeptide B were acquired at 1 scan/min throughout the run.

For 2D-LC MS/MS, the precursor ion exclusion (PIE) method[17] was applied to enhance the peptide and protein identification. In brief, an exclusion list was generated based on the peptides identified from the Mascot search program (Matrix Science, London, U.K.). The m/z value, charge state, and retention time of each identified peptide were extracted from the database search results and the corresponding raw data. The m/z value of the other charge state for each identified precursor ion was calculated. In addition, the values of all identified peptides consisted of, not only the monoisotope value, but also the three additional isotope values. Finally, all the m/z values along with their retention time information were loaded into the MS method for the new LC-MS run.

3.2.8. Protein database search and bioinformatics

Database searches for plasma membrane protein identification were performed as previously described with minor changes.[18] Raw search data were lockmass-corrected, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.3 (Waters). Peptide sequences were identified by automated database searching of peak list files using the Mascot search program. Database

searching was restricted to *Homo sapiens* (human) in Swissprot (version 4 and accession date of May 7, 2009). The following search parameters were selected for all database searching: enzyme, trypsin; missed cleavages, 1; peptide tolerance, 30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, 1+, 2+, and 3+; fixed modification, carbamidomethyl (C); variable modifications, oxidation (M). All the identified peptides with scores lower than the Mascot threshold score for identity at the confidence level of 95% were then removed from the protein list. The redundant peptides for different protein identities were deleted, and the redundant proteins identified under the same gene name but different access ID numbers were also removed from the list. Finally, a unique protein or peptide list was generated by merging all the protein and peptide lists from individual runs. Unique proteins (under unique gene names) and peptides with the highest scores were kept. The false positive peptide matching rate was gauged by using the target-decoy search strategy by searching the MS/MS spectra against the forward and reversed proteome sequences.

For the plasma membrane protein quantitation analysis in NPM-ALK expressing HEK 293 cells and NPM-ALK absent HEK 293 cells (control), MS and MS/MS data were processed and searched using Mascot Distiller. The peptide peak pairs from MS spectra were picked and the relative intensity ratios were calculated. The MS/MS spectra of the peptide peak pairs were searched with the

following parameters: database, Swissprot; taxonomy, *Homo sapiens* (human); enzyme, trypsin; missed cleavage, 1; fixed modification, carbamidomethyl (C); variable modification, Guanidinyl (K), Dimethylation d(0) (N-term, $+C_2H_4$), Dimethylation d(6) (N-term, $+^{13}C_2D_4$); MS tolerance, 30 ppm; MS/MS tolerance, 0.2 Da.

The cell location of the identified and quantified proteins was determined through the use of Ingenuity Pathway Analysis software (Ingenuity Systems, Redwood, CA).

3.3. Results and discussion

3.3.1. Plasma membrane protein profiling

Figure 3-1 shows the whole workflow used for the plasma membrane protein profiling and quantification analysis. For the plasma membrane protein profiling, two ALK+ ALCL cell lines (Karpas 299 and SUPM2) were investigated as biological duplicate experiments. The plasma membrane protein was extracted from the cells, followed by RapiGest-assisted protein solubilization and digestion. 2D LC-MS/MS was used to analyze the resultant peptide mixture. As shown in Table 3-1, the number of unique peptides and proteins identified from the Karpas 299 cells were 10304 and 3100, respectively. By using Ingenuity Pathway Analysis software for the cellular location analysis, 561 unique proteins were demonstrated to be plasma membrane proteins or extracellular proteins attached to the plasma membrane. Therefore, the enrichment efficiency was 18%. For the other ALK+ ALCL cell line, SUPM2, the number of unique peptides and proteins identified were 11318 and 2840, respectively. 552 unique proteins were identified as plasma membrane proteins or extracellular proteins based on cellular location analysis. The plasma membrane protein enrichment efficiency was 19%. In total, 726 unique plasma membrane proteins and extracellular proteins were

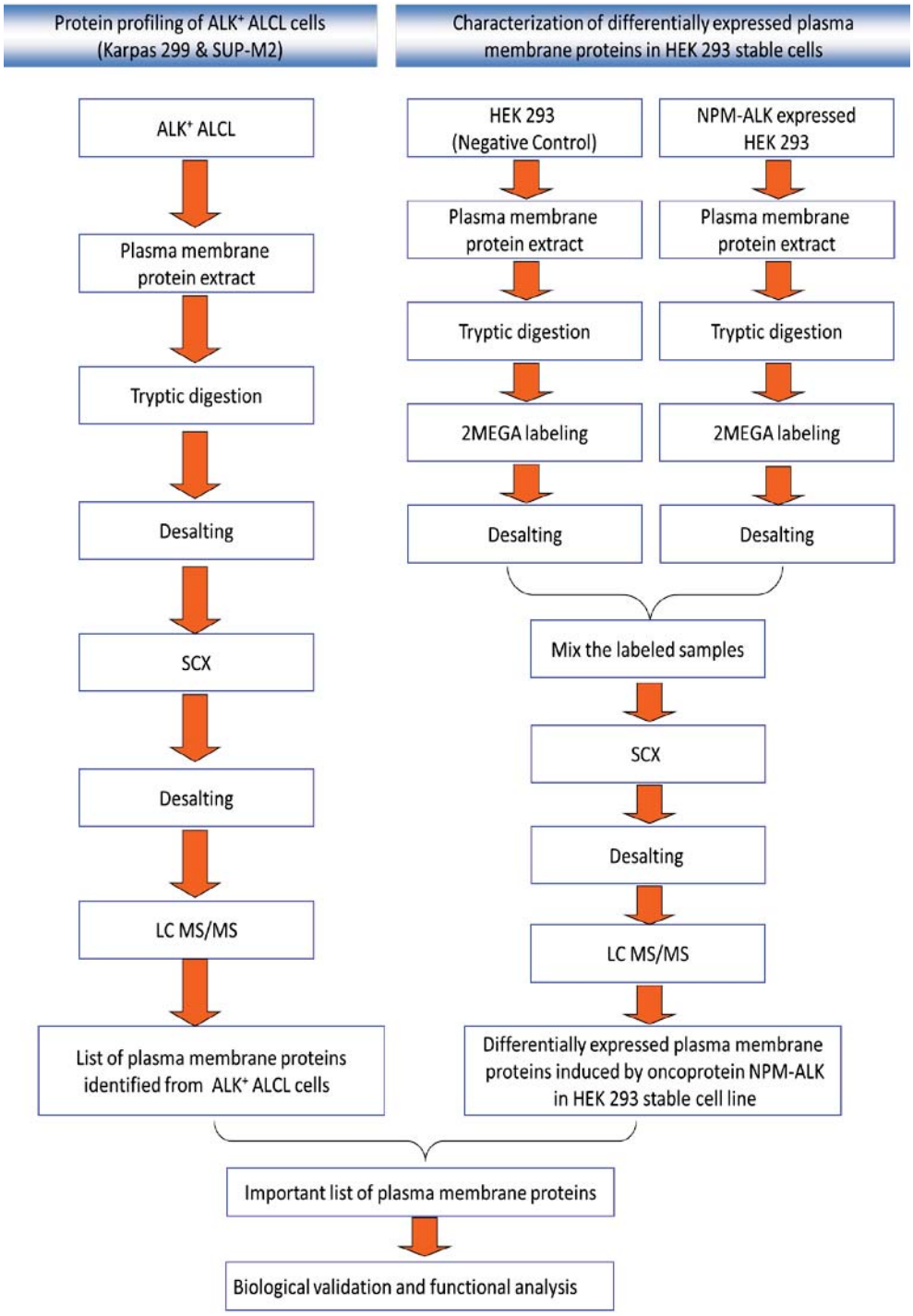


Figure 3-1. Workflow of the plasma membrane protein profiling and the 2-MEGA labeling quantitative MS analysis.

Table 3-1. Summary of the number of unique peptides, unique proteins, unique plasma membrane proteins and extracellular proteins identified and enrichment efficiency of plasma membrane proteins from biological duplicate 2D LC-MS/MS analysis of ALK+ ALCL cells (Karpas 299 and SUPM2).

Cell lines	**Number of unique peptides**	**Number of unique proteins**	**Number of unique plasma membrane proteins and extracellular proteins**	**Enrichment efficiency of plasma membrane proteins and extracellular proteins**
Karpas 299	10304	3100	561	18%
SUPM2	11318	2840	552	19%

identified from the biological duplicate samples. 387 plasma membrane proteins and extracellular proteins were identified in both the Karpas 299 and SUPM2 cells. The number of unique plasma membrane proteins and extracellular proteins identified only from one type of ALK+ ALCL cells was 174 for Karpas 299 and 165 for SUPM2. This result demonstrated the successful enrichment of the plasma membrane proteins from the ALK+ ALCL cells. The identified plasma membrane proteins and extracellular proteins will be overlapped with the following quantification result to generate the important differentially expressed plasma membrane proteins and extracellular proteins list.

3.3.2. Functional expression of NPM-ALK in stable cell line

The pTRE-TIGHT vector carrying NPM-ALK cDNA was introduced into HEK 293 stable cells via stable transfection. The resulting Tet-on HEK 293 advanced cells carrying pTRE-TIGHT/NPM-ALK were maintained in Dulbecco's modified Eagle's medium supplemented with 10% Tet-System Approved FBS, 100 μg/mL of G418 and 50 μg/mL of hyrogomycin B. The functional expression of oncoprotein NPM-ALK is shown in Figure 3-2. In order to mimic the functional expression level of NPM-ALK in the ALK+ ALCL cells, different concentrations of the doxycycline (DOX) dose (0 ng/mL, 50 ng/mL and 100 ng/mL) were added to the stable cells. The result showed that the 50 ng/mL DOX

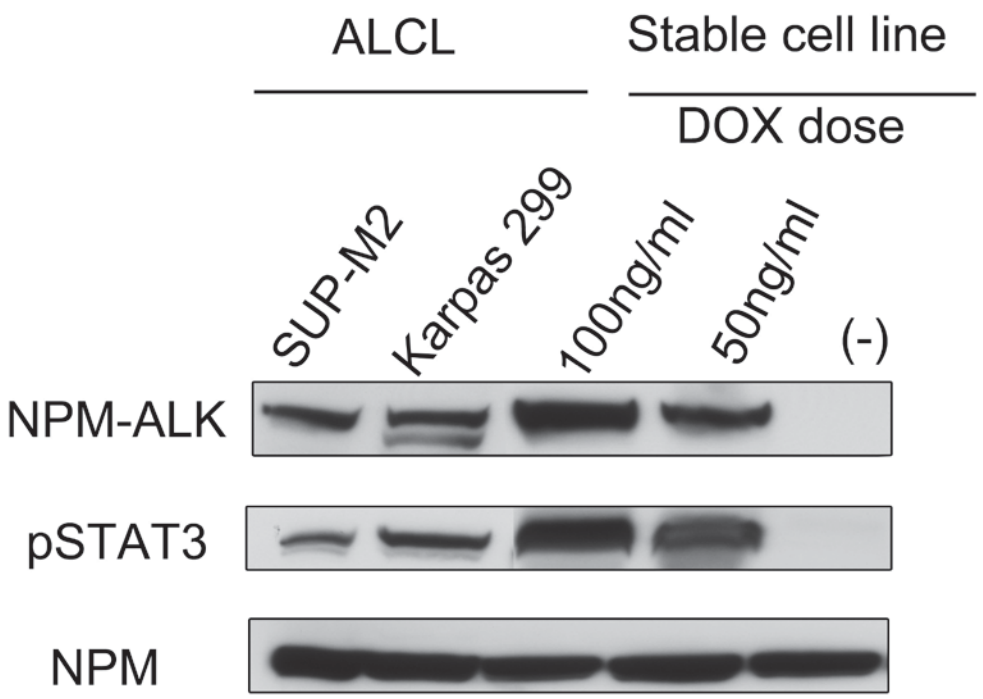


Figure 3-2. Western blot for comparison of functional expression of oncoprotein NPM-ALK in stable cell line and ALCL cell lines (Karpas299 & SUP-M2).

dose could provide similar NPM-ALK expression level to the ALK+ ALCL cells (both Karpas 299 and SUPM2). Therefore, these NPM-ALK expressing HEK 293 stable cells and NPM-ALK absent HEK 293 cells (control) were used for quantification analysis of the differentially expressed plasma membrane proteins induced by NPM-ALK in ALK+ ALCL cells.

3.3.3. Plasma membrane protein quantification

2-MEGA isotopic labeling method was used for quantification analysis of differentially expressed plasma membrane proteins induced by NPM-ALK in ALK+ ALCL cells. Figure 3-1 shows the workflow using the forward (A_HB_L) and reverse (A_LB_H) labeling strategy for MS based quantitative analysis. NPM-ALK absent HEK 293 stable cells (A, control) and NPM-ALK expressing HEK 293 stable cells (B) were separately lysed and the plasma membrane proteins of each type of the cells were extracted by using the plasma membrane protein extraction kit. The enriched plasma membrane proteins were reduced, alkylated, and digested with trypsin. The resulting peptide mixture of both samples were divided into two equal parts and chemically labeled with heavy chain (H) and light chain (L) isotopic labels. All of the labeled peptide mixtures were quantified and desalted to facilitate accurate 1:1 mixing. The heavy chain labeled NPM-ALK absent sample was mixed with the light chain labeled NPM-ALK expressing

sample to form a forward mixture of A_HB_L and the reversed labeling produced the A_LB_H mixture. The two mixtures were fractionated by using SCX and followed by LC-MS/MS analysis.

Mascot distiller was the software used for peptide identification and quantification information extraction. The peak pair of one peptide in a MS spectrum with a mass difference of 6.032 Da could be identified if both A and B mixtures contain this peptide. Pairs with a relative abundance difference of greater than 1.50 or less than 0.67 were the differentially expressed ones. The individual quantified peptides from the forward mixture of A_HB_L were matched with those from the reverse mixture of A_LB_H to find common peptides. The corresponding protein list was generated and the relative abundance ratio was calculated.

Since the 2-MEGA isotopic labeling strategy analyzes both the sample mixture A_HB_L and A_LB_H, there would be two individual quantified ratios for each peptide. These were treated as duplicate quantification experiments. In total, 793 unique peptides and 114 corresponding unique plasma membrane proteins and extracellular proteins were identified and quantified in both the forward and reverse labelled samples. By applying 1.50/0.67 as the threshold for differential expression, 48 unique plasma membrane proteins and extracellular proteins were found to be differentially expressed between NPM-ALK expressing HEK 293

cells and NPM-ALK absent HEK 293 cells. Among these proteins, 39 were up-regulated and 9 were down-regulated.

To find the most important differentially expressed plasma membrane proteins and extracellular proteins, the comparison of the identified plasma membrane proteins and extracellular proteins in two ALK+ ALCL cell lines (Karpas 299 and SUPM2) and the quantified differentially expressed plasma membrane proteins and extracellular proteins in HEK 293 stable cell line was performed (Figure 3-3). Among the 39 up-regulated proteins, 22 were both identified from the Karpas 299 cells and SUPM2 cells. Two proteins were only identified from the Karpas 299 cells and one protein was identified only from the SUPM2 cells. The other 14 proteins were not identified from the two ALK+ ALCL cell lines. For the nine down-regulated proteins, four were identified from both cell lines. One protein was only identified from Karpas 299 cells and one protein was only identified from the SUPM2 cells. The other three proteins were not identified from the two ALK+ ALCL cell lines. The final important list of plasma membrane proteins and extracellular proteins identified from ALK+ ALCL cells (Karpas 299 or SUPM2) and quantified to be differentially expressed between NPM-ALK expressing HEK 293 stable cells and NPM-ALK absent HEK 293 stable cells (control) was shown in Table 3-2, including 25 up-regulated and six down-regulated proteins.

Figure 3-3. Comparison of the identified plasma membrane proteins and extracellular proteins in two ALK+ ALCL cell lines (Karpas 299 and SUPM2) and the quantified differentially expressed plasma membrane proteins in the HEK 293 stable cell line.

Table 3-2. List of plasma membrane proteins and extracellular proteins identified from ALK+ ALCL cells (Karpas 299 or SUPM2) and quantified to be differentially expressed between NPM-ALK expressing HEK 293 stable cells and NPM-ALK absent HEK 293 stable cells (control).

Swiss-Prot ID	Name	Protein name	(# of Peptides, Ratio)
Up-regulated proteins			
P11171	41	Protein 4.1	(9, 1.5)
P08195	4F2	4F2 cell-surface antigen heavy chain	(15, 1.5)
Q9UM73	ALK	ALK tyrosine kinase receptor	(4, 30.4)
P58335	ANTR2	Anthrax toxin receptor 2	(1, 2.8)
P07355	ANXA2	Annexin A2	(18, 1.6)
P13987	CD59	CD59 glycoprotein	(4, 1.5)
P29317	EPHA2	Ephrin type-A receptor 2	(5, 1.8)
Q9Y5Y0	FLVC1	Feline leukemia virus subgroup C receptor-related protein 1	(1, 1.8)

P11166	GTR1	Solute carrier family 2, facilitated glucose transporter member 1	(5, 2.1)
P23229	ITA6	Integrin alpha-6	(4, 1.8)
O00214	LEG8	Galectin-8	(1, 1.8)
P11717	MPRI	Cation-independent mannose-6-phosphate receptor	(2, 1.7)
O15440	MRP5	Multidrug resistance-associated protein 5	(1, 1.6)
Q9Y2A7	NCKP1	Nck-associated protein 1	(2, 1.9)
Q8IXS6	PALM2	Paralemmin-2	(1, 1.6)
P41219	PERI	Peripherin	(2, 1.7)
Q9Y6M7	S4A7	Sodium bicarbonate cotransporter 3	(4, 1.8)
P19634	SL9A1	Sodium/hydrogen exchanger 1	(2, 4.3)
Q13813	SPTA2	Spectrin alpha chain, non-erythrocytic 1	(34, 1.6)
Q01082	SPTB2	Spectrin beta chain, non-erythrocytic 1	(32, 1.6)
P02786	TFR1	Transferrin receptor protein 1	(6, 1.6)

Q9NP84	TNR12	Tumor necrosis factor receptor superfamily member 12A	(1, 2.7)
P46939	UTRO	Utrophin	(5, 1.6)
Q9BV40	VAMP8	Vesicle-associated membrane protein 8	(2, 1.6)
Q9UDY2	ZO2	Tight junction protein ZO-2	(9, 1.8)
Down-regulated Proteins			
P01892	1A02	HLA class I histocompatibility antigen, A-2 alpha chain	(2, 0.6)
P02768	ALBU	Serum albumin	(1, 0.3)
Q9NVJ2	ARL8B	ADP-ribosylation factor-like protein 8B	(1, 0.6)
Q9UJZ1	STML2	Stomatin-like protein 2	(5, 0.2)
Q9UNK0	STX8	Syntaxin-8	(1, 0.2)
P53007	TXTP	Tricarboxylate transport protein	(4, 0.2)

3.3.4. Bioinformatics analysis of differentially expressed plasma membrane proteins

Ingenuity Pathway Analysis software was used for bioinformatics analysis of the plasma membrane proteins and extracellular proteins listed in Table 3-2. The top bio functions of the differentially expressed proteins were cell-to-cell signaling and interaction, cell death and survival and cell cycle in molecular and cellular functions, hematological system development and function, hematopoiesis and tissue development in physiological system development and function.

ANXA2, ITA6, TFR1 and CD59 have the cell-to cell signaling and interaction function. Among these, CD59 have the function of adhesion of red blood cells and the other three proteins have the function of binding of lymphoma cell lines. For the other two molecular and cellular functions, ALBU, ANXA2 and CD59 possess the function of cell death and survival. Protein TFR1 possesses the function of cell cycle.

For the physiological system development and function, CD59 has the function of adhesion of red blood, as mentioned above. This was also involved in the category of hematological system development and function, hematopoiesis

and tissue development. Protein 4F2 also has the function of tissue development, i.e., aggregation of lymphoma cell lines.

Considering the protein functions, differential expression levels and potential clinical relevance, the important protein list in Table 3-2 was analyzed and this analysis generated a putative biomarker candidate list of EPHA2, GTR1, ITA6, SL9A1, STML2, and TXTP.

EPHA2 (ephrin type-A receptor 2) is a member of the Eph family of receptor tyrosine kinases. Investigations have demonstrated that EPHA2 possesses many important and diverse biological functions. The expression of EPHA2 has been detected in a wide assortment of tissues, such as brain, skin, bone marrow, lung, thymus, spleen, liver, small intestine, colon, and kidney.[19] Most tissues express low levels of EPHA2. The overexpression of EPHA2 was found in tumor cells and tumor blood vessels in many types of cancer including breast cancer,[20] prostate cancer,[21] lung cancer,[22] ovarian cancer,[23] renal cancer[24] and colorectal cancer.[25] It is involved in many processes crucial to malignant progression, such as migration, invasion, metastasis, proliferation, survival and angiogenesis. Due to the overexpression of EPHA2 in several cancers, EPHA2 is an attractive target for cancer therapeutics. Two drugs targeting EPHA2 have been approved. Dasatinib was approved by the FDA in 2006 for the treatment of imatinib-resistant chronic myeloid leukemia and Philadelphia chromosome-positive acute lymphoblastic

leukemia. Regorafenib was approved by the FDA in 2012 for the treatment of metastatic colorectal cancer and advanced gastrointestinal stromal tumors. In this work, EPHA2 was identified from both the Karpas 299 and SUPM2 cell lines. It was also quantified by using 2-MEGA isotopic labeling method. The result demonstrated that EPHA2 was overexpressed in NPM-ALK expressing HEK 293 stable cells and the fold change was 1.8. Therefore, the protein EPHA2 was chosen as a biomarker candidate and it could be a potential target for ALK+ Anaplastic Large Cell Lymphoma therapeutics.

GTR1 (solute carrier family 2, facilitated glucose transporter member 1) is a major glucose transporter, the integral membrane glycoprotein involved in transporting glucose into most cells. Increased glucose transport in malignant cells is associated with increased and deregulated expression of glucose transporter proteins.[26] In human studies, high levels of GTR1 expression in tumors have been associated with poor survival.[27] Thus, this protein could be targets in cancer therapy and prevention.[28] In this work, GTR1 was identified from both the Karpas 299 and SUPM2 cell lines. It was also quantified using 2-MEGA isotopic labeling method. The result demonstrated that GTR1 was also overexpressed in NPM-ALK expressing HEK 293 stable cells and the fold change was 2.1. Therefore, the protein GTR1 was chosen as a biomarker candidate and it could be a potential target for ALK+ Anaplastic Large Cell Lymphoma therapeutics and

prevention.

ITA6 (integrin alpha-6) is the protein highly expressed in embryonic, hematopoietic and neural stem cells.[29] It was reported that the glioblastoma stem cells express high levels of ITA6, which can be an enrichment marker and a promising antiglioblastoma therapy.[30] In this work, ITA6 was identified from both Karpas 299 and SUPM2 cell lines. 2-MEGA quantification demonstrated that ITA6 was overexpressed in NPM-ALK expressing HEK 293 stable cells and the fold change was 1.8. Therefore, the protein EPHA2 was chosen as a biomarker candidate.

SL9A1 (sodium/hydrogen exchanger 1) is a protein ubiquitously expressed in all mammalian cells and a multifunctional protein which involved in normal process such as intracellular pH regulation, cell volume regulation, cell growth, cell migration, apoptosis, and pathological processes such as cancer cell invasion and heart failure.[31] SL9A1 has been reported as a key player in breast cancer cell invasion.[32] The formation of pseudopodia in invasive tumor cells resulted from the cytoskeletal changes are SL9A1-dependent and can be prevented by SL9A1 inhibition.[33] There is a reason to believe that SL9A1 could be a possible cancer treatment target. In this work, SL9A1 was identified from both Karpas 299 and SUPM2 cell lines. It was also quantified by using 2-MEGA

isotopic labeling method. The result demonstrated that SL9A1 was overexpressed in NPM-ALK expressing HEK 293 stable cells and the change was 4.3 fold. Therefore, protein SL9A1 was chosen as a potential treatment target.

STML2 (stomatin-like protein 2) is a protein which associates with the plasma membrane, where it can form high-order oligomers and interact with the peripheral membrane skeleton. STML2 is up-regulated in many different cancer types, which is associated with higher levels of mortality.[34,35] However, loss of the function of STML2 is lethal for mice.[36] In this work, STML2 was identified from both the Karpas 299 and SUPM2 cell lines. 2-MEGA quantification demonstrated that STML2 was down-regulated in NPM-ALK expressing HEK 293 stable cells by 0.2 fold. This expression level change was different from the result reported before for the other types of cancer. Further investigation is needed to clarify the function of this protein in cells.

TXTP (tricarboxylate transport protein) is the protein involved in citrate-H^+/malate exchange. It is important for the bioenergetics of hepatic cells as it provides a carbon source for fatty acid and sterol biosynthesis, and NAD^+ for the glycolytic pathway. This protein may modulate the influx/efflux of drugs from cells and thus modulate chemotherapy response and could be the key therapeutic target for improving the chemotherapy response in ovarian cancers.[37] In this work,

TXTP was identified from Karpas 299 cell line. It was also quantified by using 2-MEGA isotopic labeling method. The result demonstrated that TXTP was down-regulated in NPM-ALK expressing HEK 293 stable cells by 0.2 fold. Therefore, the protein TXTP was chosen as a biomarker candidate and it could be a potential target for ALK+ Anaplastic Large Cell Lymphoma therapeutics.

In conclusion, by using bioinformatics analysis combined with literature search, six proteins were chosen as putative biomarkers for ALK+ Anaplastic Large Cell Lymphoma. Future work will involve biological validations to further demonstrate the result and this investigation should assist in furthering our understanding of NPM-ALK-induced tumorigenesis and finding new potential drug targets.

3.4. Conclusions

We developed a high throughput plasma membrane protein purification, identification, and quantitation strategy. The plasma membrane proteins of two NPM-ALK-expressing cell lines, Karpas 299 and SUPM2, were identified by using 2D LC-MS/MS analysis. 561 unique plasma membrane proteins and extracellular proteins were identified from Karpas 299 cell line and 552 unique plasma membrane proteins and extracellular proteins were identified from SUPM2 cell line. In total, 726 unique plasma membrane proteins and extracellular proteins were identified from biological duplicate samples. 387 plasma membrane proteins and extracellular proteins were identified in both Karpas 299 and SUPM2 cells. The number of the unique plasma membrane proteins and extracellular proteins identified only from one type of ALK+ ALCL cells was 174 for Karpas 299 and 165 for SUPM2. In addition, 2-MEGA labeling method combined with 2D LC-MS/MS was used for plasma membrane protein quantitation analysis in NPM-ALK expressing HEK 293 cells and NPM-ALK absent HEK 293 cells (control). 48 unique plasma membrane proteins and extracellular proteins were found to be differentially expressed between NPM-ALK expressing HEK 293 cells and NPM-ALK absent HEK 293 cells. Among these proteins, 39 were up-regulated and 9 were down-regulated. Of the 48 proteins, six proteins were selected as the putative biomarker by using bioinformatics analysis and literature

search. Future work will be the biological validation such as western blotting and immunohistochemical analysis. The biomarker candidates in the final list may have the potential to be used for ALK+ ALCL prognosis and also as drug targets for effective therapeutics.

3.5. Literature cited

1. Delsol, G.; Ralfkiaer, E.; Stein, H.; Wright, D.; Jaffe, E., Anaplastic large cell lymphoma. *Pathology and Genetics of Tumors of Haematopoietic and Lymphoid Tissues* **2001**, 230-235.

2. Morris, S. W.; Kirstein, M. N.; Valentine, M. B.; Dittmer, K.; Shapiro, D. N.; Look, A. T.; Saltman, D. L., Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **1995,** *267* (5196), 316-7.

3. Benz-Lemoine, E.; Brizard, A.; Huret, J. L.; Babin, P.; Guilhot, F.; Couet, D.; Tanzer, J., Malignant histiocytosis: a specific t(2;5)(p23;q35) translocation? Review of the literature. *Blood* **1988,** *72* (3), 1045-7.

4. Fischer, P.; Nacheva, E.; Mason, D. Y.; Sherrington, P. D.; Hoyle, C.; Hayhoe, F. G.; Karpas, A., A Ki-1 (CD30)-positive human cell line (Karpas 299) established from a high-grade non-Hodgkin's lymphoma, showing a 2;5 translocation and rearrangement of the T-cell receptor beta-chain gene. *Blood* **1988,** *72* (1), 234-40.

5. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P., Molecular Biology of the Cell (Garland Science, New York, 2002). **1997**.

6. Suprynowicz, F. A.; Disbrow, G. L.; Krawczyk, E.; Simic, V.; Lantzky, K.; Schlegel, R., HPV-16 E5 oncoprotein upregulates lipid raft components caveolin-1 and ganglioside GM1 at the plasma membrane of cervical cells. *Oncogene* **2008,** *27* (8), 1071-8.

7. Rabilloud, T., Membrane proteins ride shotgun. *Nat. Biotechnol.* **2003,** *21* (5), 508-510.

8. Lu, B.; McClatchy, D. B.; Kim, J. Y.; Yates, J. R., 3rd, Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry. *Proteomics* **2008,** *8* (19), 3947-55.

9. Bosserhoff, A.; Wallach, J.; Frank, R. W., Micropreparative separation of peptides derived from sodium dodecyl sulphate-solubilized proteins. *J. Chromatogr.* **1989,** *473* (1), 71-7.

10. Kawasaki, H.; Suzuki, K., Separation of peptides dissolved in a sodium dodecyl sulfate solution by reversed-phase liquid chromatography: removal of sodium dodecyl sulfate from peptides using an ion-exchange precolumn. *Anal. Biochem.* **1990,** *186* (2), 264-8.

11. Loo, R. R.; Dales, N.; Andrews, P. C., Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Sci.* **1994,** *3* (11), 1975-83.

12. Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A., Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J. Proteome Res.* **2010,** *9* (6), 2863-70.

13. Wu, F.; Sun, D.; Wang, N.; Gong, Y.; Li, L., Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis. *Anal. Chim. Acta* **2011,** *698* (1-2), 36-43.

14. 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.

15. 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.

16. Wang, N.; Xie, C.; Young, J. B.; Li, L., Off-line two-dimensional liquid chromatography with maximized sample loading to reversed-phase liquid chromatography-electrospray ionization tandem mass spectrometry for shotgun proteome analysis. *Anal. Chem.* **2009,** *81* (3), 1049-60.

17. Wang, N.; Li, L., Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis. *Anal. Chem.* **2008,** *80* (12), 4696-4710.

18. Sun, D.; Wang, N.; Li, L., Integrated SDS removal and peptide separation by strong-cation exchange liquid chromatography for SDS-assisted shotgun proteome analysis. *J. Proteome Res.* **2012,** *11* (2), 818-828.

19. Hafner, C.; Schmitz, G.; Meyer, S.; Bataille, F.; Hau, P.; Langmann, T.; Dietmaier, W.; Landthaler, M.; Vogt, T., Differential gene expression of Eph receptors and ephrins in benign human tissues and cancers. *Clin. Chem.* **2004,** *50* (3), 490-9.

20. Zelinski, D. P.; Zantek, N. D.; Stewart, J. C.; Irizarry, A. R.; Kinch, M. S., EphA2 overexpression causes tumorigenesis of mammary epithelial cells. *Cancer Res.* **2001,** *61* (5), 2301-6.

21. Walker-Daniels, J.; Coffman, K.; Azimi, M.; Rhim, J. S.; Bostwick, D. G.; Snyder, P.; Kerns, B. J.; Waters, D. J.; Kinch, M. S., Overexpression of the EphA2 tyrosine kinase in prostate cancer. *Prostate* **1999,** *41* (4), 275-80.

22. Stommel, J. M.; Kimmelman, A. C.; Ying, H.; Nabioullin, R.; Ponugoti, A. H.; Wiedemeyer, R.; Stegh, A. H.; Bradner, J. E.; Ligon, K. L.; Brennan, C.; Chin, L.; DePinho, R. A., Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **2007,** *318* (5848), 287-90.

23. Lu, C.; Shahzad, M. M. K.; Wang, H.; Landen, C. N.; Kim, S. W.; Allen, J.; Nick, A. M.; Jennings, N.; Kinch, M. S.; Bar-Eli, M.; Sood, A. K., EphA2 overexpression promotes ovarian cancer growth. *Cancer Biol. Ther.* **2008,** *7* (7), 1098-1103.

24. Cho, M. C.; Cho, S. Y.; Lee, S. B.; Kwak, C.; Lee, H. W.; Kim, H. H.; Jeong, H., Role of Epha2 in Malignant Cellular Behavior in Renal Cell Carcinoma Cells. *J. Urol.* **2013,** *189* (4), E124-E124.

25. Saito, T.; Masuda, N.; Miyazaki, T.; Kanoh, K.; Suzuki, H.; Shimura, T.; Asao, T.; Kuwano, H., Expression of EphA2 and E-cadherin in colorectal cancer: correlation with cancer metastasis. *Oncol. Rep.* **2004,** *11* (3), 605-11.

26. Macheda, M. L.; Rogers, S.; Best, J. D., Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. *J. Cell. Physiol.* **2005,** *202* (3), 654-62.

27. Younes, M.; Brown, R. W.; Stephenson, M.; Gondo, M.; Cagle, P. T., Overexpression of Glut1 and Glut3 in stage I nonsmall cell lung carcinoma is associated with poor survival. *Cancer* **1997,** *80* (6), 1046-51.

28. Lu, Y.; Lemon, W.; Liu, P. Y.; Yi, Y.; Morrison, C.; Yang, P.; Sun, Z.; Szoke, J.; Gerald, W. L.; Watson, M.; Govindan, R.; You, M., A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med.* **2006,** *3* (12), e467.

29. Fortunel, N. O.; Otu, H. H.; Ng, H. H.; Chen, J.; Mu, X.; Chevassut, T.; Li, X.; Joseph, M.; Bailey, C.; Hatzfeld, J. A.; Hatzfeld, A.; Usta, F.; Vega, V. B.; Long, P. M.; Libermann, T. A.; Lim, B., Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* **2003,** *302* (5644), 393; author reply 393.

30. Lathia, J. D.; Gallagher, J.; Heddleston, J. M.; Wang, J.; Eyler, C. E.; Macswords, J.; Wu, Q.; Vasanji, A.; McLendon, R. E.; Hjelmeland, A. B.; Rich, J. N., Integrin alpha 6 regulates glioblastoma stem cells. *Cell Stem Cell* **2010,** *6* (5), 421-32.

31. Malo, M. E.; Fliegel, L., Physiological role and regulation of the Na+/H+ exchanger. *Can. J. Physiol. Pharmacol.* **2006,** *84* (11), 1081-1095.

32. Reshkin, S. J.; Bellizzi, A.; Guerra, L.; Casavola, V., Serum-deprivation increases human breast cancer cell motility and invasion by activating the Na/H exchanger (NHE1): Involvement of p38 map kinase pathway. *Pflugers Archiv-European Journal of Physiology* **2000,** *439* (5), R259-R259.

33. Lagana, A.; Vadnais, J.; Le, P. U.; Nguyen, T. N.; Laprade, R.; Nabi, I. R.; Noel, J., Regulation of the formation of tumor cell pseudopodia by the Na+/H+ exchanger NHE1. *J. Cell Sci.* **2000,** *113* (20), 3649-3662.

34. Wang, Y.; Cao, W.; Yu, Z.; Liu, Z., Downregulation of a mitochondria associated protein SLP-2 inhibits tumor cell motility, proliferation and enhances cell sensitivity to chemotherapeutic reagents. *Cancer Biol. Ther.* **2009,** *8* (17), 1651-8.

35. Luque-Garcia, J. L.; Martinez-Torrecuadrada, J. L.; Epifano, C.; Canamero, M.; Babel, I.; Casal, J. I., Differential protein expression on the cell surface of

colorectal cancer cells associated to tumor metastasis. *Proteomics* **2010,** *10* (5), 940-52.

36. Lapatsina, L.; Brand, J.; Poole, K.; Daumke, O.; Lewin, G. R., Stomatin-domain proteins. *Eur. J. Cell Biol.* **2012,** *91* (4), 240-5.

37. Cheng, L.; Lu, W.; Kulkarni, B.; Pejovic, T.; Yan, X.; Chiang, J. H.; Hood, L.; Odunsi, K.; Lin, B., Analysis of chemotherapy response programs in ovarian cancers by the next-generation sequencing technologies. *Gynecol. Oncol.* **2010,** *117* (2), 159-69.

Chapter 4 Integrated SDS Removal and Peptide Separation by Strong-Cation Exchange Liquid Chromatography for SDS-Assisted Shotgun Proteome Analysis

A version of this chapter was published as Sun, D. F.; Wang, N.; Li, L., Integrated SDS removal and peptide separation by strong-cation exchange liquid chromatography for SDS-assisted shotgun proteome analysis. *J. Proteome Res.* 2012, *11* (2), 818-828.

4.1 Introduction

Bottom-up or shotgun proteome analysis requires proteins to be solubilized in a proper solvent for enzyme or chemical digestion. During a proteome sample workup, some proteins, particularly hydrophobic proteins, such as membrane proteins, cannot be readily dissolved in conventional solvents, including buffers, chaotropic agents (e.g., urea), aqueous-organic solvents (e.g., methanol-water), and organic acids (e.g., formic acid). Surfactants can aid in dissolving proteins, and of these, sodium dodecyl sulfate (SDS), an ionic detergent with strong solubilization capability, can be used to dissolve and denature a wide range of proteins, including integral membrane proteins.[1] SDS-assisted enzyme digestion can be done without an adverse effect on digestion efficiency as long as the SDS concentration is kept low (i.e., <0.1%).[2,3] However, the resultant peptides containing SDS cannot be analyzed directly by reversed-phase (RP) LC-MS, as SDS can cause interference with the RPLC separation[4,5] and MS analysis.[6,7] Therefore, efficient removal of SDS is required before RPLC tandem MS (MS/MS) analysis. Alternatively, SDS-mimic surfactants, such as acid labile surfactants (e.g., RapiGest from Waters, PPS from Protein Discovery and Invitrosol from Invitrogen), may be used.[8-16] These reagents rapidly decompose into two easily removed products on adding an acid, such as trifluoroacetic acid (TFA), to the digested protein sample. However, the

cost of these reagents is generally high, which can be a major concern in large scale proteome analysis (e.g., proteome profiling of multiple organ or tissue samples) or in situations where a large amount of surfactant is required to process a sample (e.g., working with proteins electro-eluted from a polyacrylamide gel in SDS solution).[17]

A number of techniques have been reported for SDS removal with varying degrees of success. Most of the methods, such as the use of spin columns,[18] dialysis,[19] and ultrafiltration device,[20] involve a few steps of manual operation. On-line SDS removal using a de-SDS cartridge or column has also been described.[21, 22] Aside from the convenience and robustness of a method, one major issue on SDS removal is related to sample loss. There are two types of sample loss. One is merely losing some sample without altering the composition; using a larger starting material, if available, can overcome this type of sample loss without affecting the final proteome coverage. However, if sample loss is accompanied with a change in proteome composition (i.e., selective or uneven loss of proteins), the proteome profile generated will not accurately reflect the proteome of a biological system. Thus, in gauging the performance of an SDS removal method, vigorous testing is needed to ensure the SDS removal process does not alter the proteome profile. Recent work in our group has shown that SDS removal by SCX can result in a protein composition change which correlated to

the sample loss.[16] More recently, a proprietary resin has been introduced for removing detergents including SDS.[23, 24] According to the manufacturer, 99% of the SDS could be removed from a protein sample containing 1% SDS and the sample recovery rate was found to range from 81% for carbonic anhydrase to 100% for bovine serum albumin (BSA).

In this work, we describe an improved method based on the use of strong-cation exchange (SCX) LC for removal of SDS as well as peptide separation in an integrated chromatographic procedure. Since SCX is widely used as the first dimension of two-dimensional (2D) LC MS/MS in shotgun proteome analysis, integration of the SDS removal with SCX peptide separation does not add any extra steps to the sample handling process. There were reports of using an SCX cartridge or column for SDS removal.[21,22] However, selective sample loss can be a major problem.[16] To develop the new protocol, we systematically investigated the experimental parameters affecting sample loss and were surprised to discover that, by increasing the SDS concentration, as opposed to decreasing the concentration, the sample recovery rate could be significantly improved. By adding SDS to a digested peptide sample to reach an SDS concentration of 0.5% and selecting appropriate types and concentrations of reagents in the protein sample preparation and SCX, we could effectively remove SDS while achieving a peptide recovery rate of about 90%. This rate of recovery was essentially the same

as that obtained by running a peptide sample containing no SDS in SCX. We also compared the performance of this SDS-assisted shotgun method with the RapiGest- and PPS-assisted methods. Finally, this method was applied to the analysis of the membrane-protein-enriched fraction of a whole cell lysate of a breast cancer cell line (MCF-7).

4.2 Experimental Section

4.2.1 Materials and reagents

Iodoacetamide (IAA), ammonium bicarbonate, bovine serum albumin (BSA), potassium phosphate monobasic (KH_2PO_4), trifluoroacetic acid (TFA), and LC-MS grade formic acid (FA) were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). Dithiothreitol (DTT) and sodium dodecyl sulfate (SDS) were from Bio-Rad Laboratories (Mississauga, ON, Canada). Sequencing grade modified trypsin, LC-MS grade water, acetone, and HPLC grade acetonitrile (ACN) were from Fisher Scientific Canada (Edmonton, AB, Canada). Phosphoric acid (H_3PO_4), potassium chloride (KCl), and sodium phosphate monobasic (NaH_2PO_4) were purchased from Anachemia (Edmonton, AB, Canada). Lithium carbonate (Li_2CO_3) was from VWR (Mississauga, ON, Canada). LiH_2PO_4 was prepared by reacting Li_2CO_3 with H_3PO_4. RapiGest^{TM} SF was purchased from Waters (Milford, MA) and PPS Silent Surfactant was from Protein Discovery (Knoxville, TN). The BCA assay kit was from Pierce (Rockford, IL).

4.2.2 Cell culture and membrane protein preparation

The cell culture and preparation of the membrane protein fraction of the cell lysates of *E. coli* were similar to those reported.[16] A single colony of *Escherichia coli* K12 (ATCC no.: 47076) from frozen glycerol stocks was taken

to incubate in 2 × 100 mL Lysogeny Broth (LB) medium (1% tryptone, 0.5% yeast extract, and 1% NaCl from Sigma-Aldrich) at 37 °C with shaking at 225 rpm overnight. 2 × 30 mL of the culture were centrifuged at 3901 *g* for 10 min and then the pellets were resuspended in 2 × 500 mL LB. Cells were incubated at 37 °C with shaking at 200 rpm for 3.5 h. After cooling the medium to 4 °C, the cells were centrifuged at 11300 *g* for 20 min at 4 °C. The pellets were washed with 300 mL of 4 °C PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) on ice and then centrifuged at 11300 *g* for 20 min at 4 °C. The pellets were re-suspended in 60 mL of 4 °C water and 2 tablets of Complete Mini EDTA-free Protease Inhibitor Cocktail (Roche 1183617000) were added. The suspension was passed through a cell disruptor at 25000 psi and chased with 20 mL of 4 °C water. The lysate was centrifuged at 9820 *g* for 20 min at 4 °C to pellet the unbroken cells. 600 µL of 0.5 M EDTA was added to the supernatant. The supernatant was centrifuged at 113000 *g* for 1 h at 4 °C. The pellet was then suspended in 1.2 mL of 50 mM NH_4HCO_3 and 1.2 mL of 100 mM Na_2CO_3 and centrifuged at 119000 *g* for 1 h at 4 °C twice. The pellet was resuspended in 14 mL of 4 °C 50 mM MOPS (pH 7.3). The suspension was homogenized on ice until homogeneous and then centrifuged at 119000 *g* for 1 h at 4 °C. The pellet (i.e., the membrane protein fraction) was stored at -80 °C.

The cell culture and preparation of the membrane protein fraction of the MCF-7 cell lysates were similar to those reported.[16] The MCF-7 breast cancer cells (ATCC no.: HTB-22) were cultured in 15 cm diameter plates at 37 °C in growth medium (90% D-MEM medium, 10% fetal calf serum, 10 μg/mL insulin from bovine serum). Each plate was washed with 15 mL of 37 °C PBS. Five mL of 37 °C trypsin, 0.05% with EDTA, was added to each plate. The plates were incubated at 37 °C for 10 min. Fifteen mL of 37 °C growth medium was added to each plate. Cells were resuspended and then centrifuged at 125 *g* for 7 min. The pellets were re-suspended in 45 mL of 4 °C PBS and then centrifuged at 200 *g* for 7 min at 4 °C to remove residual medium. The pellets were re-suspended in 20 mL of 4 °C Triton X-114 lysis buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5 mM EDTA, 2% Triton X-114, Complete Mini EDTA-free protease inhibitor cocktail tablets, 1 tablet per 20 mL) and stirred overnight. The suspension was centrifuged at 20400 *g* for 30 min at 4 °C. The supernatant was incubated at 37 °C for 30 min and centrifuged at 20400 *g* for 30 min at room temperature to separate the phases. Aqueous phase and detergent phase were collected. 25 mL of 4 °C PBS was added to the detergent phase and 5 mL of 4 ˚C 10% Triton X-114 in PBS was added to the aqueous phase. After stirring at 4 °C for 1 h and incubating at 37 °C for 30 min, the aqueous phase was centrifuged at 20400 *g* for 30 min at room temperature. The detergent phase was collected and one more phase

separation was performed on the aqueous phase. In total, 14.6 mL of the detergent phase was collected to which 140 mL acetone was added and the sample was incubated at -20 °C overnight. After centrifuging the sample at 20400 *g* for 30 min at 4 °C, the pellet (i.e., the membrane protein fraction) was stored at -80 °C.

4.2.3 Protein solubilization and digestion

Figure 4-1 shows the workflow for SDS-assisted shotgun proteome analysis. Various protein or proteome samples were analyzed in this work for method development, as well as for demonstrating the performance of the method. BSA was used as a standard for the initial method development. BSA was dissolved in 50 mM NH_4HCO_3 with or without 1% SDS. For the SDS-containing samples, the solution was diluted to an SDS concentration of 0.025%, followed by reduction and alkylation. To 3000 µL of 1 µg/µL BSA solution, 60 µL (or 20 µL in the experiments involving low concentrations of DTT and IAA; see Results and Discussion) of 900 mM DTT was added to reduce the disulfide bonds. The samples were incubated at 37 °C for 1 h. Then 600 µL (or 250 µL) of 200 mM IAA was added. Alkylation was performed at room temperature for 1 h in the dark. Trypsin was added to the protein sample at a mass ratio of 1:50. Digestion was conducted at 37 °C for 48 h. The samples were dried down at 85 °C with a heater (model 110002, Boekel Scientific, Feasterville, PA) and continuously heated for 2 h to remove NH_4HCO_3 (see Results and Discussion).

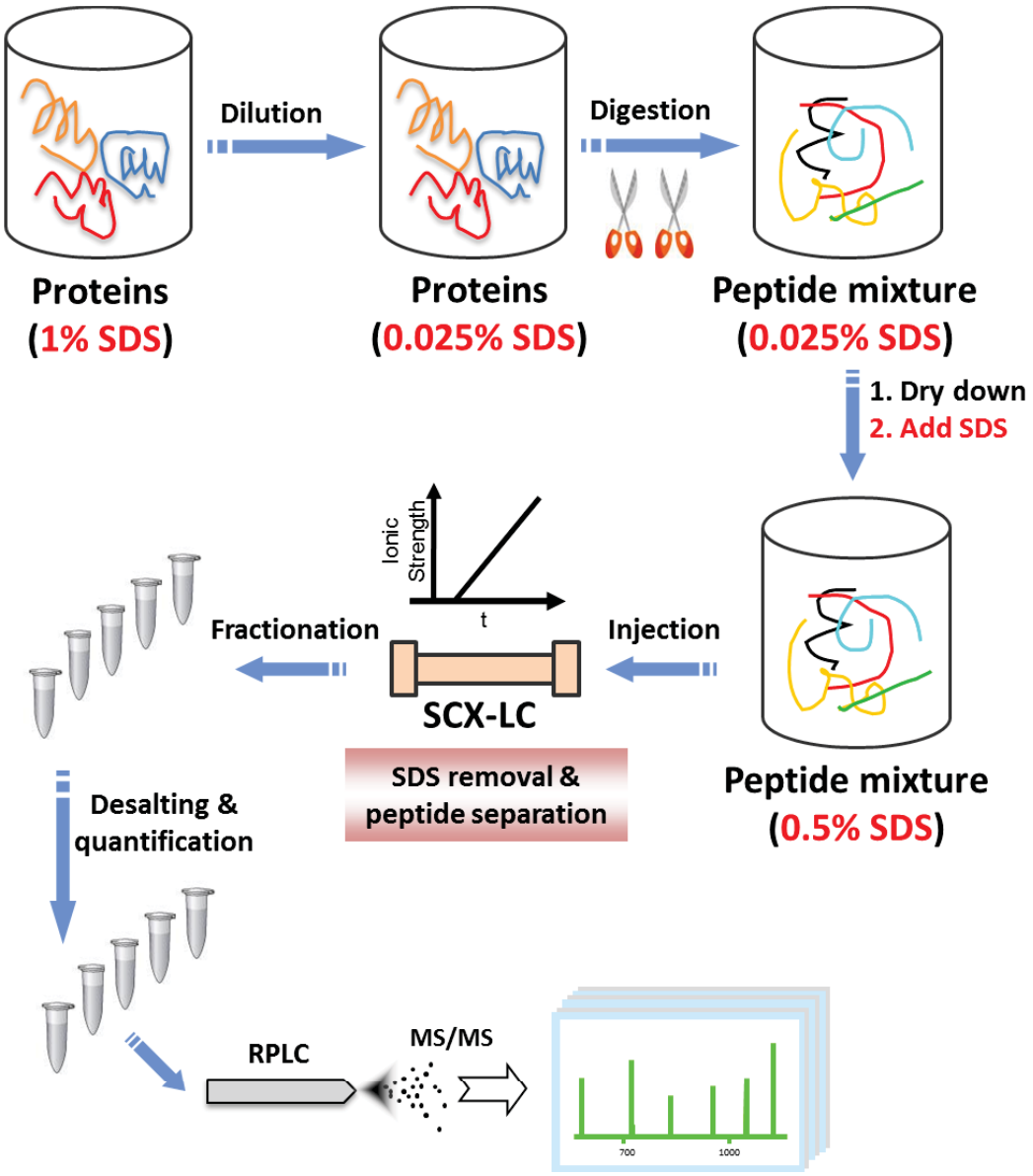


Figure 4-1. SDS-assisted proteome analysis workflow.

The *E. coli* or MCF-7 membrane protein pellets were dissolved in 1% SDS, 1% RapiGest or 1% PPS. Protein concentration was determined using the BCA assay. Reduction and alkylation were conducted after 10-fold dilution of the RapiGest and PPS samples and 40-fold dilution of the SDS sample. Twenty μL of 900 mM DTT and 250 μL of 200 mM IAA were used for 3000 μL of 1 μg/μL protein solution. Trypsin digestion was performed at a mass ratio of 1:50. RapiGest- and PPS-assisted digestions were stopped by adding 10% TFA to acidify the peptide solution. Peptide solutions were then incubated at 37 °C for 1 h followed by centrifugation at 20000 *g* for 10 min at 4 °C. SDS-assisted digestion was stopped by drying down at 85 °C.

4.2.4 SDS removal and desalting

The dried peptide mixtures were reconstituted in H_3PO_4/H_2O (pH 1 or 2). Different concentrations of SDS (0.025%, 0.05%, 0.1%, 0.25%, 0.5%, and 1%) were prepared by adding SDS to the protein digest samples. The SDS removal was performed by SCX chromatography using an Agilent 1100 HPLC system with a 2.1 × 250 mm polySULFOETHYL^{TM} A column (Part #: 252SE0503, PolyLC, Columbia, MD). Mobile phase A used in SCX was either 10 mM KH_2PO_4, 10 mM NaH_2PO_4 , or 10 mM LiH_2PO_4 at pH 2.67 (adjusted with phosphoric acid) and mobile phase B was 500 mM KCl and 10 mM KH_2PO_4 (pH 2.67). After sample loading onto the column, a 10 min column flush with 100%

mobile phase A at 0.2 mL/min was used to remove the SDS and salt in the sample. Then the column was flushed at 0.2 mL/min with 100% mobile phase B to elute all the peptides without fractionation. The peptide samples collected were dried down with a SpeedVac and reconstituted in 0.1% TFA aqueous solution.

The sample desalting step was performed by RPLC with an Agilent 1100 HPLC system fitted with a 4.6 × 50 mm C18 column (Part #: 2001-050×046, Varian, Ontario, Canada). After sample loading, the column was first flushed with 97.5% mobile phase A (0.1% TFA/H_2O) and 2.5% mobile phase B (0.1% TFA/ACN) for 5 min to remove the salts. Then, 85% mobile phase B was used to completely elute the peptides. The collected samples were dried down and reconstituted in 0.1% FA/H_2O. The amount of peptide was determined based on the UV absorbance at 214 nm.[25] The peptide recovery rate was calculated by the peptide amounts measured before and after the SCX fractionation. For the RapiGest and PPS samples, only the RPLC step was performed for desalting. The peptide amount was also determined on the basis of UV absorbance at 214 nm.

4.2.5 SDS measurement

The concentration of SDS was measured using gas chromatography (GC) MS as previously described with some modifications.[26] The 30 m × 0.54 mm 1.5 μm thickness DB-5 (5% phenyl methyl siloxane) column was used for all gas chromatographic measurements. The desalted BSA digest samples containing

different concentrations of SDS (0.00003%, 0.0003%, 0.003%, and 0.03%) were prepared to establish a calibration curve of the peak area of 1-dodecanol, the SDS pyrolysis product formed at the GC-MS inlet, vs. SDS concentration. Two μL of the peptide mixture was taken from a sample, followed by 10-fold dilution with 8 mM guanidine·HCl MeOH solution. Three μL of the solution was injected into the GC-MS. The injection temperature was 320 °C and the oven temperature was initially set to 50 °C on sample injection. After 0.5 min, the oven temperature was ramped up at 15 °C/min to a final temperature of 320 °C and held for 2 min. The peak area of 1-dodecanol in the ion chromatogram of a sample was used to determine the SDS concentration based on the established calibration curve. This method was used to measure the SDS concentration in the SCX fractions after desalting.

To measure the SDS concentration in a SCX fraction before desalting, another calibration curve was established where the BSA digests were injected into the SCX column and the peptides were eluted and then collected from the column in the same manner as the SDS removal experiment described above. In the collected BSA peptide samples, SDS was added to form a series of peptide solutions containing different concentrations of SDS (0.00006%, 0.0003%, 0.001%, 0.003%, and 0.03%). Pyrolysis GC-MS was used, as described above, to

produce a calibration curve which was used to measure the SDS concentration in the SCX fractions directly.

4.2.6 SCX in 2D-LC MS/MS

For the MCF-7 membrane protein fraction, the SDS-assisted shotgun method using the workflow shown in Figure 4-1 was applied to generate the proteome profile. In this case, the MCF-7 protein digest was reconstituted in 0.2% H_3PO_4/H_2O (pH 2). SDS was added to reach a concentration of 0.5%. SDS removal and gradient elution were performed by using mobile phase A (10 mM LiH_2PO_4 with 20% ACN, pH 2.67) and mobile phase B (500 mM KCl and 10 mM KH_2PO_4 with 20% ACN, pH 2.67) at a flow rate of 0.2 mL/min. The gradient profile was as follows: 0 min: 0% B, 10 min: 0% B, 10.01 min: 4% B, 30 min: 12% B, 40 min: 20% B, 50 min: 55% B, 52 min:100% B, 60 min: 100% B, 60.01 min: 0% B, and 80 min: 0% B. In total, 60 fractions were collected and directly desalted and quantified by RPLC-UV, as described above. The adjacent low concentration fractions were combined to produce a sufficient amount of peptide (i.e., >1.5 μg); 25 fractions were generated for LC-MS.

4.2.7 LC-MS analysis

The desalted peptide mixtures were analyzed using a quadrupole time-of-flight (QTOF) Premier mass spectrometer (Waters, Milford, MA)

equipped with a nanoACQUITY Ultra Performance LC system (Waters). In brief, a 1.5 μg sample was injected each time onto a 75 μm × 150 mm Atlantis C18 column (Part #: 186003500, Waters, Milford, MA). Mobile phase A consisted of 0.1% FA/H_2O and mobile phase B consisted of 0.1% FA/ACN. The peptide sample was separated using a 130-min gradient with the following profile: 0 min: 2% B, 2 min: 7% B, 85 min: 20% B, 110 min: 30% B, 115 min: 45% B, 120 min: 90% B, 125 min:90% B, and 130 min: 2% B. All samples were electrosprayed into the mass spectrometer at a flow rate of 350 nL/min. MS data were recorded within a m/z window of 350-1600 for 0.8 s, followed by 4 data-dependent MS/MS scans of the four most intense ions within an m/z window of 50-1990 for 0.8 s. For dynamic mass exclusion, a time window of 180 s and a mass tolerance window of 50 mDa were applied. Various collision energies were used based on the mass and charge state of the peptide. Mass scans of a mixture of mass calibrants consisting of leucine enkephalin and (Glu1)-fibrinopeptide B were acquired at 1 scan/min throughout the run.

For 2D-LC MS/MS, the precursor ion exclusion (PIE) method[27] was applied to enhance the peptide and protein identification. In brief, an exclusion list was generated based on the peptides identified from the Mascot search program (Matrix Science, London, U.K.). The m/z value, charge state, and retention time of each identified peptide were extracted from the database search results and the

corresponding raw data. The m/z value of the other charge state for each identified precursor ion was calculated. In addition, the values of all identified peptides consisted of, not only the monoisotope value, but also the three additional isotope values. Finally, all of the m/z values along with their retention time information were loaded into the MS method for the new LC-MS run.

4.2.8 Protein database search

Database searches were performed as previously described with minor changes.[28] Raw search data were lockmass-corrected, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.3 (Waters). Peptide sequences were identified by automated database searching of peak list files using the Mascot search program. Database searching was restricted to *E. coli* for the *E. coli* membrane protein digest and *Homo sapiens* (human) for the MCF-7 membrane protein digest in Swissprot (version 4 and accession date of May 7, 2009). The following search parameters were selected for all database searching: enzyme, trypsin; missed cleavages, 1; peptide tolerance, 30 ppm; MS/MS tolerance, 0.2 Da; peptide charge, 1+, 2+, and 3+; fixed modification, carbamidomethyl (C); variable modifications, oxidation (M) and deamidated (N and Q). The search results, including protein names, access IDs, molecular mass, unique peptide sequences, ion score, Mascot threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated

masses were extracted to Excel files using in-house software. All of the identified peptides with scores lower than the Mascot threshold score for identity at the confidence level of 95% were then removed from the protein list. The redundant peptides for different protein identities were deleted, and the redundant proteins identified under the same gene name but different access ID numbers were also removed from the list. Finally, a unique protein or peptide list was generated by merging all the protein and peptide lists from the individual runs. Unique proteins (under unique gene names) and peptides with the highest scores were kept. The false positive peptide matching rate was gauged by using the target-decoy search strategy[29] by searching the MS/MS spectra against the forward and reversed proteome sequences.

4.3 Results and discussion

4.3.1 Method development

The overall workflow of the SDS method is shown in Figure 4-1. Briefly, proteins solubilized in 1% SDS were diluted to reduce the adverse effect of SDS in trypsin digestion. After digestion, SCX can be used to remove SDS. However, this step is often accompanied with selective sample loss that results in the reduction of the proteome coverage. To gauge the effect of SDS on sample loss, we injected the BSA digests spiked with different concentrations of SDS into SCX-LC, followed by washing off SDS and subsequent elution of the peptides with a step gradient, as described in the Experimental section. The eluted peptides were subjected to RPLC-UV for peptide quantification. Figure 4-2 (a) plots the peptide recovery rate as a function of the SDS concentrations present in the BSA digests. The results of the protein sequence coverage generated by RPLC MS/MS analysis of the recovered peptides from the SCX column are shown in Figure 4-2 (b). Formic acid was used in this initial work to acidify the peptide samples for SCX. Two different concentrations of formic acid were investigated to gauge the pH effect on the peptide recovery rate. As Figure 4-2 (a) shows, sample loss becomes more severe as the SDS concentration increases from zero to 0.025%. Sequence coverage was also reduced. Note that the same amount of peptide from

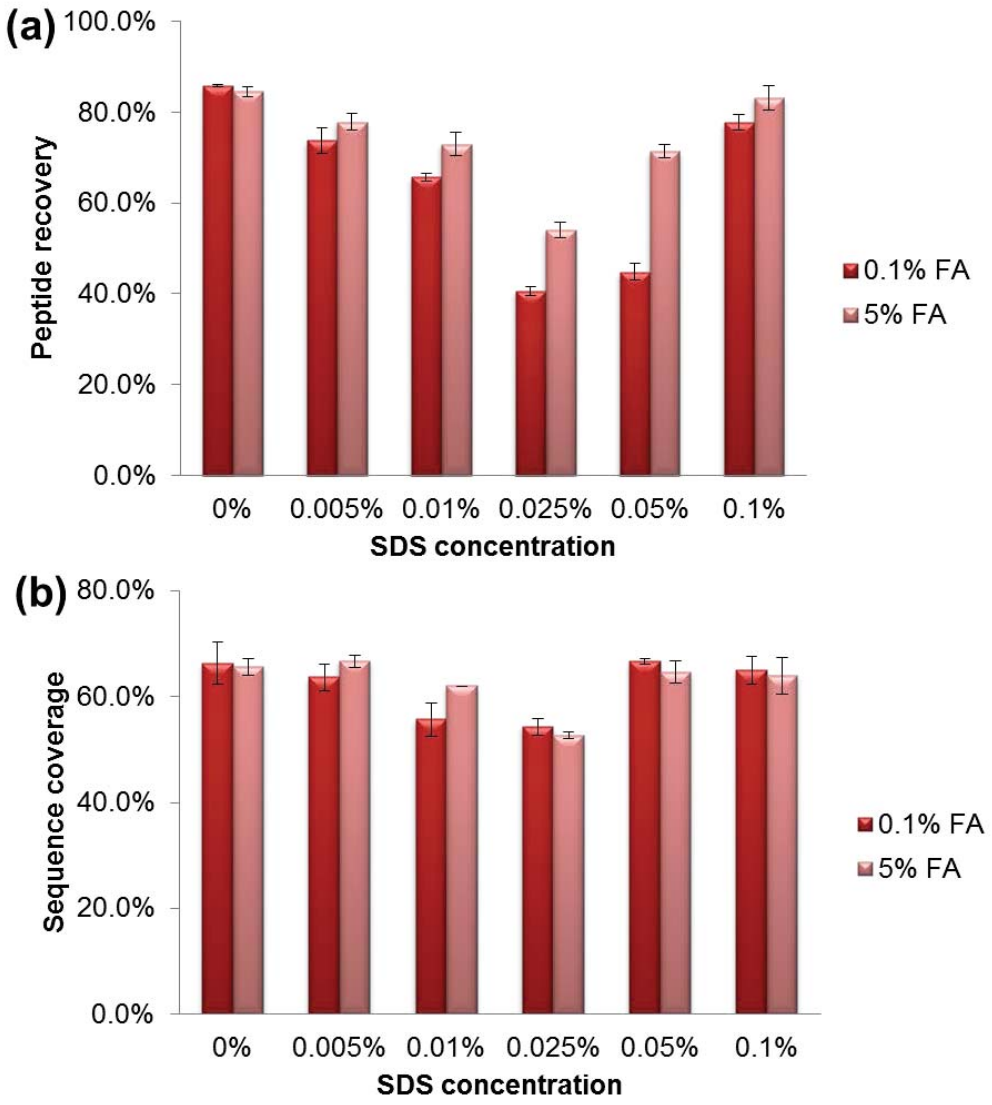


Figure 4-2. Comparison of (a) the peptide recovery rates and (b) the protein sequence coverage after SDS removal by SCX of BSA digests containing different concentrations of SDS (n=3). Error bars represent one standard deviation.

the individual samples was injected into the RPLC MS/MS. Thus, the reduction in the sequence coverage was the result of uneven or selective loss of certain peptides in the digest. Analysis of the distributions of peptides as a function of their GRAVY index or isoelectric point (pI) could not find any direct correlation between the GRAVY index or pI and the type of peptides lost.

Interestingly, as Figure 4-2 (a) shows, a further increase in SDS concentration from 0.025% to 0.05% or 0.1% SDS results in less sample loss or an increase in peptide recovery rate. This finding was somewhat surprising. However, it may be explained by considering the possible change of separation mechanisms as the SDS concentration increases. The key mechanism underlying the SDS removal by SCX is the retention of positively charged peptides on the negatively charged stationary phase while the negatively charged SDS is not retained on the column and can be washed away. However, SDS can interact with a multiply charged peptide to form a complex that can still remain positively charged. This complex could be retained by the column during the washing step and then co-elute with peptides during the peptide elution step. The peptides collected from SCX will contain SDS. On the other hand, if SDS is dissociated from the SDS-peptide complex during the washing step, the SDS will be removed and the peptides collected from SCX will not contain SDS. We have measured the SDS concentrations in the peptide fractions collected from SCX after loading of

the original samples containing varying concentrations of SDS, i.e., 0.025%, 0.05%, 0.1%, 0.25%, 0.5% and 1%, and column washing to remove SDS. After drying the individual fractions to the same volume as the original sample, we used the GC-MS method to quantify the SDS. We could not detect any SDS signals from the un-desalted SCX fractions using this method, which had a detection limit of about 0.00006%, and the desalted SCX fractions, which had a detection limit of about 0.00003%. It appears that SDS removal was quite efficient using the SCX column under the conditions used in this work. However, severe sample loss was observed when the SDS concentration in the sample was equal to or less than 0.025%. This is most likely due to the interaction of SDS with the positively charged peptides to form neutral or negatively charged complexes which will not be retained on the column. For example, a singly charged peptide bound with one negatively charged SDS molecule would form a neutral species. A doubly charged peptide may interact with three SDS molecules via two charge-charge bonds and one hydrophobic interaction to form a negatively charged complex.

As Figure 4-2 (a) shows, when the SDS concentration is 0.1%, the peptide recovery rate is similar to that of the digest containing no SDS. This may be attributed to the formation of micelles when the SDS concentration is near or above the critical micelle concentration (CMC). The CMC in pure water at 25 °C is 8.2 mM (0.24%) and decreases as the salt concentration increases due to the

screening effect of the salt cation reducing the repulsion of the ionic tails.[30] For example, the CMC is reduced to 5.3 mM (0.16%) and 3.3 mM (0.10%) when 10 mM and 30 mM NaCl or Na_3PO_4 are added, respectively, to the SDS solution.[30] The CMC of the solution used in our SCX experiment is unknown, but is expected to be lower than 0.16% as there was at least 10 mM salt present in mobile phase A. Micelles with negative charges on their surfaces and hydrophobic cores will not interact with the negatively charged SCX stationary phase. Thus, SDS can be washed away with ease. Positively charged peptides can interact with the micelle surface as well as the stationary phase. Strong interaction of peptides with the micelles would result in the elution of these peptides during the washing step. However, the data shown in Figure 4-2 (a) suggest that sample loss is not as severe as in the cases where the SDS concentration is far below the CMC. It should be noted that, when the concentration of free SDS increases above the CMC, most SDS would be present in the form of micelles, leaving little free SDS left in the solution. The net effect is equivalent to the decrease of free SDS concentration to a very low level. Thus, micelle formation is beneficial for SDS removal by SCX. It is worth noting that Tummala and Limbach have studied the effect of SDS micelles on peptide mass fingerprinting by MALDI-MS and observed an increase in the number of peptides detected from tryptic digests of several standard proteins when 0.1-0.3% SDS was purposely added to the digests

prior to MALDI analysis.[31,32] It was postulated that the formation of micelles helped the incorporation of peptides and small proteins into matrix crystals.

4.3.2 Effects of acids

Figure 4-2 (a) also shows the effect of formic acid concentration on the peptide recovery rate. For the SDS-containing samples, the use of 5% FA gave better peptide recovery rate than 0.1% FA. There was no significant difference for the samples containing no SDS. These results can be rationalized by the interplay of the interactions among SDS, peptides and stationary phase. The use of higher FA concentration increases the propensity to form multiply charged positive species that will be retained more readily on the stationary phase. For example, the C-terminal carboxyl group can be protonated. As a result, this peptide may be retained on the column instead of being washing out during the SDS removal step when the pH of the solution is higher. Thus, it is important to charge the peptides as much as possible.

We also examined the use of phosphoric acid (its salt is part of the buffer systems used in SCX), instead of formic acid, to prepare the peptide samples. The recovery rate was found to be 81.3% and 85.0% from the BSA digests that were acidified with phosphoric acid to pH 2 and 1, respectively, compared to 77.8% with 0.1% FA and 83.2% with 5% FA. The recovery rate for the phosphoric acid samples of pH 1 and pH 2 containing 0.25% or 0.50% SDS was similar, i.e., in

the range, 87.0 to 88.0%. Considering that the use of pH 1 on a routine basis may reduce the column lifetime, in the subsequent experiments, we chose to use phosphoric acid to adjust the pH to 2 for preparing the peptide samples before injection into SCX.

4.3.3 Effects of mobile phase composition

Another parameter examined was the composition of the mobile phase used to wash the column to remove the SDS. For SCX, 10 mM KH_2PO_4 (pH 2.7) is normally used as mobile phase A, while 10 mM KH_2PO_4 and 0.5 M KCl (pH 2.7) is used as mobile phase B. Mobile phase A is used to wash the column to remove SDS. It is thus desirable to create a condition that favors the binding of peptides to the column while SDS can be readily eluted. Peptide retention on the ion exchange stationary phase involves the displacement of the counter ions initially present in the column which are the same as the cations in mobile phase A after the column has been re-equilibrated using this mobile phase. The order of readiness for cation exchange is $Li^+ > H^+ > Na^+ > NH_4^+ > K^+$.[33] We tested the effect of different cations in mobile phase A on the peptide recovery rate and the results are shown in Figure 4-3 (a). The average recovery rate from the use of Li^+ appears to be slightly higher than Na^+ or K^+, but there are no statistical differences, indicating that these ions behave similarly as the cations in mobile phase A on

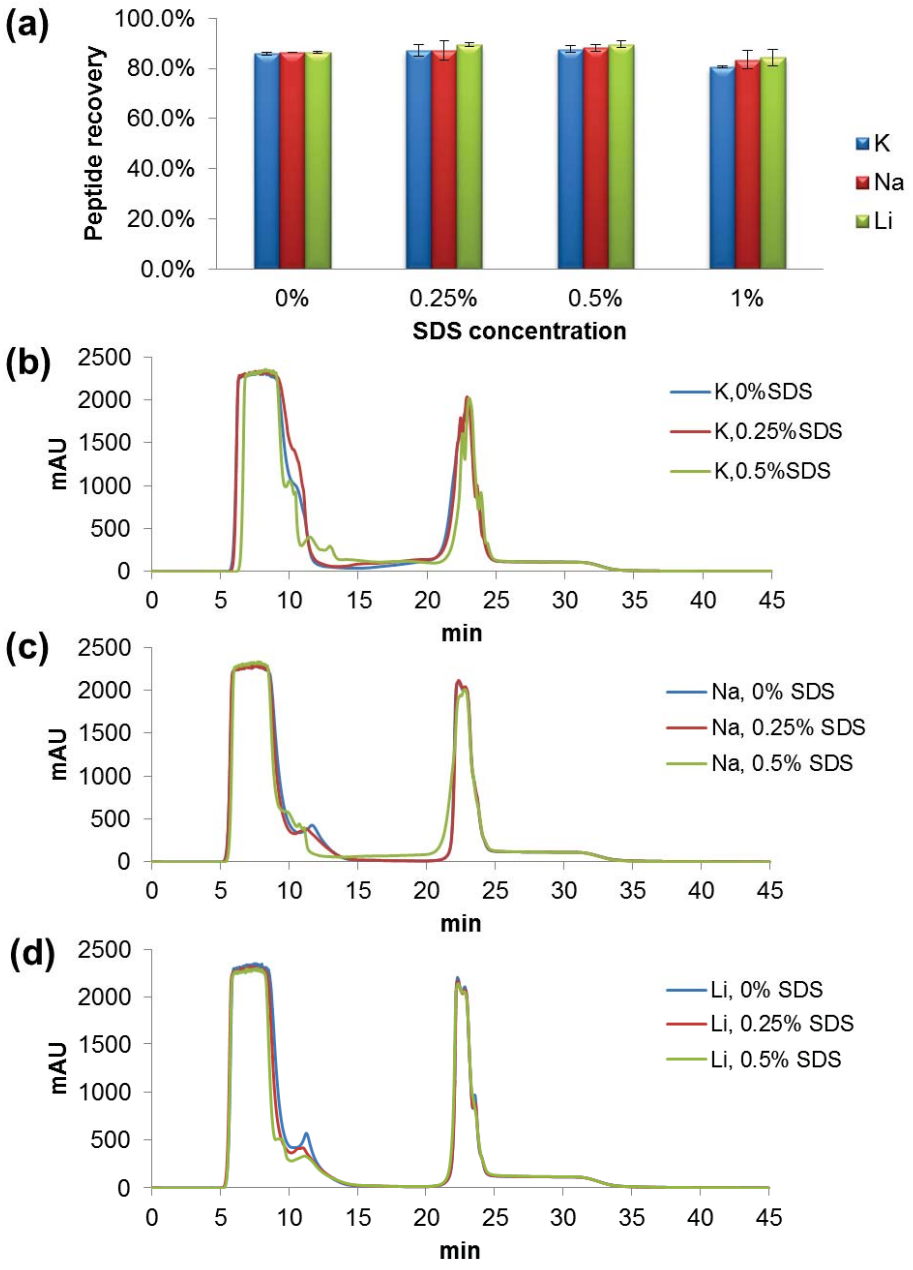


Figure 4-3. Comparison of (a) the peptide recovery rates after SCX of BSA digests containing different SDS concentrations with different mobile phase A (n=3) and (b-d) SCX-UV chromatograms obtained using different mobile phase A and different SDS concentrations. Error bars represent one standard deviation.

peptide recovery. The elution chromatograms with a step gradient as described in the Experimental are shown in Figure 4-3 (b-d). The chromatograms obtained from the Li^+ mobile phase do not vary as much as those of K^+ or Na^+. In the cases of K^+ and Na^+, the eluates collected in between the first main peak from the weakly retained compounds (retention time of ~5-10 min) and the second main peak from peptides (retention time of ~20-25 min) were found to contain some peptides, while no peptides were detected when Li^+ was used. As Figure 4-3 (a) shows, the recovery rate using Li^+ mobile phase is close to 90% for samples containing 0.25% or 0.50% SDS. Interestingly, the recovery rate is higher than that observed from the peptide samples containing no SDS. The presence of micelles seems to promote the binding of peptides to the stationary phase and reduce the elution of peptides during the washing step. This is evident in Figure 4-3 (d) where a small peak eluting at about 11.5 min increases as the sample solution changes from 0.5% SDS to 0% SDS.

Figure 4-3 (a) also illustrates that the peptide recovery rate is reduced to about 84% when the SDS concentration is 1%. This indicates that a very high concentration of SDS is not desirable, perhaps due to the increase in the size of SDS micelles[34] which can help form more SDS-peptide complexes. As was pointed out earlier, the SDS-peptide complexes can be eluted during the washing step, resulting in the loss of peptides.

4.3.4 Effects of sample composition

The above results demonstrate that, using SCX, the sample recovery rate of ~90% can be achieved while effectively removing SDS, as long as the SDS concentration in a peptide sample is kept at 0.25% or 0.5%. However, sample composition can play an important role in determining the overall sample recovery rate. Specifically, any components present in the sample that can influence the binding of peptides to the SCX column may affect the sample recovery rate. In a typical protein or proteome digest, salts are always present at relatively high concentrations, particularly when a buffer solution is used to control the pH of the solution during the sample workup and the enzymatic digestion process. Cations other than the peptide ions in the sample solution would compete for binding to the stationary phase. We found that the peptide recovery rate was 89.1±1.0% (n=3) from a BSA digest containing 0.5% SDS and 50 mM NH_4HCO_3 and the recovery rate was 90.6±1.0% (n=3) after the removal of the NH_4HCO_3. While the presence of a high concentration of NH_4HCO_3 did not affect the peptide recovery rate significantly, interference was observed in some other cases (see below). We examined three methods to remove NH_4HCO_3, namely using a SpeedVac at room temperature, SpeedVac with heating to 65 °C, and heating the sample directly to 85°C without the use of a SpeedVac. It was found that the use of a heater commonly used for incubating the sample in

enzymatic digestion can remove NH_4HCO_3 efficiently. Thus, a simple heating method can be used to overcome the potential interference of NH_4HCO_3 in SDS removal by SCX.

In addition to NH_4HCO_3, DTT and IAA are commonly used for reducing disulfide bonds and alkylating the cysteines before protein digestion by enzymes such as trypsin. The amount of DTT and IAA present in the peptide solution can have a significant effect on sample recovery. For example, when the DTT and IAA concentrations were 15 mM and 35 mM, respectively, the sample recovery rate was found to be 77.2%±0.5% (n=3) after the samples containing DTT, IAA and 0.5% SDS were subjected to SDS removal by SCX. When the DTT and IAA concentrations were reduced to 5 mM and 12 mM, respectively, which were still in excess and should not have resulted in degradation of the performance of tryptic digestion of proteins, the sample recovery rate increased to 90.6% ± 1.0% (n=3). Note that, in the case of higher concentrations of DTT and IAA (i.e., 15 mM and 35 mM, respectively), the recovery rate was 75.5±0.1% (n=3) if NH_4HCO_3 was not removed by heating. Removing NH_4HCO_3 increased the sample recovery rate to 77.2%±0.5% (n=3), indicating that, in this case, removal of NH_4HCO_3 was beneficial.

The above results and discussion indicate that the use of SCX can be very effective to remove SDS if we do an appropriate control of both the SCX and the sample solution conditions to minimize sample loss.

4.3.5 Application to SDS-assisted shotgun proteome analysis

As Figure 4-1 shows, SDS-assisted shotgun proteome analysis involves the use of SDS to dissolve proteins, followed by enzyme digestion and LC-MS/MS analysis of the resultant peptides. To demonstrate the performance of the SCX method for SDS removal in this workflow, a series of experiments using two different protein samples were carried out.

In the first example, the membrane protein fraction of the *E. coli* cell lysate was prepared as described in the Experimental. This membrane fraction was subjected to SDS-assisted protein solubilization and trypsin digestion. In brief, the protein pellets were dissolved in 1% SDS, followed by dilution of the solution to 0.025% SDS for protein reduction in 5 mM DTT, alkylation in 12 mM IAA, and digestion by trypsin (50:1 protein: enzyme). The digest was dried down by heating at 85 °C. Varying amounts of SDS were added to the digest to form a final solution containing 0.025% (i.e., no SDS was added), 0.05%, 0.1%, 0.25%, 0.5%, or 1% SDS. The experiment was done in triplicates. Each digest sample was injected onto the SCX column where mobile phase A was 10 mM LiH_2PO_4. After washing the column with mobile phase A to remove SDS, the peptides were

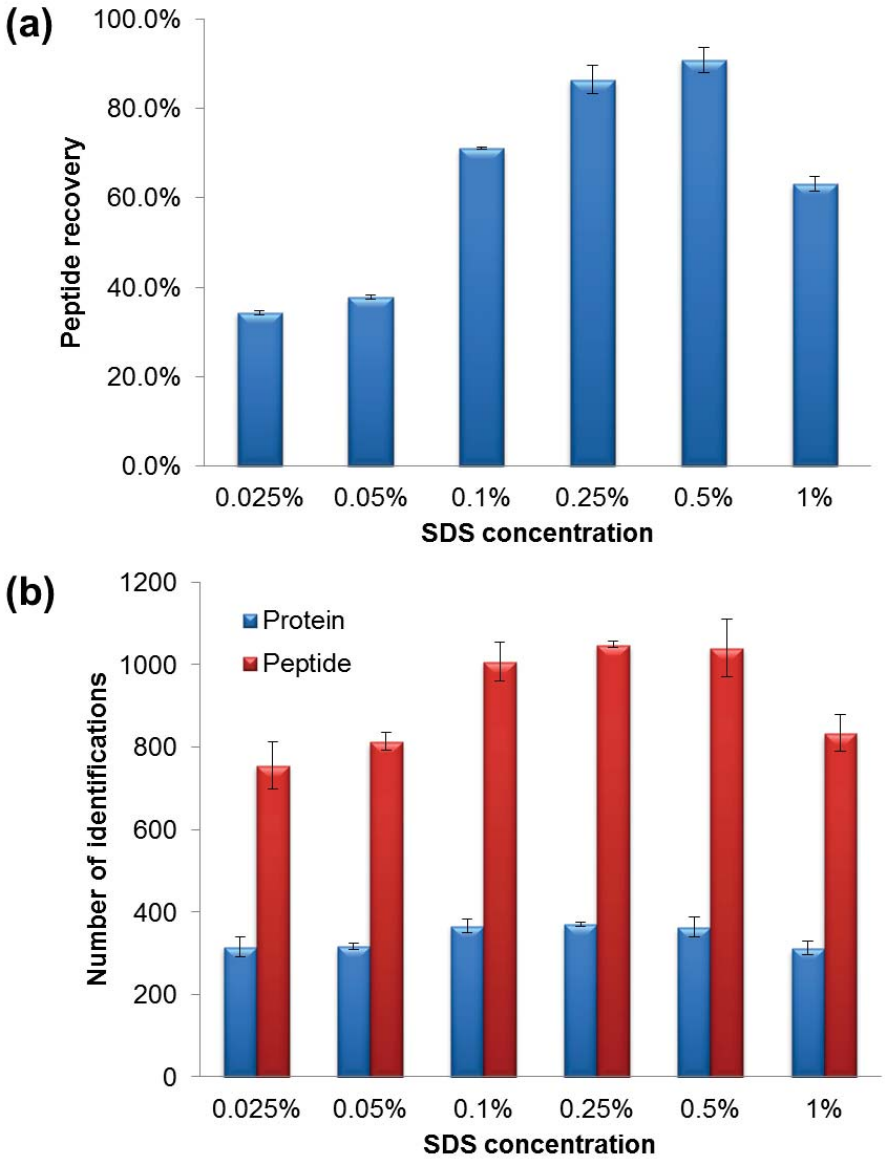


Figure 4-4. Comparison of (a) the peptide recovery rates and (b) the number of identified peptides and proteins after SDS removal by SCX of the E. coli membrane protein digests containing different concentrations of SDS (n=3). Error bars represent one standard deviation.

eluted using mobile phase B of 10 mM KH_2PO_4 and 0.5 M KCl. The peptides were collected and then subjected to desalting and quantification by LC-UV. The results obtained are shown in Figure 4-4 (a) where the sample recovery rate is plotted as a function of the SDS concentration in the final digest solution. It is clear that high sample recovery rate are obtained from the 0.5% SDS solution, which is consistent with the BSA results. Figure 4-4 (b) plots the number of unique peptides and proteins identified from the digests with varying concentrations of SDS. There is a good correlation between the sample recovery rate and the number of peptides and proteins identified. Note that the same amount of peptides (i.e., an optimal amount of 1.5 μg) was injected into LC-MS/MS for sequencing and, thus, the peptide number difference reflects the composition difference. As in the case of BSA, sample loss was selective, i.e., some peptides were preferentially lost during the SDS removal process.

In the second example, the membrane protein fractions of the MCF-7 cell lysates were subjected to SDS-assisted protein solubilization and digestion using the same protocol as described above for the *E. coli* sample. The results of sample recovery rate and the numbers of peptides and proteins identified are shown in Figure 4-5. The same conclusions as those in the case of *E. coli* can be drawn from the analysis of this MCF-7 sample. However, the difference in the number of peptides and proteins identified is not as dramatic. This is likely due to the sample

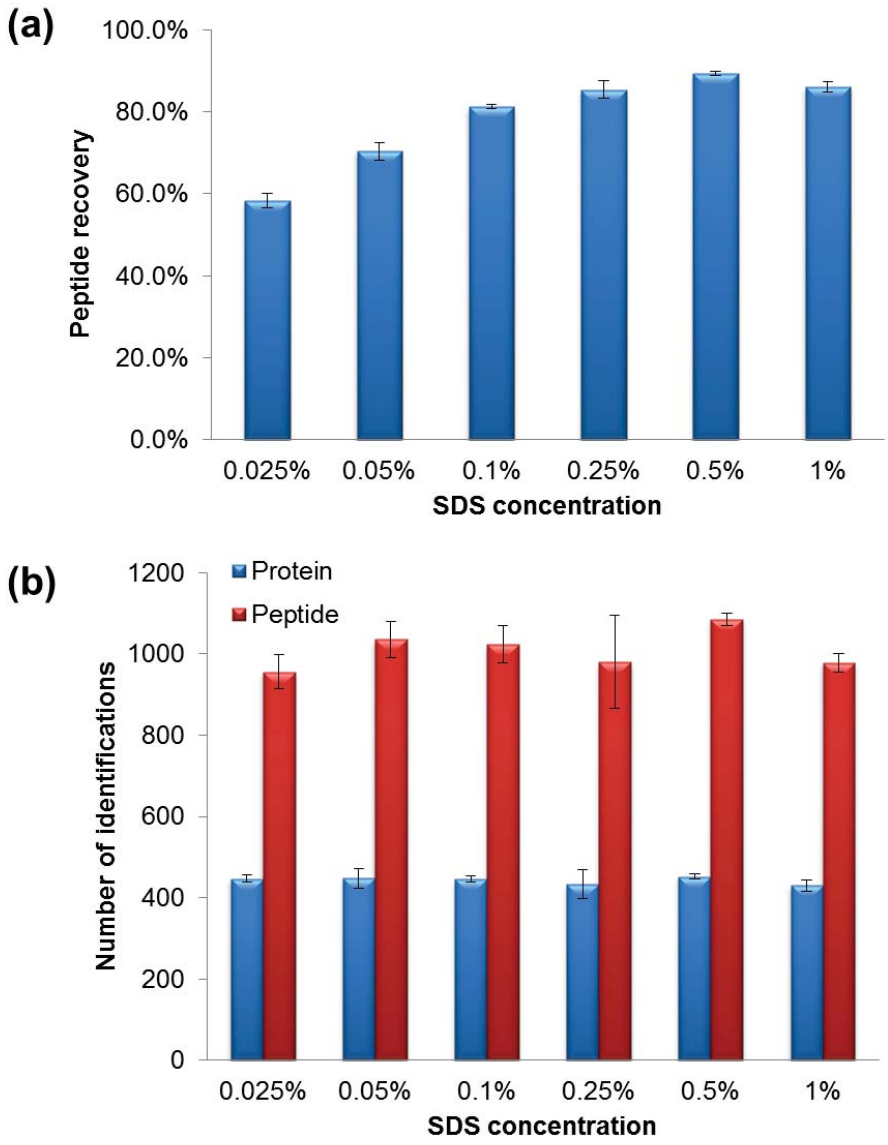


Figure 4-5. Comparison of (a) the peptide recovery rates and (b) the numbers of identified peptides and proteins after SDS removal by SCX of the MCF-7 membrane protein digests containing different concentrations of SDS concentrations after SCX (n=3). Error bars represent one standard deviation.

complexity difference: the MCF-7 digest should contain far more peptides than the *E. coli* digest. As a consequence, sample loss in the MCF-7 digest does not alter the peptide composition as significantly as that of the *E. coli* digest. In other words, LC-MS/MS only identifies a fraction of the peptides within a finite concentration dynamic range. There are many more peptides in the MCF-7 digest within the detectable range and selective loss of some peptides does not reduce the overall number of different peptides as dramatically as in the case of the *E. coli* sample.

4.3.6 Comparison with acid labile surfactants

Several types of SDS-mimic surfactants have been developed for proteome analysis. Among them, acid labile surfactants have been used in a number of proteomics studies.[8-16] Some, such as sodium-3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)-methoxyl]-1-propanesulfonate or acid labile surfactant (RapiGest), require an extra step to remove the hydrophobic product after acid hydrolysis, while others, such as 3-[3-(1,1-bisalkyloxyethyl)pyridin-1-yl]propane-1-sulfonate or PPS, hydrolyze into products that need not be removed. We have compared the performance of the SDS-assisted shotgun method with the RapiGest and PPS methods.

In our experiments, the *E. coli* membrane protein fraction was solubilized by RapiGest or PPS, followed by trypsin digestion. After degrading the

surfactants, three equal aliquots of the digests were taken for desalting and peptide quantification by LC-UV. The desalted peptides were analyzed by RPLC MS/MS. Therefore, there was no sample loss in the SCX wash step. However, for the SDS sample, we only analyzed the peptide-fraction; any peptides contained in the wash-fraction were not sequenced, a potential loss in proteome coverage if these peptides were not present in the peptide-fraction.

Table 4-1 shows the number of peptides and proteins identified from the RapiGest and PPS samples compared to the SDS sample (0.5% SDS). In the RapiGest and PPS samples, an average of 1256±38 (n=3) and 909±4 (n=3) unique peptides from 389±22 and 303±1 proteins were found, respectively. For the SDS sample, only the peptide-fraction could be analyzed by LC MS/MS, resulting in the identification of an average of 1041±69 peptides and 363±24 proteins from three replicate runs. These data indicate that the RapiGest method identified more peptides (~21% more) and proteins (~7% more) than the SDS method and both the RapiGest and SDS methods performed better than the PPS method. The performance difference observed is likely due to the differences in protein solubilization, digestion, and/or downstream peptide sample workup in the two workflows.

We have also compared the performance of the three methods for the analysis of the MCF-7 membrane protein fraction and the results are shown in

Table 4-1. Summary of the number of peptides and proteins identified by LC-QTOF-MS/MS from three replicate sample preparations of the E. coli membrane protein fraction using RapiGest, PPS or SDS.

Sample preparation	Number of peptides identified				Number of proteins identified			
	Run #1	Run #2	Run #3	Average	Run #1	Run #2	Run #3	Average
RapiGest	1275	1212	1280	1256±38	400	364	403	389±22
PPS	911	911	904	909±4	302	304	304	303±1
SDS	1119	1015	988	1041±69	390	356	343	363±24

Table 4-2. An average of 1036±31 (n=3) and 984±32 (n=3) peptides from 389±15 and 365±6 proteins were identified from the RapiGest and PPS samples, respectively. For the SDS sample, an average of 1085±16 peptides and 452±6 proteins were identified. In this case, the SDS method identified more peptides (~5%) and proteins (~24%) than the RapiGest method. Both the SDS and RapiGest methods identified more peptides and proteins than the PPS method.

The distributions of the numbers of peptides and proteins identified among the three replicates within the same method and the common peptides and proteins identified among the three methods are analyzed for the *E. coli* sample and the MCF-7 sample. For the *E. coli* sample, 1607 unique peptides from 463 unique proteins, 1100 peptides from 365 proteins, and 1384 peptides from 453 proteins were identified from the three replicates combined using the RapiGest, PPS and SDS method, respectively. The three methods combined identified a total of 2432 unique peptides from 637 unique proteins. Thus, in terms of the number of proteins identified from each method, the RapiGest method identified 73% (i.e., 463 out of 637) of the total number of proteins, the PPS method identified 57% (365 out of 637) and the SDS method identified 71% (453 out of 637). There were 236 common proteins identified from the three methods, representing 37% of the total number of proteins. For the MCF-7 sample, a total of 2545 unique peptides from 880 unique proteins were identified from the three methods combined.

Table 4-2. Summary of the number of peptides and proteins identified by LC-QTOF-MS/MS from three replicate sample preparations of the MCF-7 membrane protein fraction using RapiGest, PPS or SDS.

Sample preparation	Number of peptides identified				Number of proteins identified			
	Run #1	Run #2	Run #3	Average	Run #1	Run #2	Run #3	Average
RapiGest	1072	1018	1018	1036±31	406	382	379	389±15
PPS	1015	985	951	984±32	371	360	363	365±6
SDS	1074	1104	1078	1085±16	457	454	445	452±6

Among them, 1343 peptides from 492 proteins, 1272 peptides from 444 proteins, and 1490 peptides from 586 proteins were identified from the RapiGest, PPS and SDS method, respectively. Thus, the RapiGest method identified 56% (491 out of 880 proteins) of the total number of proteins, the PPS method identified 50% (444 out of 880) and the SDS method identified 67% (586 out of 880). There were 206 common proteins identified from the three methods, representing 23% of the total number of proteins. The smaller percentage of common proteins identified from the MCF-7 sample, compared to the *E. coli* sample, also correlates with the complexity of the sample. The MCF-7 digest contained more diverse peptides within the detectable concentration range than the *E. coli* digest and thus more random identifications were found from the MCF-7 sample.

4.3.7 Integration with peptide separation

The use of SCX for removing SDS can be integrated with peptide separation to form a fully automated procedure as the first dimension separation in the offline 2D-LC MS/MS workflow. Offline 2D-LC offers the possibility of maximizing the peptide sample loading to reversed-phase LC for MS/MS sequencing.[25] The amount injected to RPLC MS/MS is optimized beforehand; for example, in our current capillary LC QTOF setup, the optimal amount was 1.5 μg. Injection of lower than the optimal amount results in the identification of a smaller number of peptides and proteins by MS/MS. On the other hand, injecting

too much sample into RPLC MS/MS can cause a sample carryover problem, in addition to signal saturation. In offline 2D-LC, after collecting the SCX fractions, an RPLC-UV setup can be used to remove the salts and determine the peptide amounts in the individual fractions using an auto-sample-injector and a fraction collector. For an SCX fraction containing greater than 1.5 μg of peptides, 1.5 μg is injected into the RPLC MS/MS for sequencing. For the fractions containing less than 1.5 μg of peptides, adjacent fractions can be combined to reach the optimal amount for injection. This maximized sample loading strategy ensures peptide sequencing is done at the maximum capacity. Moreover, a rolling precursor ion exclusion (PIE) strategy[27] can be applied to the offline 2D-LC MS/MS to increase further the peptide and protein identification efficiency. In rolling PIE, the peptides already identified in previous SCX fractions (2 fractions) are excluded for sequencing in analyzing the new SCX fraction. Thus, more new peptides are sequenced at a given chromatographic window, compared to that without PIE where the same peptides present in multiple fractions would be sequenced multiple times, wasting the valuable instrument time. Finally, with offline 2D-LC, organic or other solvents not readily compatible with RPLC MS/MS can be used for SCX.

In handling the SDS-containing samples, such as the digest of the MCF-7 membrane protein fraction, a washing step (10 min using mobile phase A) was

added to the chromatographic elution program and the rest of the separation gradient profile was the same as that used in the conventional SCX chromatography. The large early-elution peak in the SCX-UV chromatogram, similar to those shown in Figure 4-2, was from the un-retained or weakly retained species. During the salt-gradient peptide elution between 20 and 80 min, we collected 60 fractions from the MCF-7 sample. After desalting and peptide quantification, some fractions were combined to produce 25 fractions that were subjected to RPLC MS/MS analysis. Two replicate experiments were carried out.

For the MCF-7 membrane protein fraction, 5764 unique peptides from 1988 unique proteins or protein groups were detected in the first experiment with a false peptide discovery rate of 0.76%. In the replicate experiment, 5924 peptides and 2011 proteins were detected with a false peptide discovery rate of 0.80%. In total, 6889 unique peptides corresponding to 2258 unique proteins were identified. The overlap of peptides was 70% and the overlap of proteins was 77%, illustrating that good reproducibility from replicate experiments could be achieved.

The MCF-7 membrane protein fraction was also subjected to the Rapidest sample preparation, followed by 2D-LC MS/MS shotgun analysis, in the same manner as the SDS method except that acidification of the tryptic digest was performed prior to SCX-LC fractionation. In the first experiment, 4173 unique peptides from 1588 unique proteins or protein groups were detected. In the

replicate experiment, 4654 peptides and 1670 proteins were identified. In total, 5172 unique peptides from 1847 unique proteins were identified and the peptide overlap was 71% and the protein overlap was 75%.

For this MCF-7 membrane protein fraction sample, it is clear that the SDS method identified more peptides and proteins than the RapiGest method, which is consistent with the 1D-LC MS/MS comparative results shown in Table 4-2. Combining the results obtained by the RapiGest and SDS methods, a total of 2708 unique proteins were identified. Among them, 450 unique proteins were found from the RapiGest method and 861 unique ones from the SDS method with 1397 common proteins found in both methods. This example demonstrates that SDS removal can be integrated with peptide separation in the SCX chromatographic setup for offline 2D-LC MS/MS. Since no extra step is required, this method provides a convenient way of handling the SDS-containing samples.

4.4 Conclusions

An integrated strong-cation exchange liquid chromatographic procedure for SDS removal and peptide separation was developed for SDS-assisted shotgun proteome analysis. The peptide sample recovery rate was found to be about 90% for the digests of BSA and membrane-protein-enriched fractions of the cell lysates of *E. coli* and MCF-7 breast cancer cell line. To achieve this high recovery rate, SDS was purposely added, after trypsin digestion, to reach a final concentration of 0.5% in the digest. It was found that very high concentrations of DTT and IAA in the digested peptide solution could cause severe sample loss during the SDS removal process; the use of 5 mM DTT and 12 mM IAA is recommended. Removal of NH_4HCO_3 from the peptide sample by simple heating did not affect the recovery rate significantly. Using 10 mM LiH_2PO_4, instead of KH_2PO_4 or NaH_2PO_4, as mobile phase A in SCX to wash off the SDS did not affect the recovery rate either. Compared to other methods of using acid-labile surfactants, namely RapiGest and PPS, for the MCF-7 membrane fraction sample, the SDS method identified more peptides (~5%) and proteins (~16%) than the RapiGest method, while the RapiGest method identified more peptides (~21%) and proteins (~7%) from the *E. coli* membrane fraction than the SDS method. In both cases, the two methods identified more peptides and proteins than the PPS method. However, SDS is much cheaper than the acid-labile surfactants. In addition, since

SCX is widely used as the first dimension of 2D-LC MS/MS in a shotgun proteome analysis workflow, integration of SDS removal with peptide separation in SCX will not add any extra steps to the sample handling process. Thus, the cost difference between the SDS method and the acid-labile surfactant method in 2D-LC is mainly from the difference in the cost of surfactant, which is dependent on the surfactant amount used, i.e., tens to hundreds of dollars per experiment vs. a few cents when SDS is used. However, if 1D-LC is used, the need of using SCX to remove SDS adds an additional step and cost. Finally, we demonstrated the application of this method for generating a proteome profile of the MCF-7 membrane protein fraction using off-line 2D-LC QTOF MS/MS; in total, 6889 unique peptides corresponding to 2258 unique proteins or protein groups were identified from two replicate experiments (25 fractions each) with a false peptide discovery rate of ~0.8%. In comparison, 5172 unique peptides from 1847 unique proteins were identified by the RapiGest method. To our knowledge, this is the first report of using SCX to remove SDS while keeping a high peptide recovery rate (~90%). This method provides a convenient and inexpensive means of handling proteome samples that require SDS for protein extraction, solubilization, or digestion in the 2D-LC MS/MS shotgun proteome analysis workflow.

4.5 Literature cited

1. Lu, B. W.; McClatchy, D. B.; Kim, J. Y.; Yates, J. R., Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry. *Proteomics* **2008,** *8* (19), 3947-3955.

2. Zhang, N.; Li, L., Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid Commun. Mass Spectrom.* **2004,** *18* (8), 889-896.

3. Zhang, N.; Chen, R.; Young, N.; Wishart, D.; Winter, P.; Weiner, J. H.; Li, L., Comparison of SDS- and methanol-assisted protein solubilization and digestion methods for Escherichia coli membrane proteome analysis by 2-D LC-MS/MS. *Proteomics* **2007,** *7* (4), 484-493.

4. Bosserhoff, A.; Wallach, J.; Frank, R. W., Micropreparative separation of peptides derived from sodium dodecyl sulfate-solubilized proteins. *J. Chromatogr.* **1989,** *473* (1), 71-77.

5. Kawasaki, H.; Suzuki, K., Separation of peptides dissolved in a sodium dodecyl-sulfate solution by reversed-phase liquid-chromatography - removal of sodium dodecyl-sulfate from peptides using an ion-exchange precolumn. *Anal. Biochem.* **1990,** *186* (2), 264-268.

6. Loo, R. R.; Dales, N.; Andrews, P. C., Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Sci.* **1994,** *3* (11), 1975-83.

7. Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A., Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J. Proteome Res.* **2010,** *9* (6), 2863-2870.

8. Ross, A. R. S.; Lee, P. J.; Smith, D. L.; Langridge, J. I.; Whetton, A. D.; Gaskell, S. J., Identification of proteins from two-dimensional polyacrylamide gels using a novel acid-labile surfactant. *Proteomics* **2002,** *2* (7), 928-936.

9. Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C., Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **2003,** *75* (21), 6023-6028.

10. Arnold, R. J.; Hrncirova, P.; Annaiah, K.; Novotny, M. V., Fast proteolytic digestion coupled with organelle enrichment for proteomic analysis of rat liver. *J. Proteome Res.* **2004,** *3* (3), 653-657.

11. Ruth, M. C.; Old, W. M.; Emrick, M. A.; Meyer-Arendt, K.; Aveline-Wolf, L. D.; Pierce, K. G.; Mendoza, A. M.; Sevinsky, J. R.; Hamady, M.; Knight, R. D.; Resing, K. A.; Ahn, N. G., Analysis of membrane proteins from human chronic myelogenous leukemia cells: Comparison of extraction methods for multidimensional LC-MS/MS. *J. Proteome Res.* **2006,** *5* (3), 709-719.

12. Chen, E. I.; McClatchy, D.; Park, S. K.; Yates, J. R., Comparisons of mass spectrometry compatible surfactants for global analysis of the mammalian brain proteome. *Anal. Chem.* **2008,** *80* (22), 8694-8701.

13. Mbeunkui, F.; Goshe, M. B., Investigation of solubilization and digestion methods for microsomal membrane proteome analysis using data-independent LC-MS(E). *Proteomics* **2011,** *11* (5), 898-911.

14. Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **2007,** *6* (7), 2529-2538.

15. Norrgran, J.; Williams, T. L.; Woolfitt, A. R.; Solano, M. I.; Pirkle, J. L.; Barr, J. R., Optimization of digestion parameters for protein quantification. *Anal. Biochem.* **2009,** *393* (1), 48-55.

16. Wu, F.; Sun, D. F.; Wang, N.; Gong, Y.; Li, L., Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis. *Anal. Chim. Acta* **2011,** *698* (1-2), 36-43.

17. Smith, J. A., Electroelution of proteins from stained gels. *Current Protocols in Immunology / edited by John E. Coligan* **2002,** *Chapter 8*, Unit 8.8.

18. Wisniewski, J. R.; Mann, M., Spin filter-based sample preparation for shotgun proteomics. *Nat. Methods* **2009,** *6* (11), 785-786.

19. Hixson, K. K.; Rodriguez, N.; Camp, D. G.; Strittmatter, E. F.; Lipton, M. S.; Smith, R. D., Evaluation of enzymatic digestion and liquid chromatography-mass spectrometry peptide mapping of the integral membrane protein bacteriorhodopsin. *Electrophoresis* **2002,** *23* (18), 3224-3232.

20. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* **2009,** *6* (5), 359-U60.

21. Vissers, J. P.; Chervet, J. P.; Salzmann, J. P., Sodium dodecyl sulphate removal from tryptic digest samples for on-line capillary liquid chromatography/electrospray mass spectrometry. *J. Mass Spectrom.* **1996,** *31* (9), 1021-7.

22. Timperman, A. T.; Aebersold, R., Peptide electroextraction for direct coupling of in-gel digests with capillary LC-MS/MS for protein identification and sequencing. *Anal. Chem.* **2000,** *72* (17), 4115-4121.

23. Antharavally, B. S.; Mallia, K. A.; Rosenblatt, M. M.; Salunkhe, A. M.; Rogers, J. C.; Haney, P.; Haghdoost, N., Efficient removal of detergents from proteins and peptides in a spin column format. *Anal. Biochem.* **2011,** *416* (1), 39-44.

24. Antharavally, B. S.; Rosenblatt, M. M.; Mallia, K. A.; Rogers, J. C.; Haney, P. J., A high yield method for the removal of detergents from low-concentration protein or peptide samples for mass spectrometry analysis. *in Proceedings of the American Society for Mass Spectrometry Conference, June, 2011, Denver, CO.*

25. Wang, N.; Xie, C. H.; Young, J. B.; Li, L., Off-line two-dimensional liquid chromatography with maximized sample loading to reversed-phase liquid chromatography-electrospray ionization tandem mass spectrometry for shotgun proteome analysis. *Anal. Chem.* **2009,** *81* (3), 1049-1060.

26. Campbell, R.; Winkler, M. A.; Wu, H. Q., Quantification of sodium dodecyl sulfate in microliter biochemical samples by gas chromatography. *Anal. Biochem.* **2004,** *335* (1), 98-102.

27. Wang, N.; Li, L., Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis. *Anal. Chem.* **2008,** *80* (12), 4696-4710.

28. Wang, N.; MacKenzie, L.; De Souza, A. G.; Zhong, H. Y.; Goss, G.; Li, L., Proteome profile of cytosolic component of zebrafish liver generated by LC-ESI

MS/MS combined with trypsin digestion and microwave-assisted acid hydrolysis. *J. Proteome Res.* **2007,** *6* (1), 263-272.

29. Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2005,** *2* (9), 667-675.

30. Jacquier, J. C.; Desbene, P. L., Determination of critical micelle concentration by capillary electrophoresis. *J. Chromatogr. A* **1996,** *743* (2), 307-314.

31. Tummala, R.; Limbach, P. A., Effect of sodium dodecyl sulfate micelles on peptide mass fingerprinting by matrix-assisted laser desorption/ ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004,** *18* (18), 2031-2035.

32. Tummala, R.; Limbach, P. A., Serum protein profiling using surfactant-aided matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chim. Acta* **2005,** *551* (1-2), 137-141.

33. Malmquist, G.; Lundell, N., Characterization of the influence of displacing salts on retention in gradient elution ion-exchange chromatography of proteins and peptides. *J. Chromatogr.* **1992,** *627* (1-2), 107-124.

34. Quina, F. H.; Nassar, P. M.; Bonilha, J. B. S.; Bales, B. L., Growth of sodium dodecyl-sulfate micelles with detergent concentration. *J. Phys. Chem.* **1995,** *99* (46), 17028-17031.

Chapter 5 In-gel Microwave-assisted Acid Hydrolysis of Proteins Combined with Liquid Chromatography Tandem Mass Spectrometry for Mapping Protein Sequences

A version of this chapter was published as Sun, D. F.; Wang, N.; Li, L., In-gel microwave-assisted acid hydrolysis of proteins combined with liquid chromatography tandem mass spectrometry for mapping protein sequences. *Anal. Chem.* 2014, *86* (1), 600-607.

5.1 Introduction

Protein sequence mapping is important for unambiguous identification of proteins or protein isoforms,[1,2] study of protein modifications[3,4] and discovery of amino acid substitutions from point mutations in the genome.[5] For these applications, high sequence coverage is often required. Mass spectrometry (MS) can produce a sequence map using the top-down and bottom-up methods.[6] The top-down method uses a tandem MS to dissociate a protein ion directly and offers the opportunity of examining a protein sequence quickly.[7] However, obtaining high sequence coverage (e.g., >80%) for a protein with molecular mass of higher than 25 kD can be a challenge.[8-10] The bottom-up method usually requires multiple enzymes with different specificities for protein degradation, which can increase the sequence coverage, but not necessarily generate 100% coverage.[11-13] A combination of the top-down and bottom-up methods is the middle-down approach that uses an enzyme or chemical to degrade a protein into several large peptides that are subsequently sequenced by the top-down method.[14,15] The sequence coverage of this approach depends on the availability of a suitable enzyme or chemical to degrade a protein into multiple peptides with their sizes suitable for whole sequence mapping using the top-down method.

Microwave-assisted acid hydrolysis (MAAH) MS is another method to generate high protein sequence coverage.[16,17] Using 25% trifluoroacetic acid (TFA) and microwave irradiation time of 8-10 min, a protein can be degraded into many small peptides (< 3000 Da). These peptides are ideally suited for tandem MS sequencing using a conventional and widely available mass spectrometer commonly used for bottom-up proteome analysis. In a recent work, we demonstrated the possibility of obtaining 100% sequence coverage for bovine serum albumin (BSA) (~67 kDa), which allowed the identification of an amino acid substitution in the sequence[18]. Although this method is powerful for whole sequence analysis, it requires a relatively pure protein (e.g., >80% purity)[18] in order to produce peptides of the dominant protein to read the whole sequence. To purify a protein, solution-based separation techniques (i.e., liquid chromatography) may be used, but not efficient from a complex protein sample. On the other hand, gel electrophoresis is widely used to separate protein mixtures and, with multidimensional gel electrophoresis, high resolution separation of proteins, including modified proteins such as phosphoproteins with varying phosphorylation sites, can be achieved[19]. In addition, gel electrophoresis can be used to remove salts, buffers, detergents and other impurities from a protein sample efficiently that may interfere with MS analysis. Gel electrophoresis can also be used to enrich low abundance proteins, as the sample loading capacity in a

large gel can be very high (i.e., several mg of total proteins). Gel electrophoresis is often a method of choice for protein separation in biological research laboratories, as it is compatible to many downstream biochemical analysis methods, such as Western blot, for protein or cellular function studies. Indeed, gel electrophoresis of proteins, followed by in-gel digestion of the separated proteins and MS analysis of the resultant peptides, has been widely used in proteomics.[20,21] However, this method usually does not offer a high sequence coverage of a protein.

In this work, we describe an in-gel MAAH MS method that combines the separation power of gel electrophoresis with the capability of high sequence coverage of MAAH MS for improved protein sequence mapping. The experimental workflow along with the optimization and effects of the key individual steps on the overall performance of the method is presented. This method is applied to the sequence mapping of various human plasma proteins separated by gel electrophoresis with demonstrated high sequence coverage. In addition, characterization of the phosphoprotein isoforms of bovine alpha-S1-casein separated by two dimensional (2D) gel electrophoresis was performed to demonstrate the applicability of the method for the detailed analysis of protein isoforms with varying modification sites.

5.2 Experimental section

5.2.1 Chemicals and reagents.

Dithiothreitol (DTT), tris, bromophenol blue, 2-mercaptoethanol, sodium dodecyl sulfate (SDS), bio-safe Coomassie G-250 stain and 2D starter kit were from Bio-Rad Laboratories (Mississauga, ON, Canada). LC-MS grade water, LC-MS grade acetonitrile (ACN), acetone, glycine, acetic acid and methanol (MeOH) were from Fisher Scientific Canada (Edmonton, AB, Canada). The BCA assay kit was from Pierce (Rockford, IL). Glycerol was from USB corporation (Cleveland, OH, USA). All the other chemicals were purchased from Sigma-Aldrich Canada (Markham, ON, Canada).

5.2.2 Human plasma preparation

Four mL of blood were collected from an healthy individual with the ethics approval from the University of Alberta and stored in a tube with anticoagulant reagent (EDTA). The red blood cells and the plasma were separated into two layers by centrifuging the sample at 3500 rpm (ROTOFIX 32) for 15 min. The plasma was aliquoted into 1.5 ml microcentrifudge tubes (100 μL per tube) and stored in -80°C.

5.2.3 Depletion of albumin in plasma

The EMD biosciences ProteoExtract albumin removal kit (Cat. No. 122640) was used for removal of albumin from the human plasma. To the 100 μL plasma, 5 μL of dilution buffer was added. The solution was aliquoted into three tubes (35 μL per tube). In each tube, the plasma was diluted by 10-fold with the binding buffer. The storage buffer was removed from the column and 1 mL of the binding buffer was allowed to pass the resin bed by gravity-flow. Each diluted sample was passed one column by gravity flow. The flow-through solution was collected. Six hundred μL of the binding buffer was used to wash the column and the washing fraction was collected. Column washing was repeated once. The flow-through solution and two washing fractions were combined. This combined solution contained the albumin-depleted plasma proteins. Acetone with four times of the sample volume was added and the sample was incubated at -20°C overnight. A protein pellet was collected by centrifuging the sample at 20,817 *g* for 30 min at 4°C. BCA assay was used to determine the protein solution concentration.

5.2.4 SDS-PAGE

Protein samples were separated on a 12% polyacrylamide gel using a BioRad mini-gel electrophoresis apparatus (Mississauga, ON, Canada). A constant voltage of 200 V was used for electrophoresis. After separation and

fixation, the gel was stained with bio-safe Coomassie G-250 stain and then destained with water to produce clear gel bands.

5.2.5 2D gel electrophoresis

Bovine alpha casein was first separated in the 17-cm IPG strip (pH 3.9-5.1) by IEF. The second dimension separation was performed on a 12% polyacrylamide gel using a BioRad PROTEAN Ⅱ XL apparatus (Mississauga, ON, Canada). After fixation, the gel was stained with bio-safe Coomassie G-250 stain and then destained with water to produce clear gel bands.

5.2.6 In-gel microwave-assisted acid hydrolysis

The gel band of BSA, human plasma proteins, bovine alpha casein phosphoprotein isoforms or a blank gel band was cut out and washed with 1 mL of water for 5 min twice. Then it was cut into small pieces and dehydrated by using 500 μL of ACN for 5 min twice. The dried gel pieces were desiccated using Speedvac for 15 min. To the gel pieces, 120 μL of 12 mM DTT and 40 μL of TFA (100%) were added; the TFA concentration in the hydrolysis solution was 25%. The sample vial was placed in a water bath that was placed in a household 1200 W (2450 MHz) microwave oven, according to the setup reported previously[18]. For the study of the effects of the microwave irradiation time on

MAAH, the time used was 6, 8, 10, or 12 min. The peptides from MAAH were extracted by using 300 µL of 85% ACN / 0.1% TFA solution twice, and dried down using Speedvac. To the peptide sample, 200 µL of 1 M ammonium bicarbonate was added to adjust the solution pH to ~8. Reduction and alkylation of the peptide mixture was performed by using 20 µL of DTT (90 mM) for 1 hr at 37 °C and 50 µL of IAA (200 mM) for 1 hr at room temperature in dark, respectively. Finally, 15 µL of TFA was added to adjust the pH to ~1.5 and the sample was ready for desalting and LC-MS analysis.

5.2.7 Affinity enrichment of phosphopeptides

For the analysis of the bovine alpha casein sample, the hydrolysate was subjected to immobilized metal ion affinity chromatography (IMAC) phosphopeptide enrichment using Fe-IMAC resin (Phos-Select iron affinity gel; Sigma, Ontario, Canada). The desalted and dried hydrolysate sample was re-suspended in 30% acetonitrile (ACN) mixed with 250 mM acetic acid and the peptides were loaded onto the Fe-IMAC resin. The resin was washed three times with 30% acetonitrile (ACN)/250 mM acetic acid, two times with water after overnight incubation at 4°C. The phosphopeptides were then released from the resin with 400 mM ammonium hydroxide. Both the non-phosphorylated peptides

in the wash buffer and the released phosphopeptides were collected for desalting and LC-MS/MS analysis.

5.2.8 Desalting

Sample desalting was performed by reversed phase (RP) LC using an Agilent 1100 HPLC system fitted with a 4.6 × 50 mm C18 column (Part #: 2001-050×046, Varian, Ontario, Canada). After sample loading, the column was first flushed with 97.5% mobile phase A (0.1% TFA/H_2O) and 2.5% mobile phase B (0.1% TFA/ACN) for 5 min to remove salts. Then, 85% mobile phase B was used to completely elute the peptides. The collected samples were dried down using Speedvac and reconstituted in 0.1% FA/H_2O.

5.2.9 LC-MS and MS/MS

The desalted peptide mixture was analyzed using a quadrupole time-of-flight (QTOF) Premier mass spectrometer (Waters, Milford, MA) equipped with a nanoACQUITY Ultra Performance LC system (Waters). In brief, a sample (2 µg, 1 µg or 0.5 µg) was injected each time into a 75 µm × 150 mm Atlantis C18 column (Part #: 186003500, Waters, Milford, MA). Mobile phase A consisted of 0.1% FA/H_2O and mobile phase B consisted of 0.1% FA/ACN. The peptide sample was separated using a 130-min gradient with the following profile:

0 min: 2% B, 2 min: 6% B, 105 min: 30% B, 115 min: 50% B, 125 min: 90% B, 130 min: 2% B. All samples were electrosprayed into the mass spectrometer at a flow rate of 300 nL/min. MS data were recorded within an m/z window of 300-1600 for 0.8 s, followed by 8 data-dependent MS/MS scans of the eight most intense ions within an m/z window of 50-1900 for 0.8 s. For dynamic mass exclusion, a time window of 180 s and a mass tolerance window of 50 mDa were applied. Varying collision energies were used based on the mass and charge state of the peptide ion. Mass scans of a mixture of mass calibrants consisting of leucine enkephalin and (Glu1)-fibrinopeptide B were acquired at 1 scan/min throughout the run.

5.2.10 Protein database search

The database search procedure was the same as that reported previously[18]. Raw data were lockmass-corrected, de-isotoped, and converted to peak list files by ProteinLynx Global Server 2.3 (Waters). Peptide sequences were identified by automated database searching of peak list files using the Mascot search program. Database searches were restricted to the protein sequence of the corresponding protein downloaded from the SwissProt database. The following search parameters were selected for all database searching: enzyme, none; missed cleavages, 0; peptide tolerance, 30 ppm; MS/MS tolerance, 0.2 Da; peptide charge,

1+, 2+, and 3+; fixed modification, carbamidomethyl (C); variable modifications, Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E), oxidation (M) and deamidated (N and Q). The search results, including unique peptide sequences, ion score, MASCOT threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses were extracted to Excel files using in-house software. All the identified peptides with scores lower than the Mascot threshold score for identity at the confidence level of 95% were then removed from the protein list. The protein sequence coverage was calculated by the Mascot search program. For the phosphopeptide matches, manual inspection of the MS/MS spectra and peak assignments was performed to confirm the phosphorylation site assignment.

5.3 Results and discussion

Figure 5-1 shows the overall workflow of in-gel MAAH using TFA. After protein separation and fixation in gel electrophoresis, the gel was stained with Coomassie blue and then destained with water to produce clear gel bands. The protein band was excised into pieces for in-gel MAAH. The gel pieces were dehydrated by acetonitrile twice and then dried down. To the vial, 25% TFA was added and the proteins in the gel were hydrolyzed by using microwave irradiation. After in-gel MAAH, peptides generated were extracted by using a solution containing 0.1% TFA, 85% acetonitrile and 15% water. The reduction and alkylation of peptides were then performed and the resultant peptides were desalted and then analyzed by LC-ESI MS/MS. The peptide sequences were matched by database search and a sequence map from the multiple peptide matches was finally generated for the gel-separated protein.

For the method development, a standard protein (BSA) was loaded to 12% polyacrylamide gel for SDS-PAGE and analyzed using the workflow shown in Figure 5-1. Some of the key steps involved in the workflow were optimized and the effects of the experimental conditions used on the sequence coverage and number of peptides detected were investigated as described below.

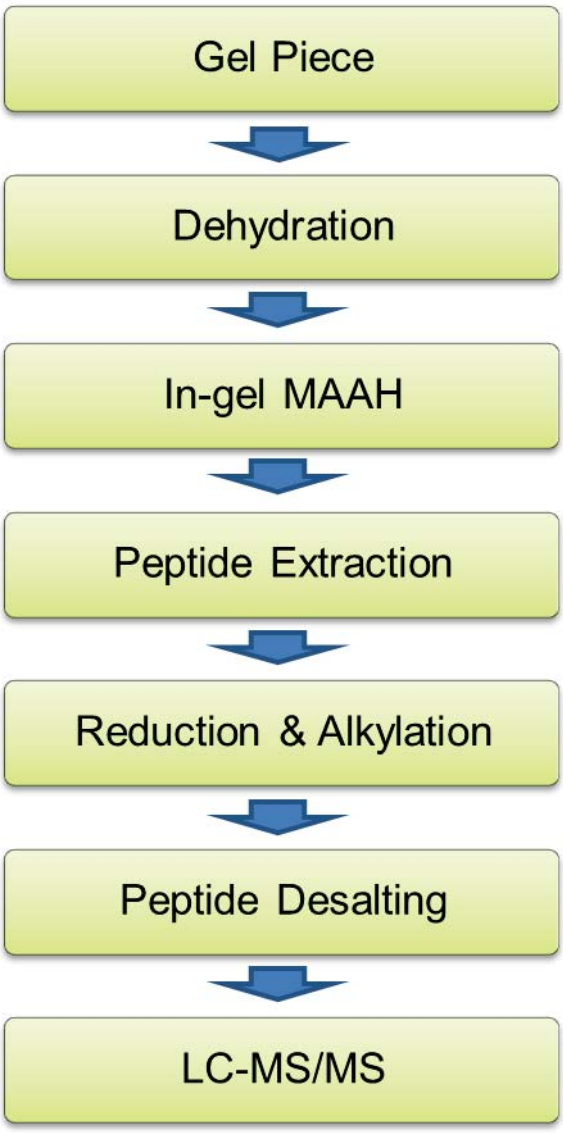


Figure 5-1. Workflow for in-gel MAAH using TFA combined with LC-MS/MS for peptide sequencing for protein sequence mapping.

5.3.1 Sequence of the workflow

In-gel tryptic digestion is a well-established method for protein identification and sequence analysis.[20,21] Compared to tryptic digestion, in-gel MAAH using TFA generates small peptides by non-specific cutting of a protein. Many of the peptides have overlapping sequences, which can be very useful to confirm peptide identities from the database search results and generate redundant sequence information to map the protein sequence and modifications with high confidence.[17,18,22] In in-gel tryptic digestion, after dehydration of the gel, reduction and alkylation are commonly performed. Two additional dehydration steps are sometimes applied to ensure efficient alkylation and in-gel digestion. Using this sequence of sample handling, in-gel MAAH MS was found to generate a sequence coverage of BSA in the range of 52 to 74% from replicate experiments, which was somewhat disappointing, considering that 100% coverage could be obtained from in-solution MAAH MS. After spending considerable efforts and time in the experimentation of different conditions for in-gel TFA MAAH, a simple workflow (Figure 5-1) was developed, which allowed 100% sequence coverage for BSA. It was found that the dehydration step(s) could affect the sequence coverage significantly, possibly related to the protein denaturing that leads to unfavorable acid hydrolysis of some peptide amide bonds. The first round of dehydration as shown in Figure 5-1 is required to take in the hydrolysis reagent.

After this step, a mixture of 120 µL of 12 mM DTT and 40 µL of TFA was added for in-gel MAAH. The peptides were then extracted, followed by reduction/alkylation of the extracted peptides; reduction/alkylation at the peptide level can release the peptides still bound by disulfide bonds after acid hydrolysis of a protein.[22] This sequence avoided two additional rounds of gel dehydration as used in some of the in-gel tryptic digestion procedures. As illustrated below, 100% sequence coverage for BSA could be obtained.

5.3.2 Microwave irradiation time

To optimize the in-gel MAAH method to generate high sequence coverage of a protein, several experimental conditions used in the workflow were examined. First of all, the effect of microwave irradiation time was investigated. For in-solution TFA MAAH, 8-10 min was found to be optimal when a household 1200 W (2450 MHz) microwave oven was used.[18] For in-gel MAAH, four different irradiation times (i.e., 6, 8, 10 and 12 min) were examined. In each test, 40 µg BSA was loaded onto the gel and the main BSA band (see lane one in the gel image shown in Figure 5-2 (a)) was cut out for in-gel MAAH. After peptide extraction and desalting, 2 µg of peptides was injected into LC-MS for analysis; this amount is the optimal amount of injection for sequencing peptides using the

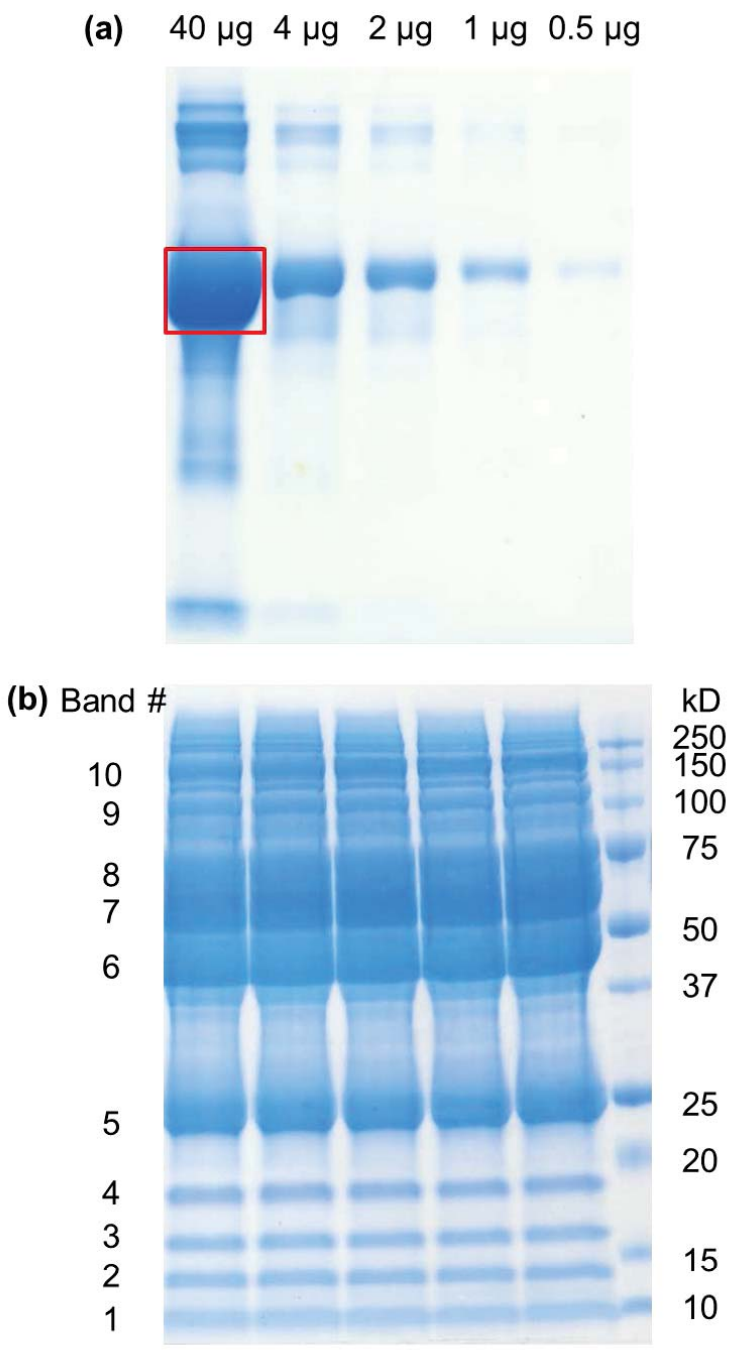


Figure 5-2. (a) SDS-PAGE image of BSA using different sample loading amounts. (b) SDS-PAGE image of human plasma proteins after albumin depletion (50 µg of total proteins was loaded per lane).

capillary LC QTOF instrument.[23] The amount of 2 µg was calculated based on the initial protein loading amount. The BSA sample used was not 100% pure, as evident from several bands, including BSA dimers, detectable in the gel (see lane one in Figure 5-2 (a)). During the in-gel MAAH and subsequent processing, some sample loss is also expected. Therefore, the actual peptide amount injected should be less than the calculated peptide amount. Nevertheless, we controlled the amount of sample injected to provide a fair comparison of the effects of different experimental conditions on the performance of in-gel MAAH.

It was found that, by using 6 or 8 min microwave irradiation, 100%±0% (n=3) sequence coverage could be reached. For 10 or 12 min irradiation, 99%±1% (n=3) sequence coverage could be obtained and in both cases the total sequence coverage was 100% from the combined results of the triplicate experiments. Figure 5-3 (a) shows the number of unique peptides identified using different irradiation time. By using 6 or 8 min irradiation, 1234±22 (n=3) and 1256±36 (n=3) unique peptides were identified, respectively. When the irradiation time increased further, the number of unique peptides decreased slightly. In the cases of 10 and 12 min irradiation, 1107±42 (n=3) and 1081±55 (n=3) unique peptides were identified, respectively. This decrease may be due to the generation of smaller peptides that were too short for sequencing by the LC-MS/MS setup used.

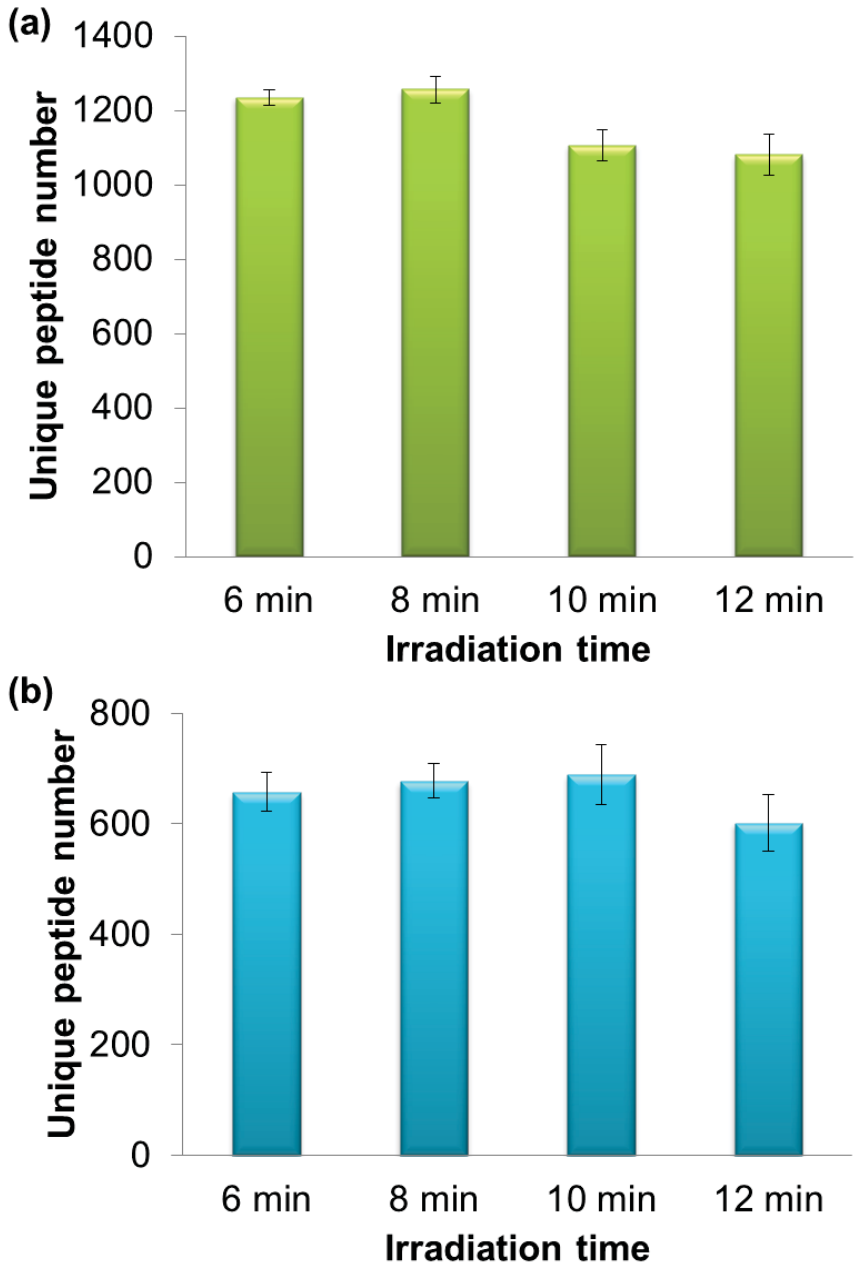


Figure 5-3. Comparison of the number of unique peptides identified from in-gel MAAH MS of BSA by using different microwave irradiation time with (a) 40 µg and (b) 4 µg of protein sample loading onto the gel. Error bars represent one standard deviation.

The results shown in Figure 5-3 (a) indicate that 8 min irradiation time and 2 μg of peptide injection amount are the optimized conditions for in-gel MAAH MS, when 40 μg loading amount is used for gel electrophoresis.

5.3.3 Effect of protein amount on microwave irradiation time

While loading a large amount of proteins onto the gel may not be an issue for characterizing a recombinant protein such as for quality control of a protein-based drug in the production or formulation process,[24] in many other real world applications, the total protein amount may be limited. In addition, for gel electrophoresis, the sample loading amount should also be optimized to avoid the saturation of the gel separation. In a complex protein mixture, the concentration dynamic range of different proteins may be very high. Thus, it is desirable for in-gel MAAH to be able to handle varying amounts of proteins. The performance of in-gel MAAH MS for sequencing proteins with varying amounts of loading was examined. Lanes 2 to 5 in Figure 5-2 (a) show the gel image of BSA with a sample loading of 4, 2, 1 and 0.5 μg, respectively. Comparison of lane 1 and lane 2 indicates that, when the loading amount decreases by ten-fold from 40 μg to 4 μg, the area of the main gel band decreases only by about two-fold. Thus, if the same condition is used for in-gel MAAH, the protein concentration in the 4 μg band should be about five times less than that in the 40 μg band. Since the

hydrolysis reagents (water and acid) are in excess, the acid hydrolysis rate of the intact protein is pseudo-first order in protein concentration. Therefore, if the protein concentration changes, the reaction rate should change, which means the optimized microwave irradiation time may be different.

To examine whether the protein loading amount would have a significant effort on the optimized microwave irradiation time, we have determined the optimized irradiation time for the 4 μg sample loading. For each test, the peptide injection amount into LC-MS was still 2 μg. It was found that, by using 6 min irradiation, 100%±1% (n=3) sequence coverage could be reached. For 8 min irradiation, 99%±0% (n=3) sequence coverage could be obtained and the total sequence coverage was 100% from the combined triplicate results. Similar results were obtained for 10 min and 12 min. The number of unique peptides identified was also investigated and the results are shown in Figure 5-3 (b). By using 6, 8 and 10 min irradiation, 657±35 (n=3), 677±32 (n=3) and 689±54 (n=3) unique peptides were identified, respectively. The difference in the peptide number is not significant for these three conditions. For the 12 min irradiation, slightly lower number of unique peptides, 601±51 (n=3), was obtained, which is consistent with the results obtained from the 40 μg sample loading, i.e., a prolonged irradiation may result in smaller peptides that are difficult to sequence using LC-MS/MS.

The above results indicate that the sample loading amount (40 μg vs. 4 μg) did not affect the optimized irradiation time significantly. Considering that, on average, slightly more peptides were detected using 10 min irradiation for the lower amount sample (i.e., the 4-μg sample; see Figure 5-3 (b)) and we dealt with mainly low micrograms of samples in most applications, we chose 10 min irradiation for the subsequent in-gel MAAH experiments.

5.3.4 Gel background interference

As Panels a and b in Figure 5-3 show, the number of unique peptides identified from the 4 μg protein loading onto the gel decreases significantly, compared to the 40 μg loading. In both cases, the injected peptide amount into LC-MS was calculated from the protein sample loading amount and they should be similar, i.e., 2 μg. To investigate the cause of this number decrease, we examined the base-peak ion chromatograms obtained from the 40 μg and 4 μg loading that are shown in Figure 5-4 (a) and 5-4 (b), respectively. Comparing the two chromatograms as well as other replicate chromatograms from each sample, it can be seen that the peak patterns are not the same from these two samples, while the peak patterns are the same for the triplicate data from each sample. For the 40 μg loading, most of the peaks were confirmed to be from the peptide ions by examining the MS/MS spectra associated with the chromatographic peaks.

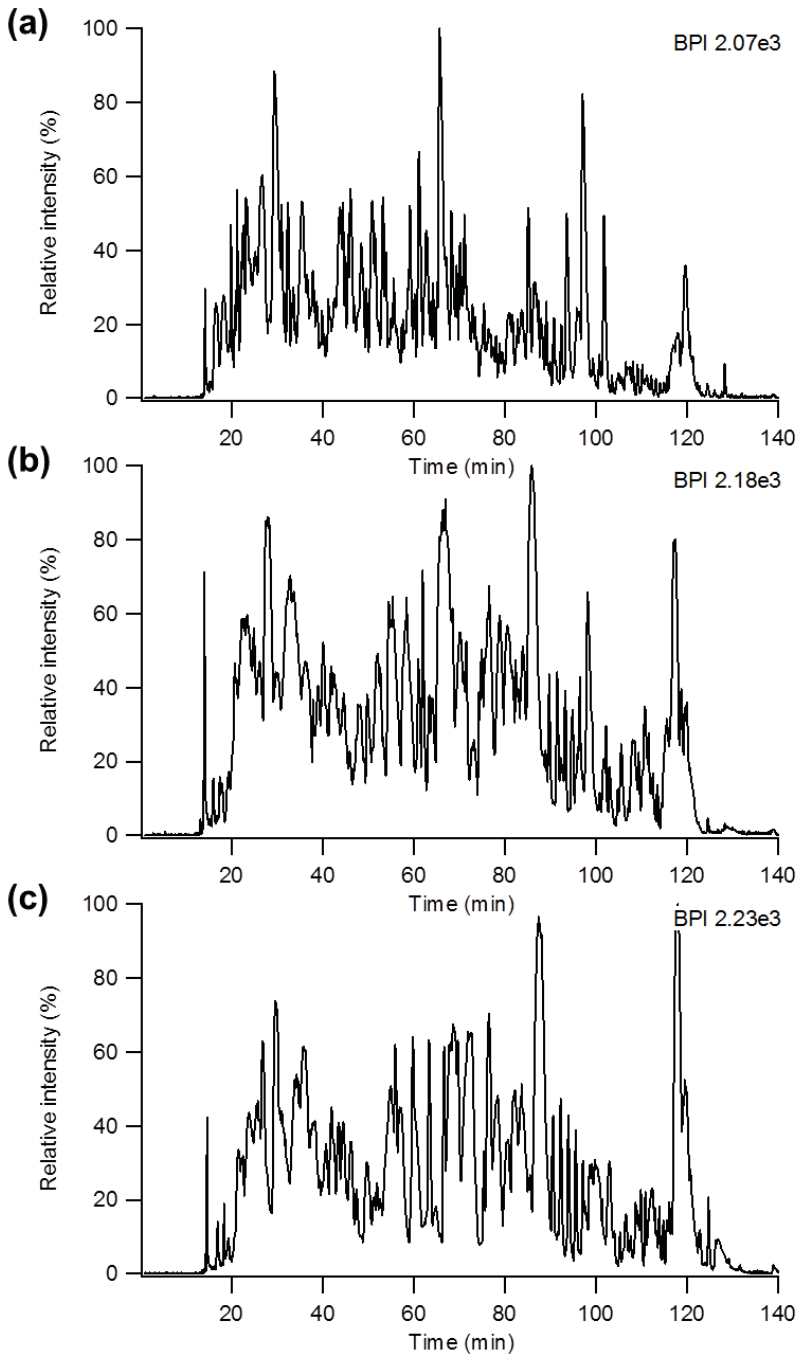


Figure 5-4. Comparison of the ion chromatograms obtained from in-gel MAAH MS of (a) a gel band from 40 µg BSA loading to the gel, (b) a gel band from 4 µg of BSA loading and (c) a blank gel band.

However, for the 4 μg loading, there are several broad chromatographic peaks with high intensity (e.g., peaks at the retention time of ~72 min and ~87 min in Figure 5-4 (b)). Suspecting that these peaks might be from the background ions, we did a set of experiments whereby a blank reagent or blank gel was subjected to the same workflow as shown in Figure 5-1. No or little background signals were detected in the blank runs of the reagents used, suggesting that the background ions were not from the reagents.

Figure 5-4 (c) shows an ion chromatogram obtained from in-gel MAAH of a blank gel. In this case, after gel electrophoresis, one piece of blank gel with the same area as the main gel piece of the 4 μg BSA loading was cut out for analysis. As Figure 5-4 (c) shows, the broad peaks found in the 4 μg loading are also present in the blank, indicating the presence of gel background signals in the analysis of the protein gel band from the 4 μg loading. However, for the 40 μg loading, the background interference was much less. This can be explained by examining the relative ratio between the peptides and the background chemicals in the peptide sample injected into LC-MS. As indicated earlier, when the protein loading amount decreased from 40 μg to 4 μg, the gel area decreased by only about two-fold. Therefore, when the same amount of peptides was injected into LC-MS, five times more of the gel background would be injected for the 4 μg loading, compared to the 40 μg loading. Thus, the background interference for the

40 μg loading was not severe. The background interference causing the reduction of the number of peptides detectable for the 4 μg loading may be explained by considering the ion suppression and sampling effects. Figure 5-5 shows the extracted ion chromatogram of a peptide ion with a sequence of LTADFAEDK (+1, m/z 768.2860) that was detectable from the in-gel hydrolysate of the 40 μg loading, but not detected from the 4 μg loading. This ion was totally suppressed by the background ions in the 4-μg loading sample.

At this stage, we have not been able to identify the chemical structures of the background species. The MS and MS/MS spectra obtained from some of the background ions did not match with the chemical structures of the expected polyacrylamide oligomers. Similar background signals and chromatographic patterns to that of the commercial BioRad gel (Figure 5-4 (c)) were observed from the gels casted by ourselves. Since the reduction of the background signals is critical to improve the overall sensitivity of the method, developing an improved gel system that would not generate interference with in-gel MAAH MS is needed. Alternatively, developing a method to separate the peptides from the interfering chemicals prior to MS analysis may also improve the overall detection sensitivity. To this end, we plan to synthesize gels with different cross-linkers and initiators to aid in deducing the chemical structures of the interfering chemicals with an aim to develop an acid-resistant gel for gel electrophoresis. However, even with the use

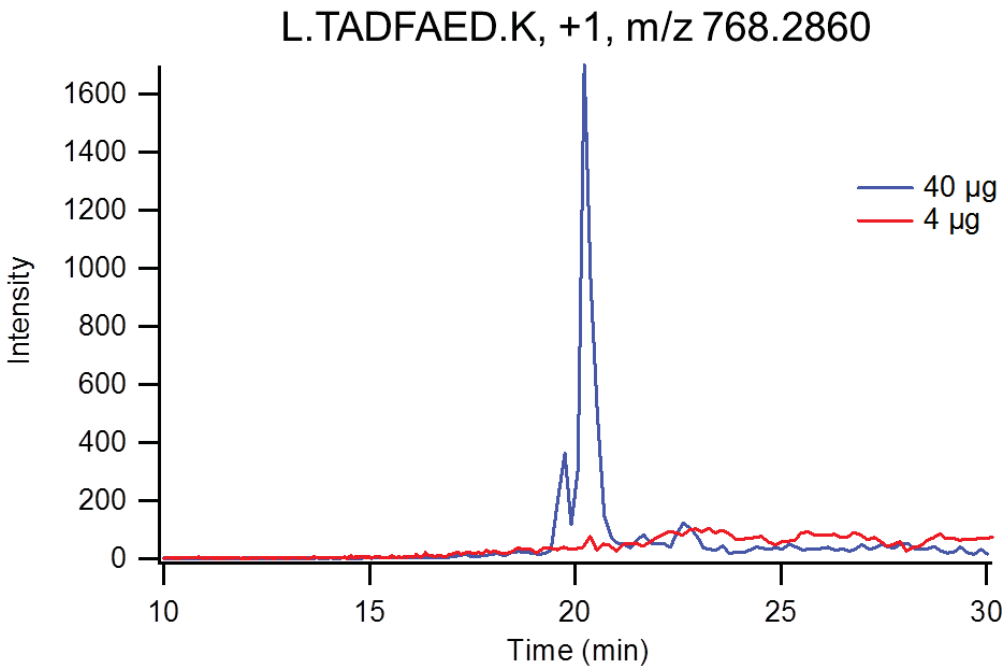


Figure 5-5. Extracted ion chromatograms of a peptide ion at m/z 768.2860 (sequence: LTADFAEDK, +1; with 0.05 Da mass window) obtained from in-gel MAAH MS of a 40 µg BSA loading gel band (blue) and a 4 µg BSA loading (red).

of the conventional form of polyacrylamide gel, high sequence coverage can still be achieved using micrograms of proteins, as demonstrated below.

5.3.5 Sensitivity

The effect of the protein amount loaded onto the gel on the detectability of the peptides generated from in-gel MAAH MS is shown in Figure 5-6. The gel images from a sample loading of 2 μg, 1 μg and 0.5 μg are shown in Figure 5-2 (a). In these cases, all the peptides extracted were injected into LC-MS for analysis. The sequence coverage of BSA decreases from 94%±3% (n=3) for 2 μg loading, to 76%±5% (n=3) for 1 μg loading, and 32%±2% (n=3) for 0.5 μg loading. The total sequence coverage from the combined triplicate results was 97%, 90% and 52%, respectively. As Figure 5-6 shows, the number of peptides identified also decreases significantly as the protein loading amount decreases, i.e., 689±54 (n=3) for 4 μg, 316±59 (n=3) for 2 μg, 136±19 (n=3) for 1 μg and 30±2 (n=3) for 0.5 μg. Comparing the results from 4 μg and 2 μg loading where the same peptide amount, i.e., 2 μg, was injected into LC-MS, the significant decrease in peptide number is mainly due to the background interference. For even lower amount of loading, both the reduced peptide concentration and the high background interference contribute to the reduction of peptide number and sequence coverage.

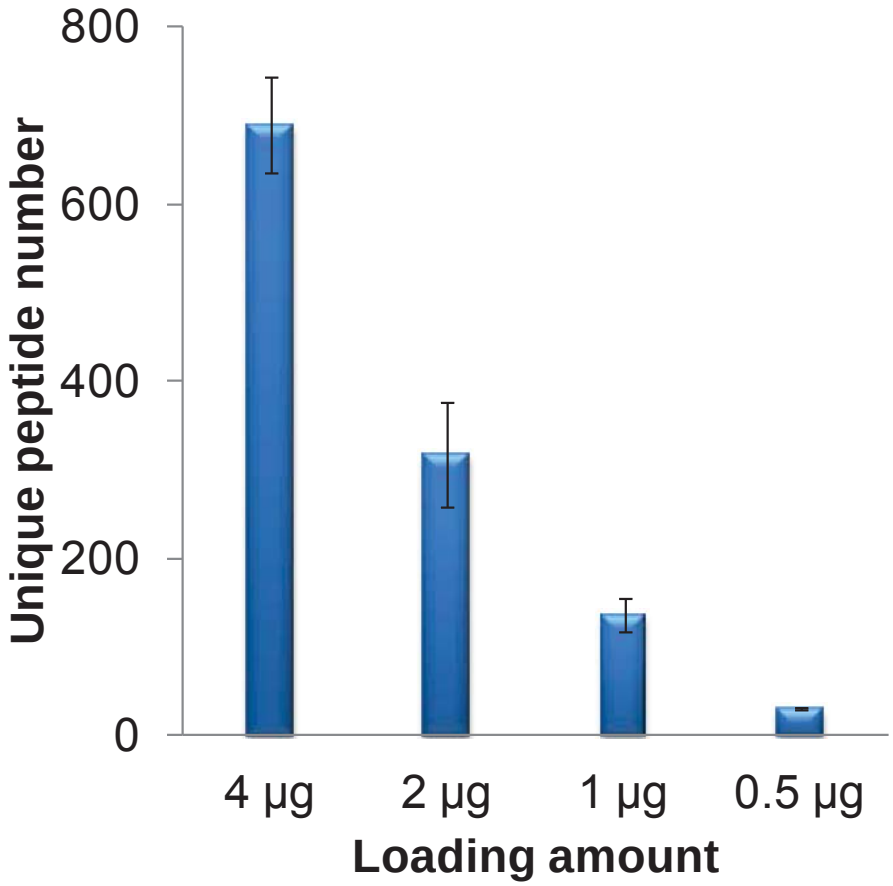


Figure 5-6. Comparison of the number of unique peptides identified from in-gel MAAH MS of BSA with different protein loading amounts to SDS-PAGE. Error bars represent one standard deviation.

The above results indicate that, for BSA, the in-gel MAAH MS method can detect unique peptides even with a protein sample loading of 0.5 μg to the gel. However, to achieve very high sequence coverage (>95%), 2 μg of sample loading was needed. The sensitivity of the method may be improved by using a more sensitive tandem MS instrument than the Premier QTOF instrument used herein, and by reducing background interference. Nevertheless, the current method can already be applied for sequence mapping of gel-separated proteins with high sequence coverage and an example is shown below.

5.3.6 Human plasma protein analysis

To demonstrate the general applicability of the in-gel TFA MAAH MS method, the performance of the method for characterizing a number of proteins in human plasma was investigated. Figure 5-2 (b) shows the SDS-PAGE gel image obtained from the albumin-depleted plasma sample with a sample loading of 50 μg total proteins per lane. Ten major gel bands of proteins with molecular weights ranging from ~8 kD to ~160 kD were excised for analysis. In total, 19 human plasma proteins were sequence-mapped and the results are summarized in Table 5-1.

Table 5-1 shows the comparison of protein sequence coverage obtained by using the in-gel MAAH MS method with those reported in the literature.

Table 5-1. Summary of the number of unique peptides identified and sequence coverage from duplicate in-gel MAAH MS experiments of ten bands from the SDS-PAGE gel of an albumin-depleted human plasma sample and the comparison with other reported work (TD: top-down method; N: no sequence coverage reported).

Band #	Protein name	MW (Da)	Number of unique peptides		Sequence coverage		Total sequence coverage	
			Run #1	Run #2	Run #1	Run #2	This work	Reported work
1	Apolipoprotein A-II	8759	134	128	100%	100%	100%	69%[34]
2	Apolipoprotein C-III	8759	33	20	92%	74%	92%	
	Hemoglobin subunit alpha	15174	161	94	100%	92%	100%	
	Hemoglobin subunit beta	15971	130	89	92%	84%	92%	
3	Transthyretin	13810	287	224	100%	100%	100%	74-91%[35] TD(N)[36]
4	Haptoglobin alpha chain	16335	133	148	97%	96%	97%	
5	Apolipoprotein A-I	28061	358	357	100%	100%	100%	TD(N)[37]
	Ig kappa chain C region	11773	175	149	100%	100%	100%	51%[38]
	Ig lambda-2 chain C regions	11458	107	76	96%	88%	97%	

6	Ig gamma-1 chain C region	36596	609	549	98%	97%	98%	58%[39]
	Ig gamma-2 chain C region	36505	516	456	94%	94%	96%	
	Ig gamma-3 chain C region	42287	488	434	79%	76%	80%	
	Ig gamma-4 chain C region	36431	486	432	91%	92%	96%	
	Alpha-1-antitrypsin	44354	224	249	82%	90%	92%	44%[39]
	Fibrinogen beta chain	51358	62	70	45%	41%	56%	
7	Serum albumin	68425	620	587	99%	99%	100%	97%[12]
8	Serotransferrin	77313	493	483	86%	80%	87%	56%[39]
9	Complement C3 alpha chain	114325	345	342	78%	78%	84%	
10	Alpha-2-macroglobulin	162134	758	668	80%	79%	84%	<45%[39-40]

Apolipoprotein A-Ⅱ was identified from band #1 with 100% sequence coverage in duplicate experiments using in-gel MAAH MS. Compared to 69% sequence coverage reported by Martosella et al. using HPLC and SDS-PAGE for protein fractionation and 2D LC-MS/MS analysis,[34] the in-gel MAAH method provides higher sequence coverage. From band #2, three proteins, Apolipoprotein C-Ⅲ, Hemoglobin subunit alpha and beta, were identified with 92%, 100% and 92% coverage, respectively. Less than 100% coverage is likely due to the lower amount of the protein present in the band. Transthyretin was identified from band #3 with 100% sequence coverage. In the work by He et al., 74-91% coverage was reported for this protein by using 2D gel electrophoresis combined with MALDI-TOF analysis.[35] Top-down analysis was recently applied to this protein and, while sequence coverage was not reported, 30 out of 127 individual amino acids could be examined.[36] In our work, 82 and 76 individual amino acids could be examined in replicate #1 and replicate #2, respectively. Thus, the in-gel MAAH MS method provides much more individual amino acid information, compared to the top-down analysis.

From band #4, haptoglobin alpha chain was identified with 96-97% coverage. The missed peptide sequence was the C-terminal NPVQ. The peptide bond between A and N was readily hydrolyzed (i.e., Gly or Ala feature).[17] The small peptide generated could not be identified using the m/z window set in our

LC-MS experiment. From band #5, Apolipoprotein A-I, Ig kappa chain C region and Ig lambda-2 chain C regions were identified with high sequence coverage (100%, 100% and 97%, respectively). In the work by Zhang et al., top-down analysis of Apolipoprotein A-I (~28 kDa) by electron capture dissociation was investigated and 45 individual amino acids could be examined with an unspecified sequence coverage.[37] In our work, 100% sequence coverage was obtained and 119-129 individual amino acids could be examined. In the work by Wasinger et al., Ig kappa chain C region was identified with 51% sequence coverage by using the Gradiflow BF400 as a fractionation tool to deplete highly abundant albumin from human plasma, followed by tryptic digestion and 2D LC-MS/MS.[38] In our work, 100% sequence coverage was obtained. For Ig lambda-2 chain C regions in band #5, the total sequence coverage was 97% from the combined duplicate results. The missed peptide sequence was the internal peptide ^{61}SNN^{64}.

Six high abundance proteins were identified from band #6. In the work by He et al., 58% sequence coverage was reported for Ig gamma-2 chain C region and 44% for Alpha-1-antitrypsin using the bottom-up method.[39] In our work, the sequence coverage for Ig gamma-2 chain C region was 94% (96% from the two runs combined) and Alpha-1-antitrypsin was identified with 82%-90% sequence coverage (92% from the combined. For Ig gamma-3 chain C region, 80% sequence coverage was obtained, possibly due to lower concentration of this

protein. compared to others in the mixture. The sequence coverage of Fibrinogen beta chain was found to be 45% and 41% from two individual runs and 56% from the combined results. This low coverage is likely due to the presence of phospho-modifications in the protein sequence. Lower concentration of this protein in the mixture may also be a factor. Future work on enriching phosphopeptides after in-gel MAAH should reduce the ion suppression effect, thereby potentially increasing the overall sequence coverage for this highly phosphorylated protein. In addition, the use of two-dimensional gel electrophoresis combined with loading of a larger amount of the sample should also assist in detecting more peptides and phosphopeptides for increasing sequence coverage.

The most abundant protein in human plasma, serum albumin, was identified in band #7. In the work by Wa et al., human serum albumin was used as a model protein to explore several approaches for obtaining high sequence coverage in protein modification studies using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS. They could reach 97% sequence coverage.[12] In our work, although depletion of albumin was performed, albumin could still be detected in the gel. For this protein, 99% sequence coverage was readily obtained from individual runs and 100% coverage was achieved from the two runs combined. In total, 587-620 unique peptides were identified.

From band #8, Serotransferrin (~77 kDa) was identified with 86% and 80% sequence coverage from individual runs and a total sequence coverage of 87% from the combined duplicate results. Compared to the sequence coverage of 56% reported by He et al.,[39] in-gel MAAH MS generates better sequence coverage. From band #9, Complement C3 alpha chain (~114 kDa) was identified. For this large protein, 78% and 78% sequence coverage from individual runs and 84% total sequence coverage were obtained. From band #10, Alpha-2-macroglobulin (~162 kDa) was identified. In previous reports, less than 45% sequence coverage could be obtained.[39,40] In our work, 80% and 79% sequence coverage could be obtained from individual runs and 84% coverage was reached from the combined duplicate results. In this case, 758 and 668 unique peptides were detected from the two runs.

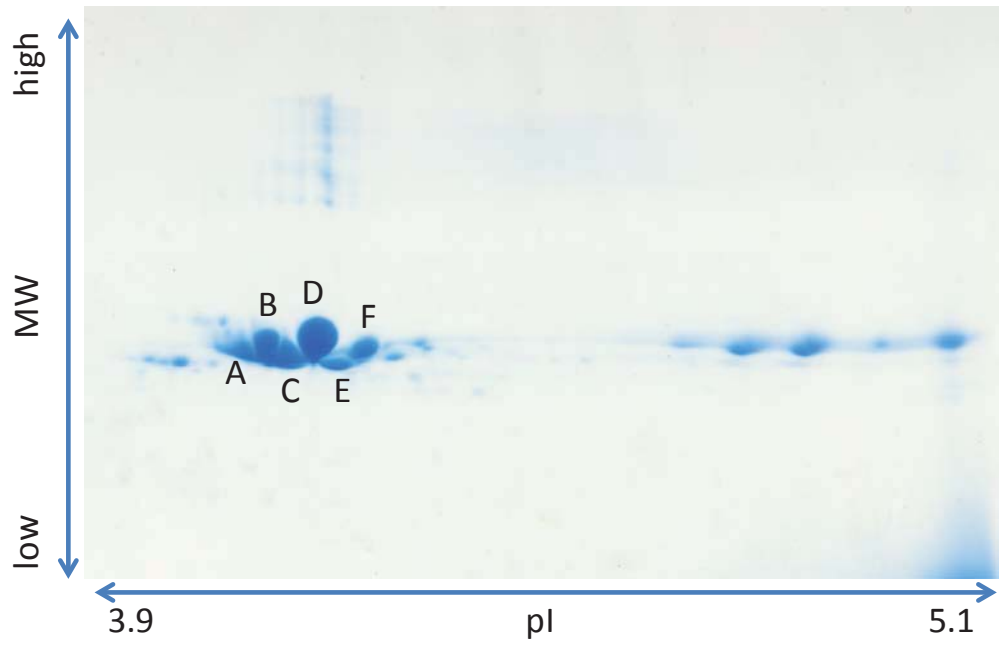
The results shown in Table 5-1 indicate that most of the 19 high abundance human plasma proteins can be identified with high sequence coverage by using the in-gel MAAH MS method. In all cases, the sequence coverage obtained was higher than those reported using other techniques. This example demonstrates that this method can be applied to proteins with a wide range of molecular weights, thereby opening the possibility of detailed protein sequence analysis that is difficult or not possible with the current bottom-up or top-down method. The major limitation of the current method is that it cannot provide the

structure or modification site information on a modification group easily hydrolyzed by MAAH (e.g., glycans in glycoproteins). However, most other commonly encountered modification groups can survive the acid hydrolysis process and thus be characterized using the sequence mapping method.

5.3.7 Characterization of bovine alpha-S1-casein phosphoprotein isoforms

While phosphoprotein identification can be readily done by detecting one or more phosphopeptides, detailed characterization of a phosphoprotein can still be a challenge task. 2D gel electrophoresis can be used to separate phosphoproteins including isoforms. Thus, the combination of in-gel MAAH with 2D gel electrophoresis of proteins may provide an improved means of phosphoprotein characterization. In this work, we have examined the application of the in-gel TFA MAAH MS method for characterizing bovine alpha-S1-casein phospho-isoforms with variable modification sites. As this sample has been widely used as a model sample for gauging the performance of a newly developed analytical technique for phosphoprotein analysis,[25-33] comparison of our results with those reported in the literature should provide a fair assessment of our method against other reported methods.

Figure 5-7 shows the gel image obtained from 2D gel electrophoresis of the bovine alpha casein sample with a sample loading of 150 μg total proteins. Six



(A) pS^{41}, pS^{46}, pS^{48}, T^{49}, pS^{64}, pS^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}
pS^{41}, pS^{46}, S^{48}, pT^{49}, pS^{64}, pS^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}

(B) pS^{41}, pS^{46}, pS^{48}, T^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, S^{88}, pS^{115}, S^{122}, S^{178}, S^{188}
pS^{41}, pS^{46}, S^{48}, pT^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, S^{88}, pS^{115}, S^{122}, S^{178}, S^{188}

(C) S^{41}, pS^{46}, pS^{48}, T^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}
S^{41}, pS^{46}, S^{48}, pT^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}

(D) S^{41}, pS^{46}, pS^{48}, T^{49}, pS^{64}, pS^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}
S^{41}, pS^{46}, S^{48}, pT^{49}, pS^{64}, pS^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}

(E) S^{41}, pS^{46}, pS^{48}, T^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}
S^{41}, pS^{46}, S^{48}, pT^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, pS^{115}, S^{122}

(F) S^{41}, pS^{46}, pS^{48}, T^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, Y^{104}, pS^{115}
S^{41}, pS^{46}, S^{48}, pT^{49}, pS^{64}, S^{66}, pS^{67}, pS^{68}, pS^{75}, Y^{104}, pS^{115}

Figure 5-7. (Top) 2D gel electrophoresis image of a bovine alpha casein sample with a sample loading of 150 µg. (Bottom) Summary of the major phospho-isoforms determined from each marked spot (p: phosphorylated site; black: known phosphorylation site; blue: new phosphorylation site).

major gel spots of alpha-S1-casein phospho-isoforms (Spot A-F) were excised for analysis. By using IMAC for phosphopeptide enrichment, phosphopeptides and non-phosphorylated peptides of the protein hydrolysate in each spot were separated for LC-MS/MS analysis. The phospho-isoforms were sequence-mapped and the results were summarized in Table 5-2. The phosphorylation site was assigned and the major isoforms in each spot were shown at the bottom of Figure 5-7. As it is shown in Table 5-2, alpha-S1-casein variant C was only identified from Spot E with 100% sequence coverage in duplicate experiments. Alpha-S1-casein variant B was identified from the other five spots with 100% sequence coverage in duplicate experiments.

The sequence analysis of phosho-isoforms from in-gel MAAH LC-MS of a given spot shown in Figure 5-7 can be illustrated in the analysis of Spot E. For Spot E, 132 and 136 unique peptides were identified from the IMAC-enriched phosphopeptide fractions in duplicate experiments. The enrichment efficiency was 91.4% and 93.0%, respectively. For all of the phosphopeptide matches, manual inspection of the MS/MS spectra and peak assignments was performed to confirm the phosphorylation site assignment. Nine phosphorylation sites (S^{41}, S^{46}, S^{48}, S^{64}, S^{66}, S^{67}, S^{68}, S^{75}, and S^{115}) that had been reported before were confirmed. At the same time, we found two new sites, T^{49} and S^{122}, which could be phosphorylated (see below). Therefore, the protein phospho-isoform in Spot E contained eleven

Table 5-2. Summary of the protein isomers/variants identified in six spots excised from the 2D gel of a bovine alpha casein sample (P: phosphopeptide fraction; N: non-phosphorylated peptide fraction).

Gel spot #	Alpha-S1-casein variant	Number of unique peptides (P)		Phosphopeptide enrichment efficiency (P)		Number of unique peptides (N)		Sequence coverage (P+N)	
		Run #1	Run #2	Run #1	Run #2	Run #1	Run #2	Run #1	Run #2
A	Variant B	213	212	94.0%	93.7%	474	350	100%	100%
B	Variant B	578	556	95.3%	93.7%	475	590	100%	100%
C	Variant B	361	384	91.8%	94.8%	592	660	100%	100%
D	Variant B	624	620	93.8%	92.8%	654	466	100%	100%
E	Variant C	132	136	91.4%	93.0%	521	601	100%	100%
F	Variant B	166	182	90.6%	94.3%	776	425	100%	100%

sites which could potentially be phosphorylated. Detailed analysis of the phosphopeptides detected (see below) indicates that the major protein components in Spot E were from phospho-isoforms containing seven phosphate groups, pS^{46}-pS^{48}-pS^{64}-pS^{67}-pS^{68}-pS^{75}-pS^{115} and pS^{46}-pT^{49}-pS^{64}-pS^{67}-pS^{68}-pS^{75}-pS^{115}.

In Spot E, there were twenty-nine peptide matches containing the sequence ^{41}SKDIGS^{46} and, among them, twenty-six matches had both serine sites phosphorylated, while two matches had only site S^{41} phosphorylated and one match had neither of the two sites phosphorylated. This indicates that the major protein phospho-isoform in Spot E contained non-phosphorylated S^{41} and phosphorylated S^{46}. For the analysis of sites S^{48} and T^{49}, we found twenty-four matches contained the sequence ^{48}ST^{49}. Among them, four matches had only serine phosphorylated and three matches had only threonine phosphorylated. For the other fourteen matches containing one phosphorylation site, manual interpretation of the matched spectra failed to pinpoint which site was phosphorylated. Among the remaining three matches, one had both two sites phosphorylated and two had neither of the two sites phosphorylated.

The major protein isoform in Spot E contained one phosphorylation site in ^{48}ST^{49}, i.e., either phosphorylated S^{48} or phosphorylated T^{49}. All of the fourteen peptide matches containing the S^{64} site had this site phosphorylated, indicating

that this site was always phosphorylated. Similarly, we examined the 114 matches which contained S^{75} and found that this site was also always phosphorylated. At the same time, site S^{66} was found always non-phosphorylated. For sites S^{67} and S^{68}, there were twenty-two matches containing both of the two sites. Fifteen of them had both of the two sites phosphorylated; four of them only had site S^{67} phosphorylated. The other three matches had only one phosphorylation modification, but we could not pinpoint which site was phosphorylated by manual interpretation. Therefore, the major protein in Spot E should have both of these two sites phosphorylated. In the end, for sites S^{115} and S^{122}, only one match contained non-phosphorylated site S^{115} and phosphorylated site S^{122}. All the other ninety-three matches contained phosphorylated site S^{115}.

In the same manner, the phosphorylation sites and the major protein phospho-isoforms in the other five spots were determined and summarized in Figure 5-7. As Figure 5-7 shows, the two new phosphorylation sites, T^{49} and S^{122}, were also identified from Spot A. The major protein phospho-isoforms containing nine phosphate groups were pS^{41}-pS^{46}-pS^{48}-pS^{64}-pS^{66}-pS^{67}-pS^{68}-pS^{75}-pS^{115} and pS^{41}-pS^{46}-pT^{49}-pS^{64}-pS^{66}-pS^{67}-pS^{68}-pS^{75}-pS^{115} in Spot A. For Spot B, five new phosphorylation sites (T^{49}, S^{88}, S^{122}, S^{178} and S^{188}) were identified. The major protein components in Spot B were the phospho-isoforms with eight phosphate groups, pS^{41}-pS^{46}-pS^{48}-pS^{64}-pS^{67}-pS^{68}-pS^{75}-pS^{115} and pS^{41}-pS^{46}-pT^{49}-pS^{64}-pS^{67}

-pS^{68}-pS^{75}-pS^{115}. For Spots C and D, the two new phosphorylation sites, T^{49} and S^{122}, were identified. The major protein phospho-isoforms in Spot C were pS^{46}-pS^{48}-pS^{64}-pS^{67}-pS^{68}-pS^{75}-pS^{115} and pS^{46}-pT^{49}-pS^{64}-pS^{67}-pS^{68}-pS^{75}-pS^{115}, with seven phosphate groups. In Spot D, the major phospho-isoforms containing eight phosphate groups were pS^{46}-pS^{48}-pS^{64}-pS^{66}-pS^{67}-pS^{68}-pS^{75}-pS^{115} and pS^{46}-pT^{49}-pS^{64}-pS^{66}-pS^{67}-pS^{68}-pS^{75}-pS^{115}. For the last spot, Spot F, T^{49} and Y^{104} were identified as the new sites which could be phosphorylated. The major components in the spot were seven-phosphorylated, pS^{46}-pS^{48}-pS^{64}-pS^{67} -pS^{68}-pS^{75}-pS^{115} and pS^{46}-pT^{49}-pS^{64}-pS^{67}-pS^{68}-pS^{75}-pS^{115}.

From the 2D gel analysis, we successfully separated six different phospho-isoforms of bovine alpha-S1-casein. We also mapped the protein sequence and detected all known phosphorylation sites. In addition, six new phosphorylation sites were identified. Considering that the bovine alpha-S1-casein sample has been widely characterized using many techniques, identification of 6 new phosphorylation sites demonstrates the enabling analytical power of the in-gel MAAH LC-MS method for detailed analysis of protein-isoforms with variable modification sites. This success can be attributed to the combination of high separation resolution of 2D gel electrophoresis and high sequence coverage as well as a great number of overlapping peptides produced and detected by in-gel MAAH LC-MS.

5.4 Conclusions

We have developed a method based on the use of gel electrophoresis to separate proteins, followed by in-gel MAAH of the gel-separated proteins using TFA to produce peptides and LC-MS/MS of the resultant peptides. For BSA (~67 kDa), 100% sequence coverage can be readily obtained with a sample loading of 4 µg onto a gel. The sequence coverage of BSA decreases from 94%±3% (n=3) for 2 µg loading, to 76%±5% (n=3) for 1 µg loading, and 32%±2% (n=3) for 0.5 µg loading, mainly due to the interference of the gel background chemicals produced in MAAH. The total sequence coverage from the combined triplicate results are 97%, 90% and 52%, respectively. The in-gel MAAH MS method in this current form is useful for mapping gel-separated proteins with high sequence coverage. We have demonstrated the application of this method for sequencing 19 relative high abundant human plasma proteins separated by SDS-PAGE, indicating that this method is applicable to characterize a variety of proteins with virtually no molecular weight limit. In addition, we have combined 2D gel electrophoresis for phosphoprotein separation with in-gel MAAH MS for characterizing bovine alpha-S1-casien phospho-isoforms and determined six new phosphorylation sites, which illustrates that this method can be used for detailed characterization of protein modifications. Future work will involve the determination of the chemical structures of the interfering chemicals so to improve the overall detection

sensitivity by developing an acid-resistant gel and/or a means of separating the peptides from the background chemicals prior to MS analysis.

5.5 Literature cited

1. Meyer, B.; Papasotiriou, D. G.; Karas, M., 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2011,** *41* (2), 291-310.

2. Zhu, K.; Kim, J.; Yoo, C.; Miller, F. R.; Lubman, D. M., High sequence coverage of proteins isolated from liquid separations of breast cancer cells using capillary electrophoresis-time-of-flight MS and MALDI-TOF MS mapping. *Anal. Chem.* **2003,** *75* (22), 6209-17.

3. Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G., Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **2007,** *4* (10), 798-806.

4. Larsen, M. R.; Trelle, M. B.; Thingholm, T. E.; Jensen, O. N., Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* **2006,** *40* (6), 790-798.

5. Bergquist, J.; Andersen, O.; Westman, A., Rapid method to characterize mutations in transthyretin in cerebrospinal fluid from familial amyloidotic polyneuropathy patients by use of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chem.* **2000,** *46* (9), 1293-1300.

6. Chait, B. T., Mass spectrometry: Bottom-up or top-down? *Science* **2006,** *314* (5796), 65-66.

7. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007,** *4* (10), 817-821.

8. Tipton, J. D.; Tran, J. C.; Catherman, A. D.; Ahlf, D. R.; Durbin, K. R.; Lee, J. E.; Kellie, J. F.; Kelleher, N. L.; Hendrickson, C. L.; Marshall, A. G., Nano-LC

FTICR tandem mass spectrometry for top-down proteomics: routine baseline unit mass resolution of whole cell lysate proteins up to 72 kDa. *Anal. Chem.* **2012,** *84* (5), 2111-2117.

9. Rozman, M.; Gaskell, S. J., Charge state dependent top-down characterisation using electron transfer dissociation. *Rapid Commun. Mass Spectrom.* **2012,** *26* (3), 282-286.

10. Wu, C.; Tran, J. C.; Zamdborg, L.; Durbin, K. R.; Li, M. X.; Ahlf, D. R.; Early, B. P.; Thomas, P. M.; Sweedler, J. V.; Kelleher, N. L., A protease for 'middle-down' proteomics. *Nat. Methods* **2012,** *9* (8), 822-824.

11. Choudhary, G.; Wu, S. L.; Shieh, P.; Hancock, W. S., Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *J. Proteome Res.* **2003,** *2* (1), 59-67.

12. Wa, C. L.; Cerny, R.; Hage, D. S., Obtaining high sequence coverage in matrix-assisted laser desorption time-of-flight mass spectrometry for studies of protein modification: Analysis of human serum albumin as a model. *Anal. Biochem.* **2006,** *349* (2), 229-241.

13. Wang, D. X.; Baudys, J.; Rees, J.; Marshall, K. M.; Kalb, S. R.; Parks, B. A.; Nowaczyk, L.; Pirkle, J. L.; Barr, J. R., Subtyping botulinum neurotoxins by sequential multiple endoproteases in-gel digestion coupled with mass spectrometry. *Anal. Chem.* **2012,** *84* (11), 4652-4658.

14. Forbes, A. J.; Mazur, M. T.; Patel, H. M.; Walsh, C. T.; Kelleher, N. L., Toward efficient analysis of <70 kDa proteins with 100% sequence coverage. *Proteomics* **2001,** *1* (8), 927-933.

15. Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C., High-throughput middle-down analysis using an orbitrap. *J. Proteome Res.* **2012,** *9* (8), 3886-3890.

16. Zhong, H. Y.; Zhang, Y.; Wen, Z. H.; Li, L., Protein sequencing by mass analysis of polypeptide ladders after controlled protein hydrolysis. *Nat. Biotechnol.* **2004,** *22* (10), 1291-1296.

17. Zhong, H. Y.; 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.

18. 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.* **2010,** *21* (9), 1573-1587.

19. Lopez, J. L., Two-dimensional electrophoresis in proteome expression analysis. *J. Chromatogr. B* **2007,** *849* (1-2), 190-202.

20. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **1996,** *68* (5), 850-858.

21. Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006,** *1* (6), 2856-2860.

22. Reiz, B.; Li, L., Microwave-Assisted Acid and Base Hydrolysis of Intact Proteins Containing Disulfide Bonds for Protein Sequence Analysis by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2011,** *21* (9), 1596-1605.

23. Wang, N.; Xie, C. H.; Young, J. B.; Li, L., Off-line two-dimensional liquid chromatography with maximized sample loading to reversed-phase liquid chromatography-electrospray ionization tandem mass spectrometry for shotgun proteome analysis. *Anal. Chem.* **2009,** *81* (3), 1049-1060.

24. Staub, A.; Guillarme, D.; Schappler, J.; Veuthey, J. L.; Rudaz, S., Intact protein analysis in the biopharmaceutical field. *J. Pharm. Biomed. Anal.* **2011,** *55* (4), 810-822.

25. Shi, S. D. H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W., Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal. Chem.* **2001,** *73* (1), 19-22.

26. Sze, S. K.; Ge, Y.; Oh, H. B.; McLafferty, F. W., Plasma electron capture characterization of large dissociation for the proteins by top down mass spectrometry. *Anal. Chem.* **2003,** *75* (7), 1599-1603.

27. Wu, S. L.; Kim, J.; Hancock, W. S.; Karger, B., Extended Range Proteomic Analysis (ERPA): A new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J. Proteome Res.* **2005,** *4* (4), 1155-1170.

28. Kweon, H. K.; Hakansson, K., Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* **2006,** *78* (6), 1743-1749.

29. Kjeldsen, F.; Savitski, M. M.; Nielsen, M. L.; Shi, L.; Zubarev, R. A., On studying protein phosphorylation patterns using bottom-up LC-MS/MS: the case of human alpha-casein. *Analyst* **2007,** *132* (8), 768-776.

30. Young, J. B.; Li, L., Impulse-driven heated-droplet deposition interface for capillary and microbore LC-MALDI MS and MS/MS. *Anal. Chem.* **2007,** *79*, 5927-5934.

31. Poth, A. G.; Deeth, H. C.; Alewood, P. F.; Holland, J. W., Analysis of the human casein phosphoproteome by 2-D electrophoresis and MALDI-TOF/TOF MS reveals new phosphoforms. *J. Proteome Res.* **2008,** *7* (11), 5017-5027.

32. Wu, S.; Yang, F.; Zhao, R.; Tolic, N.; Robinson, E. W.; Camp, D. G.; Smith, R. D.; Pasa-Tolic, L., Integrated workflow for characterizing intact phosphoproteins from complex mixtures. *Anal. Chem.* **2009,** *81* (11), 4210-4219.

33. Eriksson, A.; Bergquist, J.; Edwards, K.; Hagfeldt, A.; Malmstrom, D.; Hernandez, V. A., Mesoporous TiO_2-based experimental layout for on-target enrichment and separation of multi- and monophosphorylated peptides prior to analysis with matrix-assisted laser desorption-ionization mass spectrometry. *Anal. Chem.* **2011,** *83* (3), 761-766.

34. Martosella, J.; Zolotarjova, N.; Liu, H.; Nicol, G.; Boyes, B. E., Reversed-phase high-performance liquid chromatographic prefractionation of immunodepleted human serum proteins to enhance mass spectrometry identification of lower-abundant proteins. *J. Proteome Res.* **2005,** *4* (5), 1522-37.

35. He, Q. Y.; Lau, G. K.; Zhou, Y.; Yuen, S. T.; Lin, M. C.; Kung, H. F.; Chiu, J. F., Serum biomarkers of hepatitis B virus infected liver inflammation: a proteomic study. *Proteomics* **2003,** *3* (5), 666-74.

36. Theberge, R.; Infusini, G.; Tong, W. W.; McComb, M. E.; Costello, C. E., Top-down analysis of small plasma proteins using an LTQ-Orbitrap. Potential for

mass spectrometry-based clinical. assays for transthyretin and hemoglobin. *Int. J. Mass Spectrom.* **2011,** *300* (2-3), 130-142.

37. Zhang, J.; Zabrouskov, V.; McClure, T. D., Top-down analysis of intact proteins by electron capture dissociation (ECD). *Thermo Scientific LC/MS application note* **2007,** *365*, 1-6.

38. Wasinger, V. C.; Locke, V. L.; Raftery, M. J.; Larance, M.; Rothemund, D.; Liew, A.; Bate, I.; Guilhaus, M., Two-dimensional liquid chromatography/tandem mass spectrometry analysis of Gradiflow fractionated native human plasma. *Proteomics* **2005,** *5* (13), 3397-401.

39. He, P.; He, H. Z.; Dai, J.; Wang, Y.; Sheng, Q. H.; Zhou, L. P.; Zhang, Z. S.; Sun, Y. L.; Liu, F.; Wang, K.; Zhang, J. S.; Wang, H. X.; Song, Z. M.; Zhang, H. R.; Zeng, R.; Zhao, X., The human plasma proteome: analysis of Chinese serum using shotgun strategy. *Proteomics* **2005,** *5* (13), 3442-53.

40. Holm, L.; Ackland, G. L.; Edwards, M. R.; Breckenridge, R. A.; Sim, R. B.; Offer, J., Chemical labelling of active serum thioester proteins for quantification. *Immunobiology* **2012,** *217* (2), 256-64.

Chapter 6 Conclusions and Future Work

In proteomics analysis, ideally all the proteins or the entire proteome in a given sample should be profiled and quantified. However, this is a very challenging task, especially for membrane proteins that are difficult to handle due to their high hydrophobicity. The overall goal of this thesis research was to develop new mass spectrometric techniques for comprehensive membrane proteome analysis and protein sequence mapping.

In the past several years, shotgun proteomics has been rapidly developed as a complementary method for membrane protein identification.[1,2] These methods usually require the digestion of solubilized proteins into complex peptide mixtures which are subsequently analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS).[3] The overall performance of a shotgun membrane proteome analysis is very much dependent on the efficiencies of protein solubilization. After a brief introduction of a number of key techniques and methods related to my thesis work in Chapter 1, we report the comparison of the performance of RapiGest-, PPS-, and SDS-based sample preparation methods for shotgun membrane proteome analysis in Chapter 2.[4] Using the membrane fractions of *E. coli* and MCF7 cell extracts as models, we demonstrated that the use of RapiGest allows identification of more peptides and proteins than PPS or SDS. In comparing the SDS and RapiGest methods, the RapiGest method does not require an additional step for processing the digest, while the SDS method

requires the removal of SDS that resulted in peptide sample loss. Overall, the RapiGest method is superior to the SDS and PPS methods in handling membrane proteome samples.

We developed a high throughput plasma membrane protein identification and quantitation strategy based on the RapiGest-assisted membrane protein solubilization method in Chapter 3. The plasma membrane proteins of two NPM-ALK-expressing cell lines, Karpas 299 and SUPM2, were identified using 2D LC-MS/MS analysis. 561 unique plasma membrane proteins and extracellular proteins were identified from the Karpas 299 cell line and 552 unique plasma membrane proteins and extracellular proteins were identified from the SUPM2 cell line. In addition, 2-MEGA labeling combined with 2D LC-MS/MS was used for plasma membrane protein quantitation analysis in NPM-ALK expressing HEK 293 cells and NPM-ALK absent HEK 293 cells (control). 48 unique plasma membrane proteins and extracellular proteins were found to be differentially expressed between the NPM-ALK expressing HEK 293 cells and the NPM-ALK absent HEK 293 cells. Among these proteins, 39 were up-regulated and 9 were down-regulated. Of the 48 proteins, six proteins were selected as putative biomarkers by using bioinformatics analysis and literature search.

Although the RapiGest-assisted membrane protein solubilization method has been demonstrated to be superior to the SDS and PPS methods in handling membrane proteome samples, the high cost of it compared to SDS can be a major concern in large scale proteome analysis or in situations where a large amount of surfactant is required to process a sample (e.g., working with proteins electro-eluted from a polyacrylamide gel in SDS solution).[5] In Chapter 4, we developed an integrated strong-cation exchange liquid chromatographic procedure for SDS removal and peptide separation for SDS-assisted shotgun proteome analysis.[6] The peptide sample recovery rate was found to be about 90% for the digests of BSA and membrane-protein-enriched fractions of the cell lysates of *E. coli* and MCF-7 breast cancer cell line. To achieve this high recovery rate, SDS was purposely added, after trypsin digestion, to reach a final concentration of 0.5% in the digest. Compared to other methods of using acid-labile surfactants, namely RapiGest and PPS, for the MCF-7 membrane fraction sample, the SDS method identified more peptides (~5%) and proteins (~16%) than the RapiGest method and both methods identified more peptides and proteins than the PPS method. Since SCX is widely used as the first dimension of 2D-LC MS/MS in a shotgun proteome analysis workflow, integration of SDS removal with peptide separation in SCX will not add any extra steps to the sample handling process. To our knowledge, this is the first report of using SCX to remove SDS while keeping a

high peptide recovery rate (~90%). This method provides convenient and inexpensive means of handling proteome samples that require SDS for protein extraction, solubilization, or digestion in the 2D-LC MS/MS shotgun proteome analysis workflow.

Aside from protein identification and quantification, proteomics work also employs protein sequence mapping for unambiguous identification of proteins or protein isoforms,[7,8] study of protein modifications[9,10] and discovery of amino acid substitutions from point mutations in the genome.[11] For these applications, methods which could generate high sequence coverage of proteins are needed. In Chapter 5, we developed a method based on the use of gel electrophoresis to separate proteins, followed by in-gel MAAH of the gel-separated proteins using TFA to produce peptides and LC-MS/MS of the resultant peptides. This method could provide high sequence coverage of proteins separated from a complex protein mixture. For BSA (~67 kDa), 100% sequence coverage can be readily obtained with a sample loading of 4 μg onto a gel. The sequence coverage of BSA decreased from 94%±3% (n=3) for 2 μg loading, to 76%±5% (n=3) for 1 μg loading, and 32%±2% (n=3) for 0.5 μg loading, mainly due to the interference of the gel background chemicals produced in MAAH. The in-gel MAAH MS method in this current form is useful for mapping gel-separated proteins with high sequence coverage. We demonstrated the application of this

method for sequencing 19 relative high abundant human plasma proteins separated by SDS-PAGE, indicating that this method is applicable to characterize a variety of proteins with virtually no molecular weight limit. In addition, we combined 2D gel electrophoresis for phosphoprotein separation with in-gel MAAH MS for characterizing bovine alpha-S1-casein phospho-isoforms and determined six new phosphorylation sites, which illustrates that this method can be used for detailed characterization of protein modifications.

In summary, several techniques and procedures have been developed for membrane proteome qualitative, quantitative analysis, and protein sequence mapping. However, there are still a lot of work remained to be done.

First, biological validation, such as western blotting and immunohistochemical analysis, of the six proteins selected as putative biomarkers of ALK+ ALCL in Chapter 3 needs to be done. The biomarker candidates demonstrated may have the potential to be used for ALK+ ALCL diagnosis and also as drug targets for effective cancer therapeutics.

Second, the methodology of metabolomics analysis of cancer cell lines could be established to validate the result generated by proteomics analysis. Also, the metabolomics analysis itself could provide a biomarker candidate list in the metabolite aspect.

Third, for the in-gel MAAH method developed in Chapter 5, future work will involve the determination of chemical structures of interfering chemicals to improve the overall detection sensitivity by developing an acid-resistant gel and/or means of separating the peptides from background chemicals prior to MS analysis.

Fourth, once the background interference issue is solved, in-gel MAAH method with higher sensitivity could be further applied for sequence mapping of complex protein mixture separated by 2D gel electrophoresis such as whole cell lysate of cancer cell line or tissue. Also, the modification of proteins could be identified without enrichment, such as protein PEGylation which is usually designed for protein drug delivery.[12]

6.1 Literature cited

1. Poetsch, A.; Wdlters, D., Bacterial membrane proteomics. *Proteomics* **2008,** *8* (19), 4100-4122.

2. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009,** *11*, 49-79.

3. Chen, E. I.; Hewel, J.; Felding-Habermann, B.; Yates, J. R., III, Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). *Mol. Cell. Proteomics* **2006,** *5* (1), 53-56.

4. Wu, F.; Sun, D. F.; Wang, N.; Gong, Y.; Li, L., Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis. *Anal. Chim. Acta* **2011,** *698* (1-2), 36-43.

5. Smith, J. A., Electroelution of proteins from stained gels. *Current Protocols in Immunology / edited by John E. Coligan* **2002,** *Chapter 8*, Unit 8.8.

6. Sun, D. F.; Wang, N.; Li, L., Integrated SDS removal and peptide separation by strong-cation exchange liquid chromatography for SDS-assisted shotgun proteome analysis. *J. Proteome Res.* **2012,** *11* (2), 818-828.

7. Meyer, B.; Papasotiriou, D. G.; Karas, M., 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2011,** *41* (2), 291-310.

8. Zhu, K.; Kim, J.; Yoo, C.; Miller, F. R.; Lubman, D. M., High sequence coverage of proteins isolated from liquid separations of breast cancer cells using capillary electrophoresis-time-of-flight MS and MALDI-TOF MS mapping. *Anal.*

Chem. **2003,** *75* (22), 6209-6217.

9. Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G., Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **2007,** *4* (10), 798-806.

10. Larsen, M. R.; Trelle, M. B.; Thingholm, T. E.; Jensen, O. N., Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* **2006,** *40* (6), 790-798.

11. Bergquist, J.; Andersen, O.; Westman, A., Rapid method to characterize mutations in transthyretin in cerebrospinal fluid from familial amyloidotic polyneuropathy patients by use of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chem.* **2000,** *46* (9), 1293-1300.

12. Harris, J. M.; Chess, R. B., Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discovery* **2003,** *2* (3), 214-221.