MULTIPLEXED FRAGMENTATION AND PROTEIN

INTERACTION REPORTER TECHNOLOGY

APPLICATION TO HUMAN CELLS

By

HYE IN NAM

A dissertation submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE IN CHEMISTRY

WASHINGTON STATE UNIVERSITY
Department of Chemistry

AUGUST 2009

To the Faculty of Washington State University:

    The members of the Committee appointed to examine the

thesis of HYE IN NAM find it satisfactory and recommend that it be accepted.

_____

James E. Bruce, Ph.D., Chair

_____

Kenneth L. Nash, Ph.D.

_____

ChulHee Kang, Ph.D.

ACKNOWLEDGMENT

Dr. Bruce has been instrumental in ensuring my academic, professional, financial and moral well being ever since I joined his lab. In every sense, none of this work would have been possible without him. Many thanks also go to my committee members Dr. Kenneth L. Nash and Dr. ChulHee Kang for their valuable time and suggestions, which significantly improved the quality of this dissertation. Many people have assisted me in so many ways during my work at WSU. I would like to thank my friends and colleagues.

I am grateful to my father and mother. They always respect my decision on every moment in my life. My sister, Hye Won and brother-in-law, Young hoon Kim are always my close friends and mentor. I would like to thank my father-in-law and mother-in-law. They always pray for me and believe in me.

My final, and most heartfelt, acknowledgment is to my husband Chang hun You. His support, encouragement, and companionship have turned my journey through graduate school into pleasure.

MULTIPLEXED FRAGMENTATION AND PROTEIN

INTERACTION REPORTER TECHNOLOGY

APPLICATION TO HUMAN CELLS

Abstract

by HYE IN NAM, MS
Washington State University
August 2009

Chair: James E. Bruce

Chemical cross-linking combined with mass spectrometry is a feasible approach to study the protein-protein interactions. However, system complexity and cross-linking product heterogeneity have precluded widespread chemical cross-linking use for large-scale identification of protein-protein interactions. Mass spectrometry identifiable cross-linkers protein interaction reporters (PIRs) were developed to overcome these problems. PIR couple with mass spectrometry was applied to study the protein-protein interactions of prokaryotic cells, such as *Shewanella oneidensis MR-1*. However, the biological features of eukaryotic cells are different with prokaryotic cells. In this report, PIR coupled with mass spectrometry was applied to human cancer cells, such as MCF7 and HeLa cells. 109 proteins in MCF7 and 127 proteins in HeLa were labeled with PIRs and identified.

The bottom-up LC/MS method involves the protein denaturation, protease digestion and LC/MS/MS to identify proteins and peptides. The most common data acquisition method for LC/MS/MS, can allow the identification of relatively abundant proteins. In biological systems, functional proteins, such as enzymes and membrane proteins are of relatively low abundance,

thus data-dependent acquisition is generally insufficient to identify many functional proteins. Multiplexed fragmentation was developed. Multiplexed fragmentation with PIR strategy is a potential approach to study protein-protein interactions. However, ambiguous identification of PIR labeled peptides is the major challenge. To overcome this problem, this report demonstrated the feasibility of LC/3-stage multiplexed fragmentation approach to study protein-protein interactions with PIRs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Protein-Protein Interaction

Proteins can be grouped into two groups: structural and functional proteins. Structural proteins provide the structural base to cells and organisms. For instance, proteins in cytoskeletons are used to maintain cell shape. Functional proteins are related to specific cellular functions, such as enzyme catalysis, cell signaling and membrane transportation. Most of the functional proteins cannot work alone. They usually form protein complexes or interact with other proteins to conduct particular functions.

The interaction between proteins is important for many cellular functions. Protein modifications are strongly involved with protein-protein interactions. The representative examples are the interactions of protein kinases or protein phosphatases with their target proteins. When these enzymes modify their substrate proteins, interactions between enzymes and substrates are triggered [1]. Protein-protein interactions are also related with many diseases such as cancers [2, 3] and Alzheimer's disease [4, 5]. Therefore, information about protein-protein interactions better explains their biological functions and help us understand the diseases process.

Methods of Protein-Protein Interaction

There are various approaches to investigate protein-protein interactions. These techniques are divided into three categories. Each category has own strengths and weaknesses. This section discusses both aspect of each approach.

Category 1 includes the techniques requiring some manipulation at the genetic level, such

1

as yeast two-hybrid [6], split ubiquitin [7], in-vivo fluorescence resonance energy transfer (FRET) [8], tandem affinity purification (TAP) tag [9]. These techniques use fusion proteins, which are produced by the fusion of the foreign piece to the protein of interest. The fusion protein is used as the bait. When the protein is bound and/or interacts with the bait, the detectable signal is produced and the certain protein-protein interaction is recognized. The strengths of this category are ease of scaling up, good sensitivity, high throughput, and semi-automated experimental procedures. These methods are necessary to generate many yeast strains with genetic pieces coding for the desired attachment combined to every single open reading frame (ORF), require many pair wise screening tests to discover interactions, and to isolate and sequence DNA for identifying interacting partners. In addition, these approaches are inclined to produce the false positive and false negative results due to the fusion of foreign pieces to protein of interest. The fusion of foreign pieces to proteins of interest can change the protein tertiary structure. The structural change can make different proteins bind to the fused proteins. When the foreign protein are large enough to cover regions of the protein where a protein-protein interaction takes place, the protein-protein interaction can neither occur nor be detected. Moreover, the protein-protein interaction cannot be detected if the interaction is related to or requires the presence of post-translational modifications (PTMs) and/or enzymes, each of which does not originally exists in yeast cells.

Category 2 contains the approaches that do not require genetic manipulation. Protein-protein interactions are detected at the protein level without any changes that are introduced on the genetic level. The representative examples are protein chip [10, 11], co-immunoprecipitation (Co-IP), and Pull-down assays [12]. These approaches are based on antibody-antigen interactions. The most popular protein chip is the antibody array. Antibodies are spotted on the protein chip

and are used as bait to detect proteins from cell lysate. In CO-IP, an antibody is used to isolate the protein of interest bound with interacting partners. These techniques are high throughput and allow us to detect interactions related with PTMs. The proteins are dissolved in a certain buffer to performing the experiments. In this case, the folding of proteins can be changed. These structural changes of the proteins can produce false positive and false negative results. Also, common buffer solutions contain the detergent. The detergents in the buffer solution can affect the protein-protein interactions. Especially, the association of the detergents with hydrophobic proteins can prevent the interaction being present *in vitro*. The major disadvantage is the approaches require certain antibodies. If the protein of interest does not interact with any antibody, the protein-protein interaction can not be detected.

Category 3 includes chemical cross-linking coupled with mass spectrometry [13-17]. This approach is based on the close proximity of the binding partners. The proteins are labeled with the chemical cross-linker *in vivo*. The labeled proteins are purified, digested and identified by the mass spectrometry. The potential advantages of this method are high throughput and large-scale discovery. They can also allow discovery of protein-protein interactions without any biological hypothesis and thus, presents an unbiased approach. The major challenge is the system complexity. Various types of the cross-linked products can be produced in the cross-linking reactions, such as, dead-end, intra-, and inter-cross-linked peptides, and unlabeled peptides. The inter-cross-linked products are relatively small amount in the reaction mixture. Therefore, the detection and identification of the inter-cross-linked products are the great challenges [16-18]. The novel types of strategies [19, 20] and cross-linkers have been developed to overcome these limitations, such as isotopic labeled cross-linkers [21-23], cross-linkers contained affinity tags[17, 18, 24, 25] and mass spectrometry cleavable (MS cleavable) bonds[16-18, 26, 27].

Protein Interaction Reporter (PIR)

Protein Interaction Reporter (PIR) [14, 18] is a type of chemical cross-linker that contains an affinity tag and MS cleavable bonds. The low energy collisions with tandem mass spectrometry can break the MS labile bond in the cross-linkers. Then, the cross-linked products release the reporter ions and the modified peptides without disrupting peptide backbones. Figure 1-1 shows the connectional molecular design of PIRs.

Different types of the cross-linked products can generate the different fragmentation patterns. The mathematical relationships between masses of fragments are illustrated in Figure 1-2. The relationships are used to obtain the unambiguous identification of the cross-linked products [18]. Practically, the masses of the cross-linked products (m_c), are obtained during MS scan. On the contrary, masses of the reporter ions (m_r) and modified peptides (m_{p1}, m_{p2}) are measured during low energy MS/MS scan.

PIR coupled with Mass Spectrometry to Study Protein-Protein Interactions

The feasibility of PIR coupled with mass spectrometry to study protein structures and protein-protein interactions is established using the standard protein, *Shewanella oneidensis MR-1*, and Gram-negative anaerobic bacteria [14, 16-18]. But there have been few studies of the use PIR in higher organisms like *Homo sapiens*. Human cells and bacterial cells are clearly different. The former is a large eukaryotic cell and the latter is a small prokaryotic cell shown in Table 1-1 [28]. The cell size of each cell is shown in Figure 1-3. The size of typical human cells is approximately 10μm, but the size of major bacterial cells is 1 to 5 μm. For example, *Escherichia coli* is approximately 2μm.

4
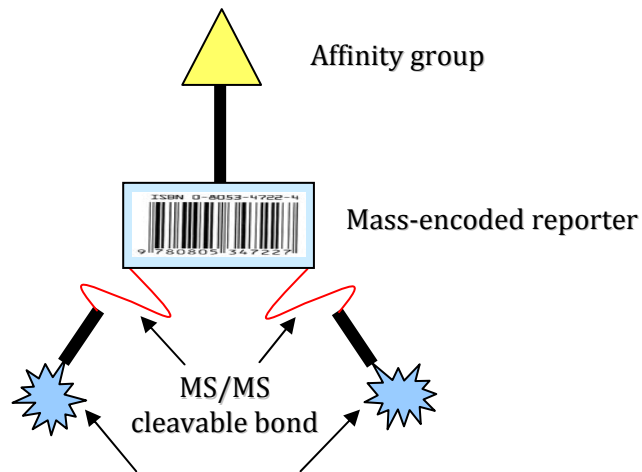
Figure 1-1. Conceptual modular design of Protein interaction reporter (PIR).
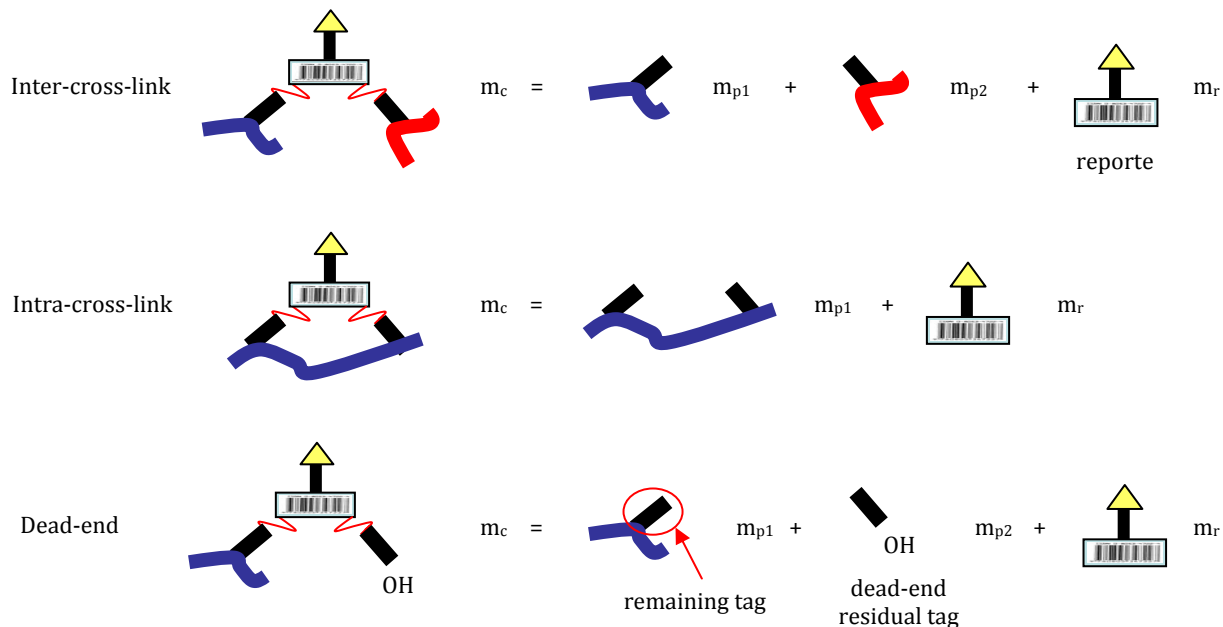
Figure 1-2. Specific fragmentation patterns of PIR labeled peptides helps distinguish dead-end, intra-, and inter-cross-linked peptides. The neutral mass of the precursor ion equals the sum of the neutral masses of its product ions [18].
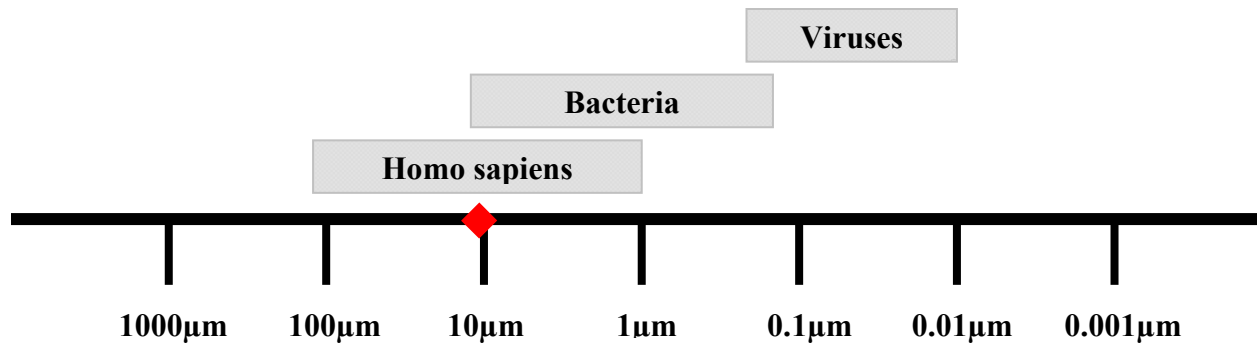
Figure 1-3. Size range of Homo sapiens, bacteria and viruses cells. The diamond (♦) represents the size of typical Homo sapiens cell.

Table 1-1. Comparison of Prokaryotic and Eukaryotic Organisms [28].

	Prokaryotes	Eukaryotes
Organisms	bacteria and cyanobacteria	protists, fungi, plants and animal
Cell size	generally 1 to 10 μm in linear dimension	generally 5 to 100 μm in linear dimension
Organelles	few or none	nucleus, mitochondria, chloroplasts, endoplasmic reticulum, etc.
DNA	circular DNA in cytoplasm	very long linear DNA molecules containing many noncoding regions; bounded by nuclear envelope
RNA	RAN and protein synthesized in same compartment	RAN synthesized and processed in nucleus; proteins synthesized in cytoplasm
Cytoplasm	no cytoskeleton; cytoplasmic streaming, endocytosis and exocytosis all absent	cytoskeleton composed of protein filaments; cytoplasmic streaming; endocytosis and exocytosis
Cell division	chromosomes pulled apart by attachments to plasma membrane	chromosomes pulled apart by cytoskeletal spindle apparatus
Cellular organization	mainly unicellular	mainly multicellular, with differentiation of many cell types

Liquid Chromatography/Tandem Mass Spectrometry (LS/MS/MS) for Proteome Research

LC/MS is a general method in the study of proteomics, such as the detection and identifications of proteins in complex mixtures like cell lysate. The bottom-up LC/MS method involves the protein denaturation, protease digestion and LC/MS/MS to identify proteins and peptides [29]. There are two necessary processes to accomplish LC/MS/MS experiments for complex protease digests. One is that mass spectra need to be analyzed in real time and the ions should be selected for collision-induced dissociation (CID). This LC/MS/MS approach is called data-dependent data acquisition [30, 31]. The advantages of this approach are high speed, high throughput and not requiring a highly accurate instrument. This method can provide information about relatively abundant proteins. In biological systems, functional proteins, such as enzymes and membrane proteins, are present at relatively low abundance. Thus, data-dependent acquisition method is generally insufficient to identify protein-protein interactions of many functional proteins. Novel MS/MS acquisition methods, such as multiplexed fragmentation method [32-34] and MS^E have been developed to overcome this problem [35, 36]. These methods are achieved by the parallel alternating scans which are acquired at either low or high collision energy.

Purposes of This Study

PIR coupled mass spectrometry is applicable to study protein-protein interactions for bacteria, such as *Shewanella oneidensis MR-1* and Gram-negative anaerobic bacteria [16-18]. In this report, PIR coupled with mass spectrometry was applied to proteome study of human cell lines to verify the feasibility of this method for studying both eukaryotic and prokaryotic cells.

As discussed in the previous section, the data-dependent LC/MS/MS technique has a

limitation. When PIR experiments are carried out, inter-cross-linked product are often of relatively low abundance. Multiplexed fragmentation approach was applied to identify low abundant species. The identification of modified peptide is performed by peptide mass fingerprinting (PMF) using mass of modified peptides obtained at high collision energy. The major challenge of this approach relates to unambiguous identification by peptide mass alone. In this report, 3-stage multiplexed fragmentation approach with PIR experiments was developed and demonstrated to help overcome this problem.

CHAPTER TWO

EXPERIMENTAL SECTION

Chemicals and Reagents

All chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA) unless otherwise noted. Fluorenylmethyloxycarbonyl-RINK (Fmoc-RINK), 4-hydroxymethyl-3-methoxyphenoxybutyric acid MBHA (HMPB-MBHA) resins, Fmoc-Arg(Pbf)-NovaSyn TGT resin, and Fmoc-protected amino acids were purchased from Novabiochem (San Diego, CA, USA). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Water used for preparing solutions and solvents was 18-MΩ deionized water produced with a Barnstead Nanopure Water System.

Synthesis of Cross-linkers

PIRs were synthesized using a 431A Peptide Synthesizer (Applied Biosystem, Foster City, CA USA). First, Fmoc-Lys(biotinyl-e-aminocaproyl)-OH was coupled to 4-hydroxymethyl-3-methoxyphenoxybutyric acid MBHA resin (HMPB-MBHA resin) using the standard symmetric anhydride method. Then, Fmoc-Lys(Fmoc)-OH was coupled to lysine and formed a branch point for PIRs. Two Fmoc-Pro-OH and Fmoc-Asp-OH were sequentially coupled to lysine to synthesize the first PIR (PIR I). Carboxyl groups were then introduced by reacting the primary amines of proline with succinyl anhydride. Subsequently, the two carboxyl groups were activated by forming the esters with N-hydroxysuccinimide (NHS). To synthesize the second PIR (PIR II), Two RINK groups were coupled to Lys by treating the Fmoc-RINK linkers. Then, succinic anhydride and NHS were coupled as described previously.

9

The crude product was cleaved from the superacid sensitive resin with 0.5% TFA in chloroform [37]. The chloroform and TFA pyridine salts were removed under a vacuum. The synthesized PIRs were purified using reversed-phase HPLC and the final product had a purity of approximately 90%.

2-1. Protein Interaction Reporter Technology Application to Human Cells

Cell Culturing, Harvesting and Labeling

HeLa and MCF7 cell lines were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). Eagle's minimum essential medium (EMEM), newborn Bovine Calf Serum (NBCS) and trypsin-EDTA were purchased from HyClone (Logan, UT, USA). Fetal bovine serum (FBS) and Penicillin-Streptomycin were purchased from GIBCO Invitrogen (Carlsbad, CA, USA).

HeLa cells were cultured in EMEM supplemented with 5% FBS, 5% NBCS and 1% Penicillin-Streptomycin. MCF7 cells were cultured in EMEM supplemented with 10% FBS, 1% Penicillin-Streptomycin and 0.1% bovine insulin. The cells were grown in a humidified chamber maintained at 37°C and 5 % carbon dioxide (CO_2). After reaching 80% confluence, the culture medium was removed, and the cells were washed with phosphate buffered saline (PBS) at pH 7.4. Then, trypsin-EDTA was added for approximately 3 minutes to detach the cells. The culture medium was added to quench trypsin-EDTA digestion. Detached cells were washed 3 times with PBS. To perform the cell labeling, the cells were reacted with 1mM PIR II in PBS for 1 hour at 4°C. PIR labeled cells were washed 3 times with PBS at room temperature (RT) to remove unreacted PIRs.

Microscopy Sample Preparations

MCF7 cells were grown onto the sterilized microscopic cover glass at 37°C and 5 % CO_2 until 80% confluence and washed with PBS by gently shaking. The cells were reacted with 1mM PIR II in PBS for 1 hour at 4°C. PIR labeled cells were treated with 3.7% formaldehyde in PBS for 30 minutes at RT and washed 3 times with PBS. Then, the cells were permeabilized by incubation with 1mL of 0.2% triton X-100 for 5 minutes at RT and washed 3 times with PBS. To reduce non-specific binding, the cells were blocked with 6% solution of Bovine Serum Albumin (BSA) in PBS for 30 minutes at RT and washed 3 times with PBS. The cells were incubated with 1:50 dilution of anti-biotin antibody (Sigma C7653) in 1% solution of BSA in PBS (BSA-PBS) for 30 minutes. Next, the cells were washed 3 times with BSA-PBS and incubated with 10ug/ml solution of Alexa Fluor 488 rabbit anti-mouse IgG (moluculare probes, A-11054) enough to cover all cells for 30 minutes at RT. The cells were washed 3 times with 1ml of BSA-PBS and incubated with 10ug/ml Alexa Fluor 488 goat anti-rabbit IgG enough to cover all cells for 30 minutes at RT. The cells were washed 3 times with 1mL of BSA-PBS. The cells are treated with propidium iodide to visualize the nuclei to prepare microscope slide. All images were acquired using the Zeiss LBS 510 Laser Confocal Microscope at the WSU Microscopy Center.

Sample Preparation for Identification of PIR Labeled Proteins

The unlabeled and labeled cells were lysed in Radio-Immunoprecipitation Assay (RIPA) buffer by a vortex mixer. The cell lysates were centrifuged at 1800 rpm for 45 minutes at 4°C to remove debris, which are broken cell membranes. Bradford assay was performed to determine protein concentrations.

The supernatants of cell lysates were divided into two portions. One portion was used for

SDS-PAGE and anti-biotin western blot analysis. The other was used to identify PIR labeled proteins. Monomeric avidin beads (Pierce Rockford, IL, USA) were added into the cell lysate and incubated for 12 hours 4°C to enrich the PIR labeled proteins. The beads were washed 3 times with PBS for reducing non-specific binding. The proteins bound on the beads were eluted by boiling with Laemmili sample buffer (Bio-Rad Laboratories Inc. Hercules, CA, USA) and separated by SDS-PAGE. The gels obtained from SDS-PAGE was used to identify the PIR labeled proteins

SDS-PAGE and Anti-Biotin Western Blot Analysis

Cell lysate was mixed with 1:1 ratio with Laemmli sample buffer that was prepared according to manufacture's instructions. The mixture was heated at 95°C for 10 minutes to reduce disulfide bonds. Approximately 20ug of protein was as loaded onto SDS-PAGE for separation by the gel with 5% stacking buffer and 8% resolving buffer. SDS-PAGE was performed under constant voltage condition at 180 V. The gels were washed 3 times with deionized water (DI water) for 20minutes at RT to remove detergents in the running buffer. Then, the gel was stained with Coomassie Blue R250 (Bio-Rad, Hercules, CA, USA) for 1hour at RT. The gel was washed with DI water overnight. Finally, the gels were imaged with a Densitometer (Molecular Dynamics, Sunnyvale, CA, USA).

To acquire anti-biotin western blot images, the proteins separated by SPS-PAGE were transferred on the nitrocellulose membrane using transfer buffer by trans-blot SD semi-dry transfer cell (Bio-Rad laboratories Inc. Hercules, CA, USA). The blotting was performed for 2 hours at 15V. The membrane was air-dry overnight to fix the proteins. Then, the membrane was blocked in 15% (w/v) dry milk Tris Buffered Saline (TBS) buffer for 1 hour, incubated with

mouse anti-biotin antibody at 1:10000 dilution in 15% (w/v) dry milk TBS buffer, and washed 3 times with TBS buffer. The membrane was incubated with goat anti-mouse HRP conjugate antibody (Bio-Rad laboratory inc 170-5047) at 1:20000 dilution in 15% (w/v) dry milk TBS buffer for 1 hr, and washed 3 times with TBS buffer. The membrane was reacted with Chemoluminescent Peroxidies Substrate Kit (Sigma CPS-1-60) for approximately 1 minute to produce chemoluminescence. KODAK film for X-ray was exposed to the membrane and automatically processed. The exposed KODAK film was imaged with a Densitometer.

In-Gel Digestion

Gel bands of interest were excised and distained 50% methanol and 5% acetic acid overnight. The gel pieces were washed by acetonitrile (ACN), and completely dried in SpeedVac. The disulfide bonds of proteins in gel pieces were reduced by 10mM dithiothreitol (DTT) in 100mM ammonium bicarbonate (Ambic) for 30 minutes at RT. The reduced disulfide bonds were alkylated by 50mM iodoacetamide (IAA) in 100mM Ammonium bicarbonate for 30 minutes at RT in the dark. The gel was washed with100mM Ambic and ACN, and completely dried in SpeedVac. The gel pieces were hydrated by 20ng/µL trypsin in 50mM Ambic for 10 minutes on the ice. After the hydration of gel pieces, the excess of trypsin solution was removed. The gel was incubated at 37°C overnight. After in-gel digestion, the samples were subjected to nano LC/MS/MS.

Nano LC/MS/MS and Data Analysis for Protein Identification

Data-dependent nano LC/MS/MS of in-gel digest was performed using ion trap mass spectrometer (Esquire HCT, Bruker Daltonics, Billerica, MA, USA) equipped with a nano ESI

source and nano HPLC systems (Ultimate, Dionex, Sunnyvale, CA, USA). The peptides were injected by autosampler and trapped into the precolumn. The peptides were eluted from precolumn by washing with solution A (0.1% formic acid, 2% ACN in water) at a flow rate of $50\mu L/min$ for 3 minutes. These peptides were trapped into nano C18 reversed phase column (C18 PepMap, $75\mu m$ X 150mm, $3\mu m$, 100Å, Dionex, Sunnyvale, CA, USA) Then, the peptides were eluted using the following gradient: 0% Solution B (0.1% formic acid, 98% ACN in water) for 0-3 minutes, 0-15% Solution B for 3-15 minutes, 15-25% Solution B for 15-60 minutes, 25-40% Solution B for 60-90 minutes, 95% Solution B for 90-105 minutes and 0% Solution B for 135 minutes.

Data analysis was performed using Bruker Daltonics DataAnalysis software (version 3.1, Billerica, MA, USA). Proteins were identified by searching against NCBI human protein database using Mascot44 (version 2.1.0, MatrixScience Ltd., London, UK) licensed in house. Database search parameters were used as previously reported [16, 17] Database search parameters were shown in Figure 2-1. The auto hits option was selected to allow reporting of all the protein hits with the probability-based mascot scores that exceeded their thresholds ($p < 0.05$), indicating significance above the 95% confidence level. The subcellular locations of identified proteins were obtained from DAVID Bioinformatics Resources of National Institute of Allergy and Infectious Diseases (NIAID, http://david.abcc.ncifcrf.gov/home.jsp).

Figure 2-1. MASCOT search parameters for protein identification. Database search parameters are briefly: search database, NCBI; toxonomy, *Homo sapiens*; enzyme, trypsin; allowed missed cleavages, up to 3; fixed modifications, carbamidomethyl (C); variable modifications, oxidation (M); peptide tolerance, 1.6 Da; and MS/MS tolerance, 0.8 Da.

2-2. Multiplexed Fragmentation

Synthesis of Model Peptides

Model peptides were prepared in-house based on F-moc solid phase peptide synthesis chemistry. Fmoc amino acids were sequentially coupled to Fmoc-Arg(Pbf)-NovaSyn TGT resin (Novabiochem, San Diego, CA, USA). The first model peptide (Peptide I) was synthesized by sequential coupling with Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Phe-OH, Fmoc-Val-OH, Fmoc-Arg(Pbf)—OH, Fmoc-Ser-OH, and Fmoc-Val-OH. The sequence of Peptide I is Val-Ser-Arg-Val-Phe-Leu-Lys-Gly-Ala-Arg (VSRVFLKGAR).

The other peptide (Peptide II) was synthesized by sequential coupling with Fmoc-Gln-OH, Fmoc-Asn-OH, Fmoc-Glu-OH, Fmoc-Phe-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Gly-OH, Fmoc-Pro-OH, Fmoc-Leu-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser-OH, and Fmoc-Val-OH. The sequence of Peptide II is Val-Ser-Arg-Leu-Pro-Gly-Leu-Lys-Phe-Glu-Asn-Gln VSRLPGLKFENQR).

The crude product was treated with acetic anhydride to protect N-terminus of model peptide. Acetic acid and extra acetic anhydride were removed under a vacuum. Then, peptides were treated with TFA in chloroform, and then neutralized with pyridine to cleave the resin and protecting groups on Lys and Arg. The chloroform and TFA pyridine salts were removed under a vacuum. The synthesized peptides were purified using reversed-phase HPLC and the final product had a purity of approximately 90%. The schematic mechanism of peptide synthesis is summarized in Figure 2-2. The molecular structures of Peptide I and II are shown in Figure 2-3.

1st step: activation of Fmoc amino acids
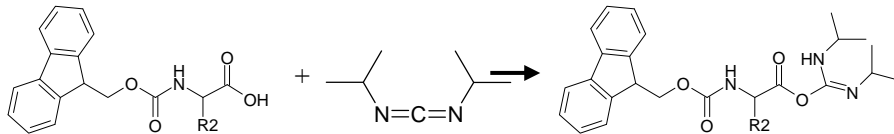
2nd step: coupling reaction between activated Fmoc amino acid and amino acid resin
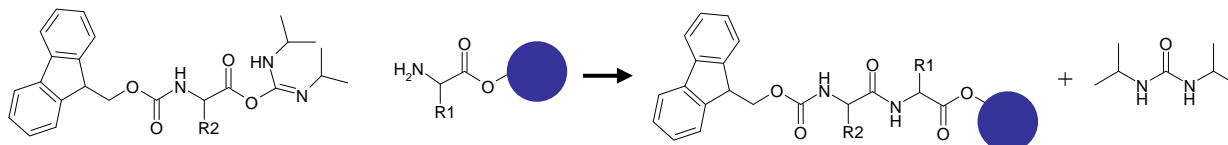
3th step: deprotection of Fmoc group
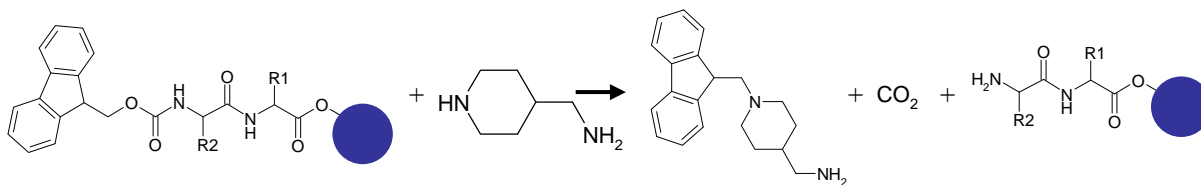
4th step: acetylation of N-terminus

5th step: removal of resin
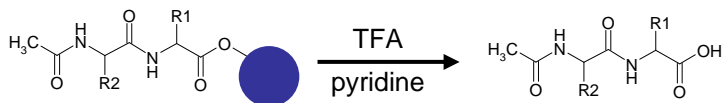
Figure 2-2. Schematic mechanisms of peptide synthesis. The solid circle (●) represents the superacid sensitive resin

(a) Peptides I
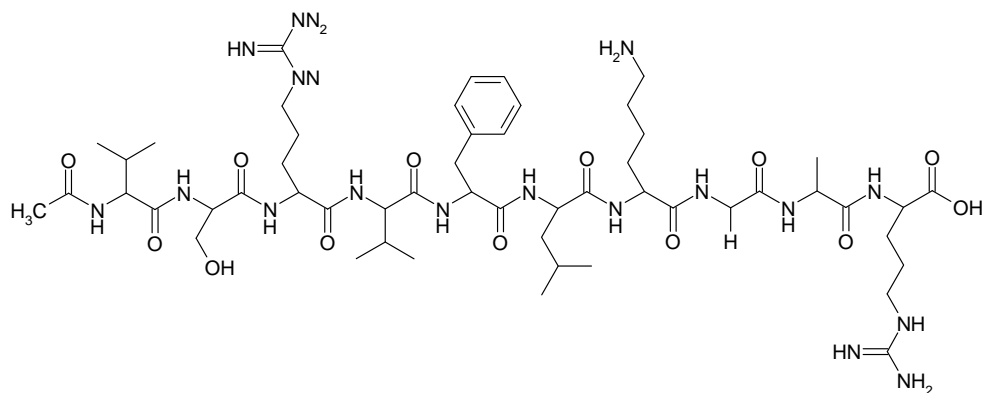
(b) Peptides II
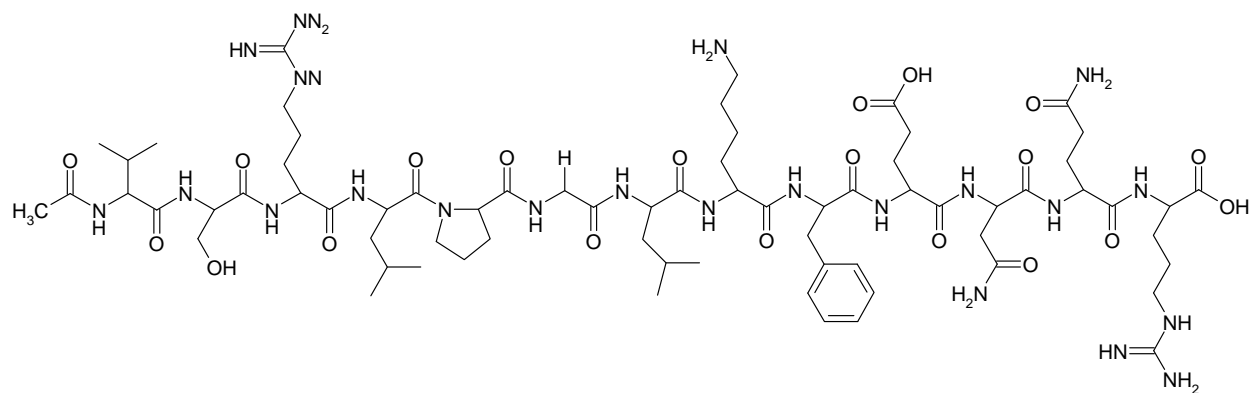
Figure 2-3. Molecular structures of Peptide I and Peptide II. (a) Peptide I; Ac-VSRVFLKGAR; monoisotopic mass: 1173.70 Da, (b) Peptide II; Ac-VSRLPGLKFENQR, monoisotopic mass: 1584.88 Da.

Cross-linking Reaction and In-Solution Digestion

Peptides I and II were completely dried in SpeedVac overnight. Peptides were mixed with 2:1 ratio with Disuccinimidyl suberate (DSS), or PIRs that was dissolved in dried DMSO. The reaction mixture was incubated at RT overnight. Without quenching the reaction, DMSO was removed in SpeedVac overnight. The labeled peptides were suspended in 100mM Ambic. Reduction of disulfide bonds were sequentially performed using DTT in 100mM Ambic for 30 minutes at 37°C. Then, alkylation was carried out adding IAA in 100mM Ambic for 30 minutes at 37°C in the dark. Trypsin in 50mM Ambic was added into the mixture, and incubated at 37°C overnight. The digests were subjected to the multiplexed fragmentation experiments. The efficiencies of cross-linking reaction and trypsin digestion were determined by taking Matrix-assisted laser desorption/ionization – time of flight mass spectra (MALDI-TOF MS, Omni Flex MALDI-TOF mass spectrometer, Bruker Daltonics, Billerica, MA, USA).

Instrument

Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (FTICR-MS, Apex-Q FTICR MS, Bruker Daltonics, Billerica, MA, USA) with a 7.0 T superconducting magnet was used for developing multiplexed fragmentation approach with PIR labeled peptides. Schematic diagram of Apex-Q 7T FTICR MS is shown in Figure 2-4. A 2.5µL syringe pump was used to introduce sample solution to electrospray ionization (ESI) source, which was the choice of ionization source, at a flow rate of 30µL/hour. The voltage on ESI source was set at 1800V. The other voltages applied to each segment of FTICR-MS were shown in Figure 2-5. The instrument was operated using Bruker XMASS software version 7.0.6 (Bruker Daltonics, Billerica, MA, USA).
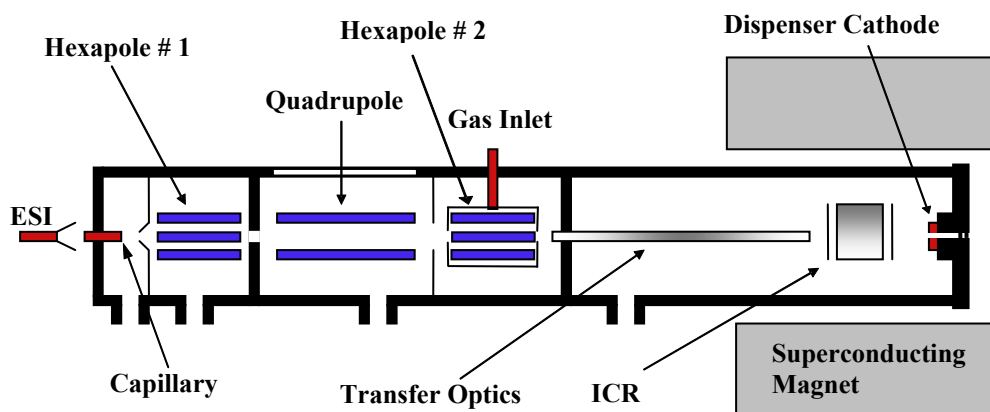
Figure 2-4. Schematic diagram of Apex-Q 7T Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (Apex-Q 7T FTICR MS).
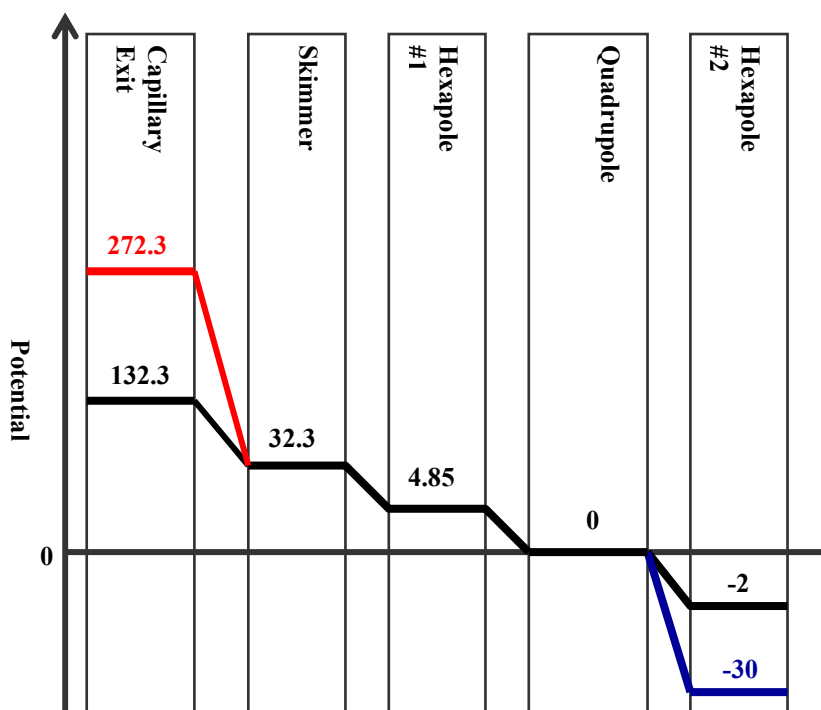
Figure 2-5. Multiplexed fragmentation method in Apex-Q 7T FTICR MS. The black line indicates the potential applied to each region of the mass spectrometer for MS scan. ISCID scan is shown by the red line and the double activation scan is shown by the red and blue line.

In-Source Collision-Induced Dissociation (ISCID) and Double Activation Analysis

In-Source Collision-Induced Dissociation (ISCID) was used to release report ions and modified peptides from cross-linked products. ISCID experiments were performed by changing the potential difference between Capillary Exit and Skimmer. The fragments of modified peptides are generated by the double activation. The double activation experiments were achieved by combining ISCID and collision induced dissociation (CID) without the ion isolation [38]. As explained in the instrument section, Apex-Q 7T FTICR MS was operated in positive ESI mode with an applied voltage of 1800 kV. During MS, ISCID, and double activation scanning, data were acquired for 15 scans (approximately 15 s). Figure 2-5 shows the voltages setting to perform MS, ISCID, and the double activation.

LC/Multiplexed Fragmentation Experiment and Data Analysis

LC/multiplexed fragmentation experiments were performed using ESI Apex-Q 7T FTICR MS with equipped with a nano HPLC systems (Ultimate, Dionex, Sunnyvale, CA, USA). HPLC was operated in the same manner as explained previously. Various cross-linked species were introduced by autosampler, and separated by nano C18 reversed phase column using the gradient elution. The eluted species from the HPLC system were directly introduced by ESI Apex-Q 7T FTICR MS system to perform multiplexed fragmentation experiment. Apex-Q 7T FTICR MS was operated using Bruker XMASS software version 7.0.6. The applied voltage on each segment of the instrument was automatically changed by Bruker XMASS software to acquire 3-stage multiplexed fragmentation data. During the data acquisition, data were acquired for 4 scans (approximately 2seconds). LC/multiplexed fragmentation data were analyzed using ICR-2LS software [37].

False Discovery Analysis of Multiplexed Fragmentation

All simulations of false discovery rates for PIR relationship differentiation and accurate peptide mass-based protein identification were carried out using the X-links application. The term "false discovery" is used to indicate the identification of peptide masses without other information. The application X-links that was developed using Microsoft's Visual Basic at Environmental Molecular Sciences Laboratory (EMSL), Pacific Northwest National Laboratory (PNNL) [18]. The genome of *Escherichia coli K12* was downloaded from NCBI web site (http://www.ncbi.nlm.nih.gov/) and was used with the X-links application to calculate all possible tryptic peptides with variable allowed missed cleavages in the mass range from 500 to 5000 Da. Full database consists all possible tryptic peptides. Restricted database is the list of the tryptic peptides having at least one internal Lys.

The restricted database was used to verify the identification of modified peptides using 3-stage multiplexed fragmentation data. The term "candidate peptide" is used to indicate tryptic peptide having exactly same mass with Peptide I and II. The candidate peptides were selected from the restricted database. Then, the theoretical MS/MS patterns of these candidates were generated using GPMAW software version 6.00 (Lighthouse Data, Odense, Denmark).

CHAPTER THREE

RESULTS AND DISCUSSION

Protein Interaction Reporter (PIR)

PIR is a homobifunctional, MS cleavable cross-linker. It contains an amine-reactive NHS ester at each end of spacer arm. NHS esters react with primary amines at pH 7-9 to form stable amide bonds, along with release of the N-hydroxysuccinimide leaving group. Proteins generally have several primary amines in the side chain of Lys residues and the N-terminus of each polypeptide that are available as targets for NHS-ester cross-linker. The reaction mechanism between primary amines and NHS esters is shown in Figure 3-1. As explained in the introduction, PIR contains an affinity tag, and MS cleavable bonds. The structures of two PIRs (PIR I and PIR II) are shown in Figure 3-2. A tag, Biotin, is used for the affinity isolation of labeled proteins. Biotin group is specifically bound to avidin, which is a tdrametic protein produced in the oviducts of birds [39]. Two PIRs used for this research havs two different MS cleavable bonds. RINK groups were utilized to make PIR I [14. 37, 40]. Asp-Pro (DP) amino acids are seed to form MS cleavable bonds of PIR II [41, 42]. The fragmentation mechanisms of PIRs are shown in Figure 3-3.
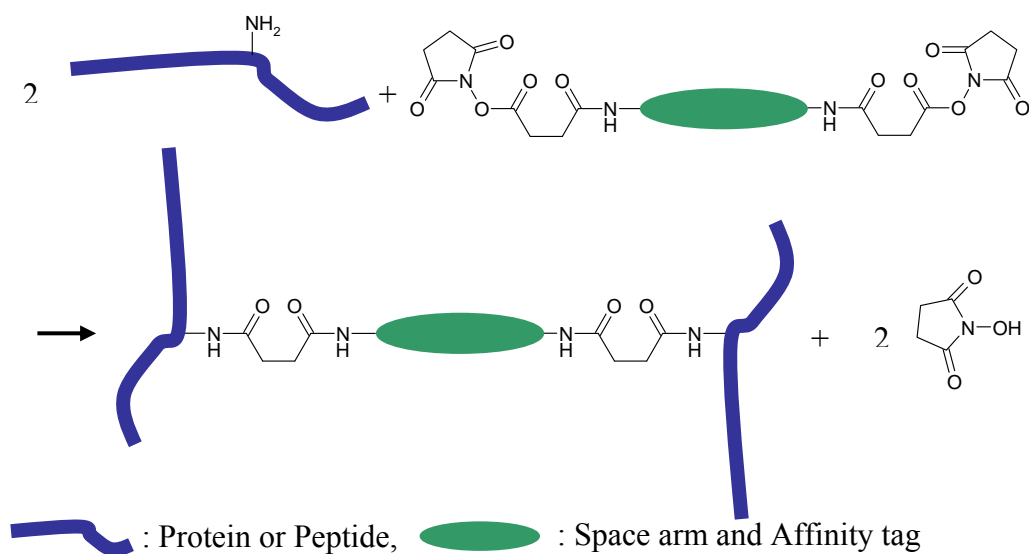
Figure 3-1. Cross-linking reaction mechanism between the primary amine and NHS ester.

PIR I
Biotin-DP-NHS
Formula: $C_{57}H_{77}N_{13}O_{25}S$
Monoisotiopic Mass: 1375.49

PIR II
Biotin-Rink-NHS
Formula: $C_{74}H_{91}N_{11}O_{24}S$
Monoisotiopic Mass: 1548.59
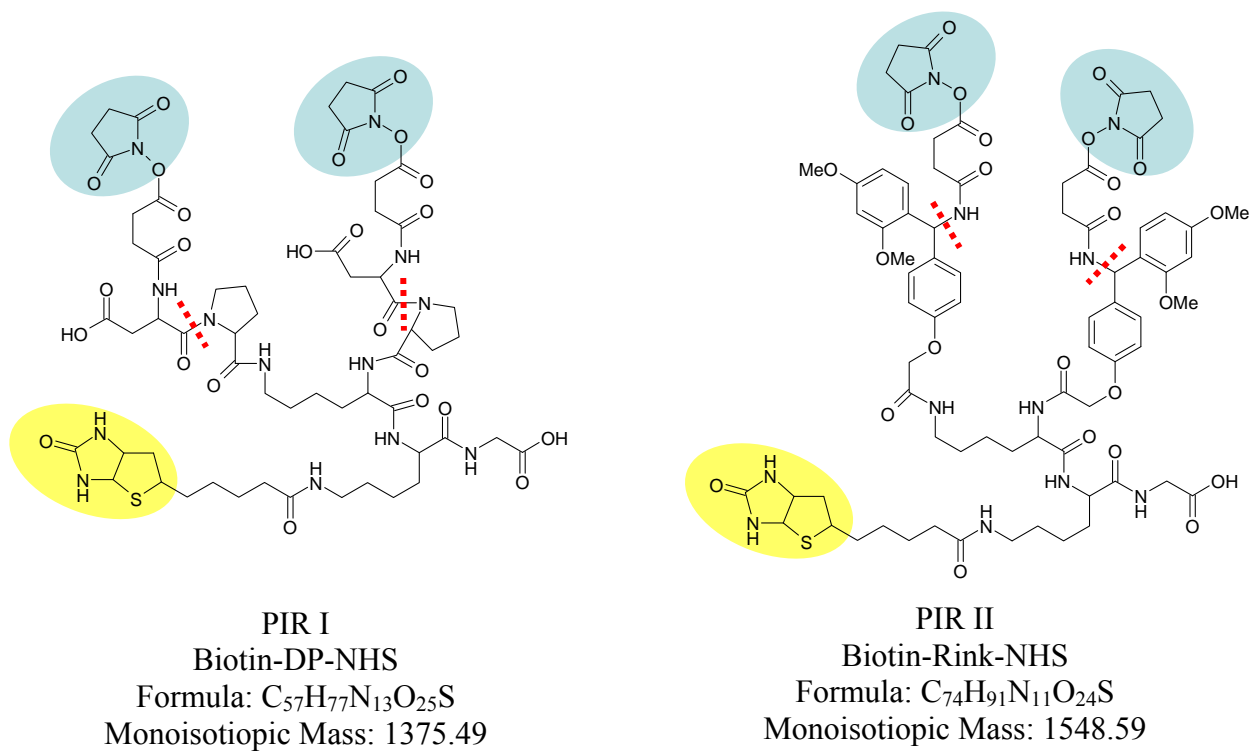
Figure 3-2. Molecular structures of PIR I and II The blue shadows represent NHS ester.

The yellow shadows indicate biotin. The red dotted lines represent MS cleavable bonds.

(a) PIR I

modified tag $C_8H_7N_1O_5$
monoisotopic mass : 197.03

reporter ion $C_{34}H_{58}N_9O_8S_1$
monoisotopic mass : 752.41

(b) PIR II

modified tag $C_4H_5N_1O_2$
monoisotopic mass : 99.03

reporter ion $C_{58}H_{71}N_7O_{14}S_1$
monoisotopic mass : 1122.48
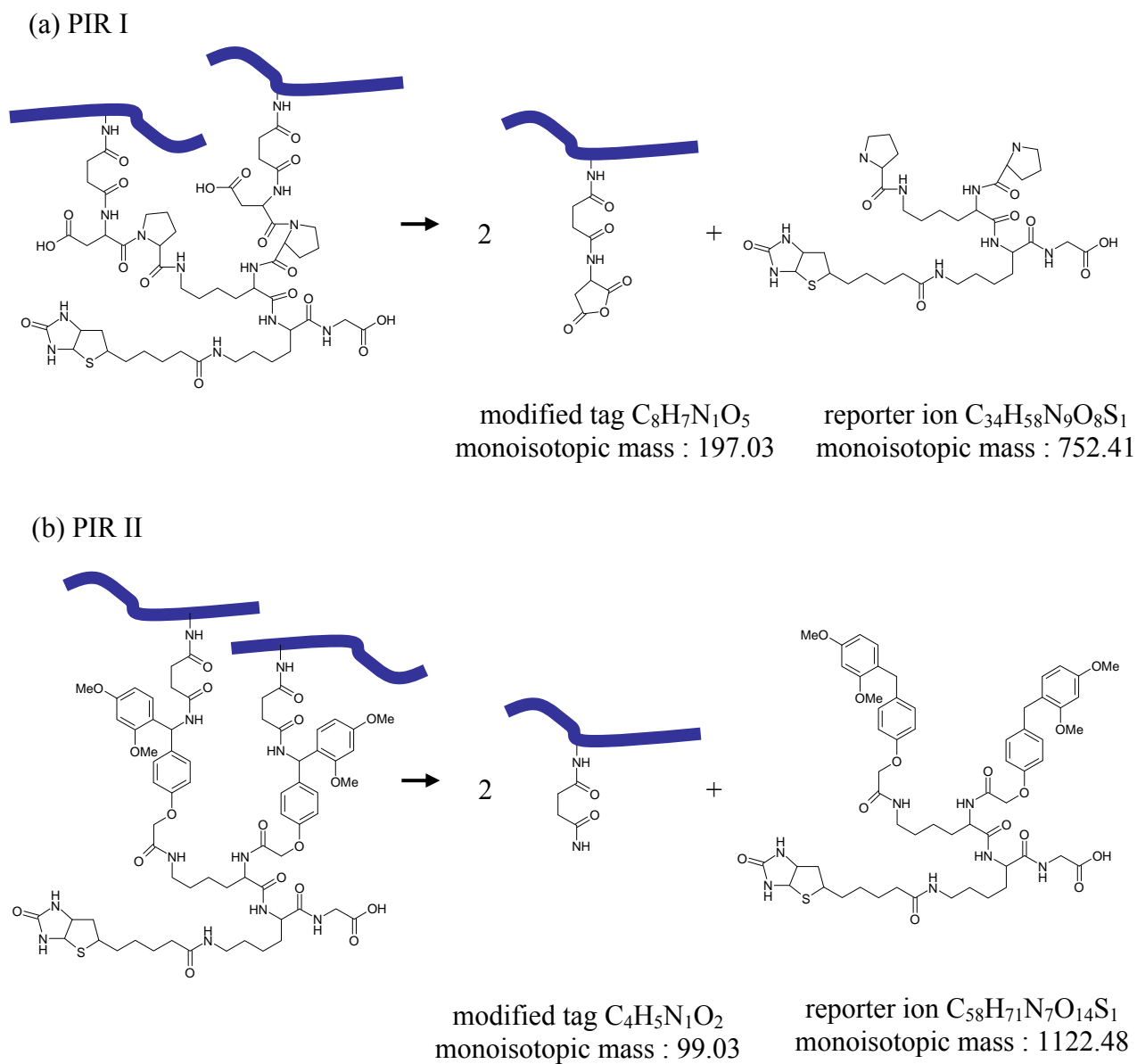
Figure 3-3. Fragmentation mechanisms at MS cleavable bonds of (a) PIR I, and (b) PIR II.
Two labile bonds preferably cleaved by ISCID. Cleavage of both labile bonds gives rise to
reporter ions at m/z (a) 752.41, and (b) 1122.48.

3-1. Protein Interaction Reporter Technology Application to Human Cells

PIR coupled with mass spectrometry was demonstrated to study protein-protein interactions of prokaryotic cells, such as bacteria. But the feasibility of this method applied to eukaryotic cells was not verified. Two human cells, MCF7 and HeLa were used to demonstrate the feasibility of PIR Strategy with eukaryotic cells. MCF7 is a breast cancer cell, and HeLa is a cervical cancer cell. These cells were cultured and labeled with PIR II to profile the proteomic data of MCF7 and HeLa cells using protein interaction reporter Strategy.

Human Cell Labeling with PIR

Confocal microscope was used to verify the feasibility of the human cell labeling with PIRs. PIR labeled cells were reacted with anti-biotin antibodies. Then, anti-biotin antibodies were sequentially interacted with Alexa Fluor 488 rabbit anti-mouse IgG and Alexa Fluor 488 goat anti-rabbit IgG. There antigen-antibody reactions produce and amplify the green fluorescence, which is biotin-specific signal. The nuclei of human cells are dyed by propidium iodide and produce the red fluorescence. The green and red fluorescence was used to visualize confocal microscopic images. Schematic mechanisms of the confocal microscopy are shown in Figure 3-4 [43].

In this report, PIR II labeled MCF7 cells were used to take microscopic images. The green and red fluorescence was separately scanned to take microscopic images. Then, two fluorescence images overlapped to obtain whole cell images. The fluorescence images of cells with and without PIR labeling are shown in Figure 3-5. Unlabeled cell produced only the red fluorescence, thus this sample does not contained biotin related species inside and outside cells. But, labeled sample produced both red and green fluorescence. The biotin-specific signal is

produced by PIR used for cell labeling. Therefore, PIRs can label human cells without lyzing

cells. Also, the green fluorescence was expressed outside of cell membranes. This result can

indicate that anti-biotin antibodies used to produce the fluorescence can not penetrate the cell

membrane.

: Anti-biotin antibody

: Alexa Fluor 488 rabbit anti-mouse IgG and Alexa Fluor 488 goat anti-
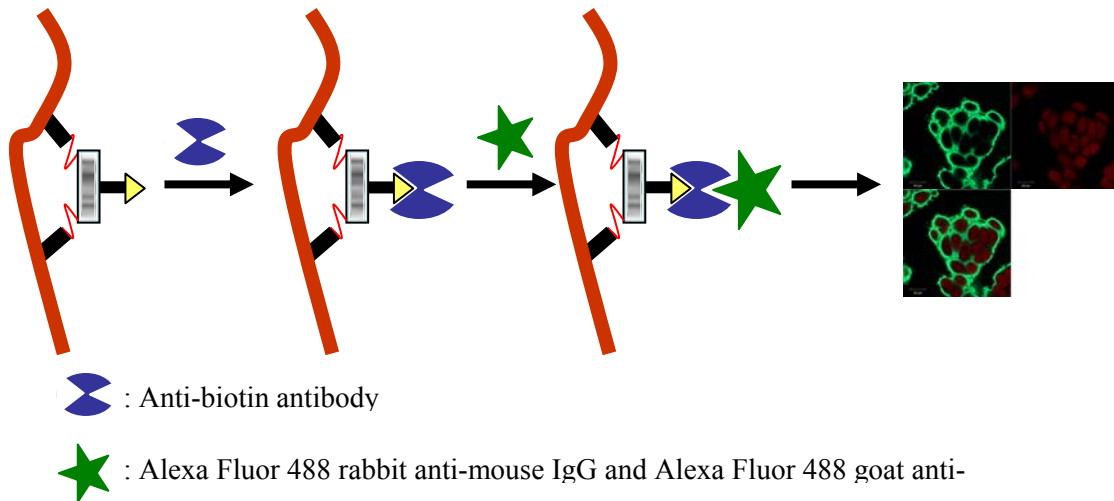
Figure 3-4. Schematic mechanisms of confocal microscopy [43].

(a)   (b)
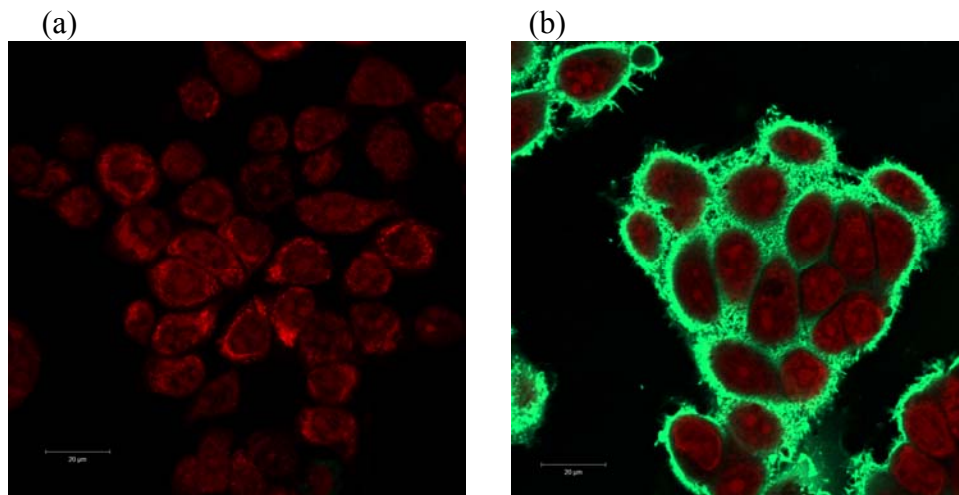
Figure 3-5. Confocal fluorescence images of MCF7 cells (a) without PIR II labeling, and (b)

with PIR II labeling. Both images are the superposition of fluorescent signals from Alexa Fluor

488 conjugated IgG (green; indicates biotin) and propidium iodide (red; indicates nuclei).

SDS-PAGE and Anti-Biotin Western Blot Analysis

The feasibility of human cell labeling with PIRs was verified using confocal microscopy, but these images are insufficient to determine PIR labeling efficiency. SPS-PAGE and anti-biotin western blot analysis were performed. MCF7.

The human cells were cultured and divided into two portions. The one portion of cells was lysed without PIR labeling. The unlabeled sample is called "Control" sample. The other was labeled with PIR II and lysed. This sample is called "Labeled" sample. 8% SDS polyacrylamide gel was run to separate proteins in these samples. Then, the gels were visualized by Coomassie Blue stain. The gel images are shown in Figure 3-6. The cell lysates of control and labeled samples produced identical gel images, and numerous protein bands were observed on the gels. Therefore, both cell lysates contain a large variety of proteins, but SDS-PAGE is not adequate to distinguish two different samples.

To determine the labeling efficiency, anti-biotin western blot analysis was carried out using both control and labeled samples. The proteins in the cell lystaes were separated by SDS-PAGE, and then transferred on the nitrocellulose membrane. This membrane was treated with anti-biotin and commercial chemoluminescent substrates. The membrane was visualized by exposing the film to chemoluminescence. Figure 3-7 shows the obtained western blot images. The images of the labeled samples contain dark spots. The dark spot is biotin-specific signal generated by chemoluminescence. The control samples did not produce biotin-specific signals. Thus, PIRs are efficiently reacted with human cells, and these results are in good agreement with the confocal microscopic images.
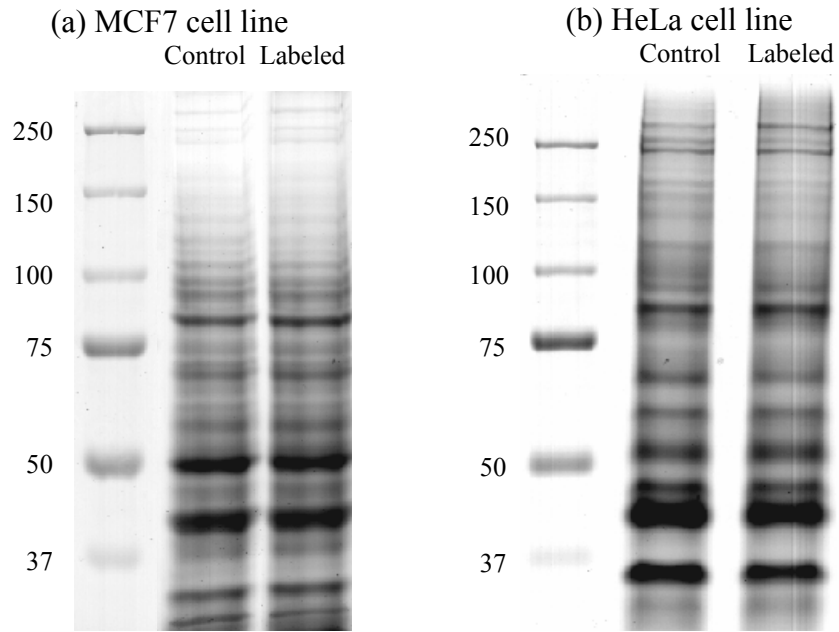
Figure 3-6. SDS-PAGE analysis of (a) MCF7 and (b) HeLa cell lysates.
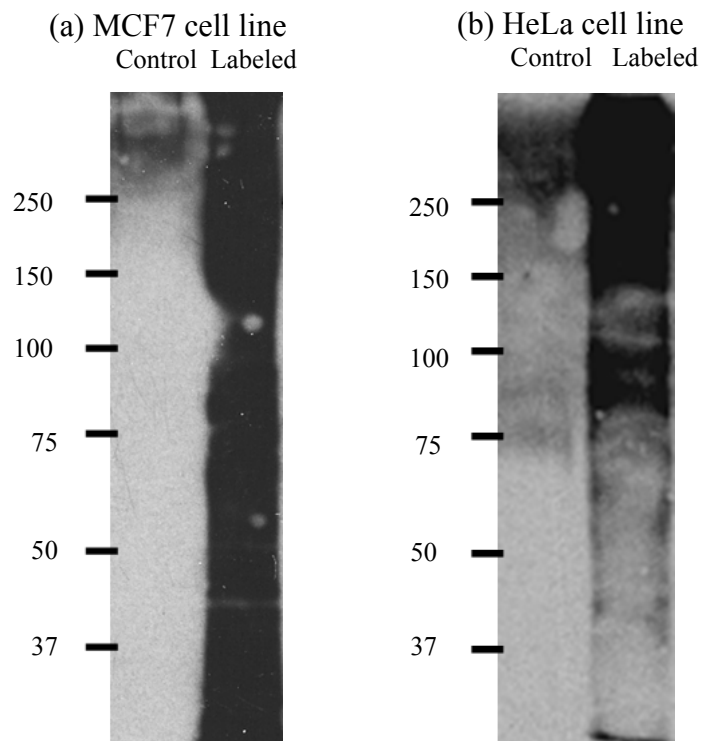
Figure 3-7. Anti-biotin weastern blot analysis of (a) MCF7 and (b) HeLa cell lysates.

Identification of PIR II Labeled Proteins

PIR labeled proteins were enriched from cell lystea using monomertic avidin beads and eluted in Laemmli sample buffer by boiling to profile the proteomic data of MCF7 and HeLa. The proteins eluted from avidin beads were separated by SDS-PAGE. The gel images are shown in Figure 3-8. The control samples produce few protein bands in the gel, but labeled samples contain more than 25 protein bands. As the result of the gel images, PIR labeled proteins were efficiently captured by avidin beads. The proteins bands present in the gels were excised and destained under acidic condition. The proteins in the gel pieces were reduced and alkylated by DTT and IAA. Then, these proteins were digested by trypsin. In-gel digests were subjected to nano LC/MS/MS, and the proteins identified by Mascot search of LC/MS/MS data. In labeled MCF7 cells, 109 proteins were identified. 127 proteins are identified using labeled HeLa cells as a sample. The identified proteins were listed in Table 4-1 and 4-2 in Chapter 4.

The labeled proteins were compared with proteome data obtained from literatures [44-46]. According to the literatures, the proteins present in MCF7 and HeLa were analyzed by two-dimensional (2D) gel electrophoresis and identified by MALFI-TOF MS on the basis of PMF, following in-gel digestion. The protein profiles contain 1859 proteins for MCF7 and approximately 1200 proteins for HeLa. As a result of comparison, PIR labeled proteins identified from MCF7 and HeLa cell lines are subsets of reported proteome data. 5.86 % of the detectable proteins present in MCF7 and 10.58 % of the detectable proteins in HeLa are labeled by PIR II. Subcellular locations of PIR labeled proteins are obtained from DAVID Bioinformatics Resources and summarized in Figure 3-9. The subcelluar locations of 50% of PIR labeled proteins are unknown. The other proteins are located in the variety of organelles and cytoplasm. Thus, PIR II can penetrate the cell membrane. But a small number of proteins in human cells can

be labeled with PIR II.
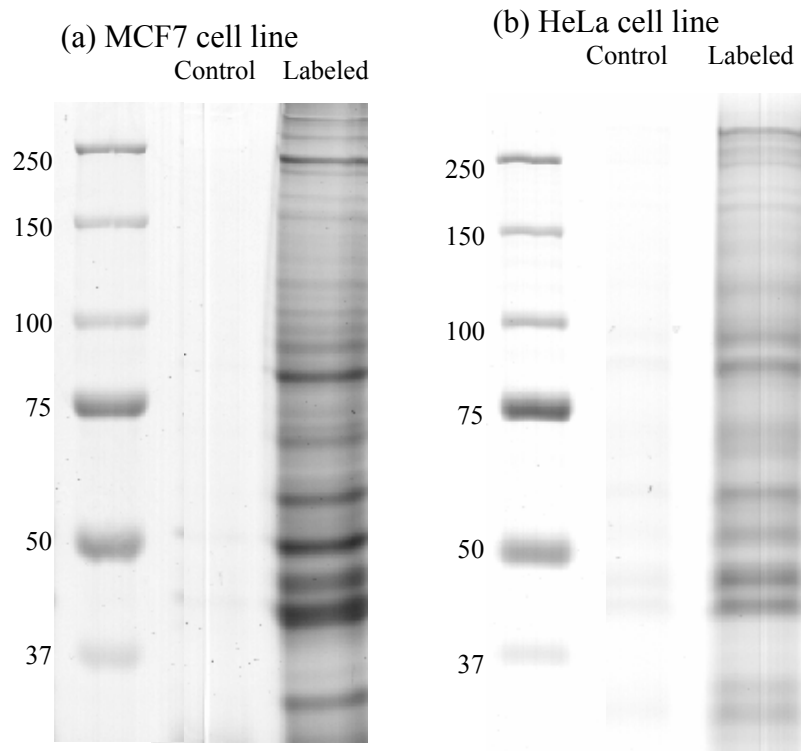
Figure 3-8. SDS-PAGE analysis of (a) MCF7 and (b) HeLa cell lysates after enrichment process.

(a)

Unknown
51.38% (56)

Nucleus
6.42% (7)

Mitochondrion
6.42% (7)

Membrane
8.26% (9)

Cytoplasm
15.60% (17)

Golgi Apparatus
1.83% (2)

Extracellular
0.92% (1)

Cytoskeleton
3.67% (4)

Endoplasmic Reticulum
5.50% (6)

(b)

Unknown
48.82% (62)

Nucleus
4.72% (6)

Mitochondrion
4.72% (6)

Membrane
14.96% (19)

Secreted
7.09% (9)

Cytoplasm
15.75% (20)

Melanosome
0.97% (1)

Cell Junction
0.97% (1)

Cytoskeleton
1.57% (2)

Endoplasmic Reticulum
0.97% (1)
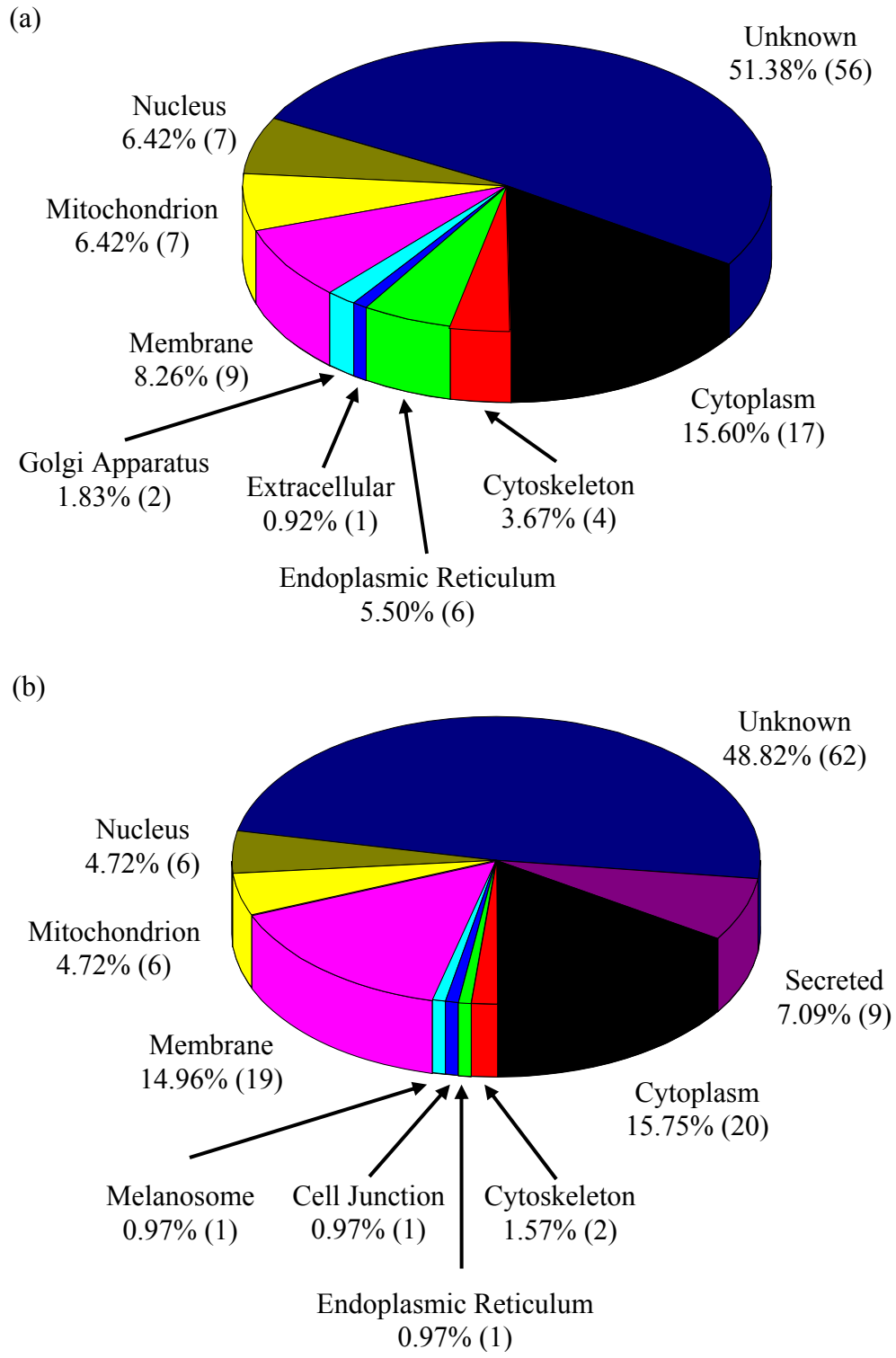
Figure 3-9. Protein categorization of the identified protein by subcellular locations, (a) MCF7 and (b) HeLa cell line

3-2. Multiplexed Fragmentation

Sample Preparations using PIRs and Model Peptides

Multiplexed fragmentation method with PIR Strategy was developed to study protein-protein interaction. Cross-linked tryptic peptides is required to demonstrate PIR coupled with 3-stage multiplexed fragments experiments. Commercial standard peptides, such as angiotensin I and bradykinin, can not be used for this experiment, because the standard peptides are not tryptic peptides. Peptide I and II were synthesized to perform this experiment.

As shown in Figure 2-3, the model peptides contain one internal Arg and Lys. C-terminus of the peptides is Arg and N-terminus of the peptide is blocked by acetyl group. PIRs react with internal Lys. After cross-linking, the reaction mixture is digested by trypsin. Thus cross-linked peptides have same structure of tryptic peptides. The proposed reaction mechanisms of inter-cross-linking are shown in Figure 3-10.

Model peptide I was reacted with commercial cross-linker, disuccinimidyl suberate (DSS), to prove efficiency of sample preparation process. MALDI-TOF MS spectrum of the reaction mixture was obtained during each process. The results show in figure 3-11. The MALDI-TOF MS spectrum shows Peptide I at *m/z*: 1174.706 before the sample preparation. The cross-linking mixture contains inter-cross-linked products at *m/z* 2486.501, and dead ends at *m/z* 1330.709. Inter-cross-lined tryptic peptides at *m/z* 1718.051, dead-end at *m/z* 945.470, and unreacted peptide at *m/z* 1174.710 were produced by try trypsin digestion. These MS spectra indicate that the sample preparation is efficiently carried out using model peptides.

1st step: Cross-Linking Reaction

Inter-cross-linked
product

Dead-end

Unreacted
Peptide
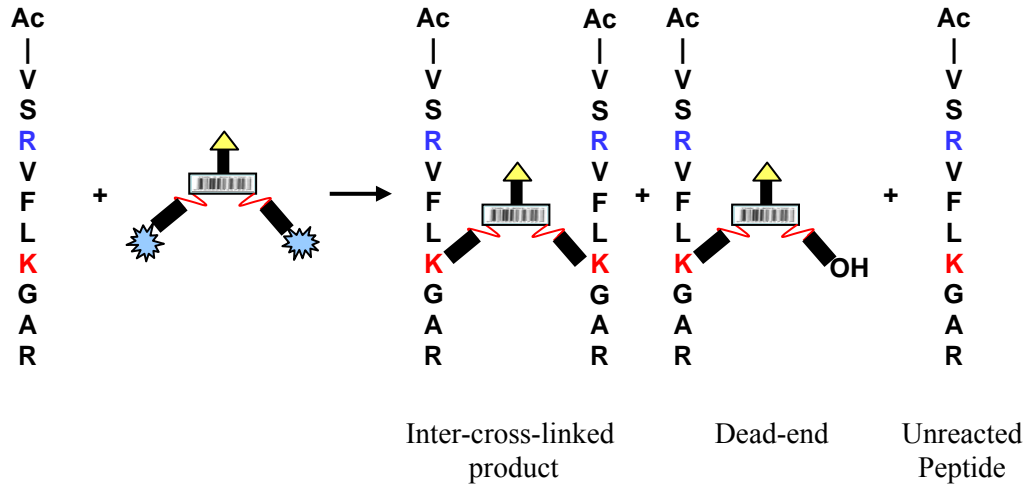
2nd step: Trypsin Digestion
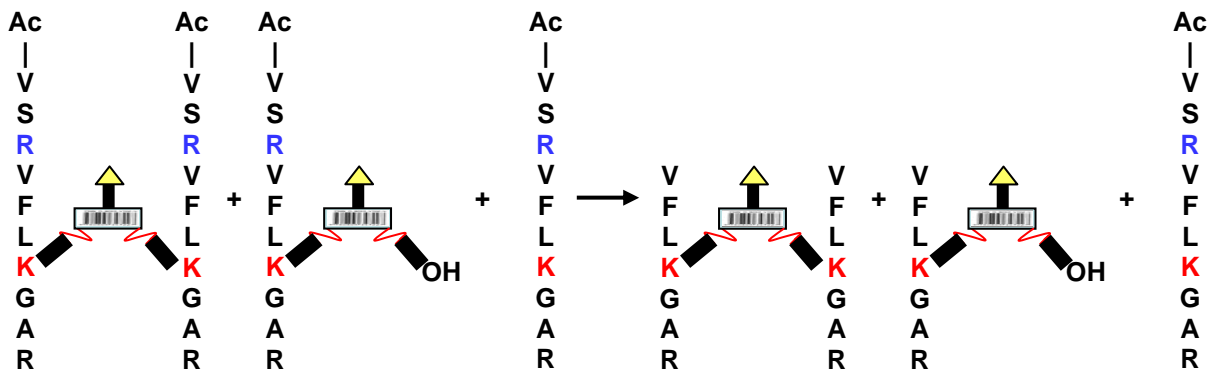
Figure 3-10. Schematic mechanisms of cross-linking reaction between PIR and Peptide I and trypsin digestion.
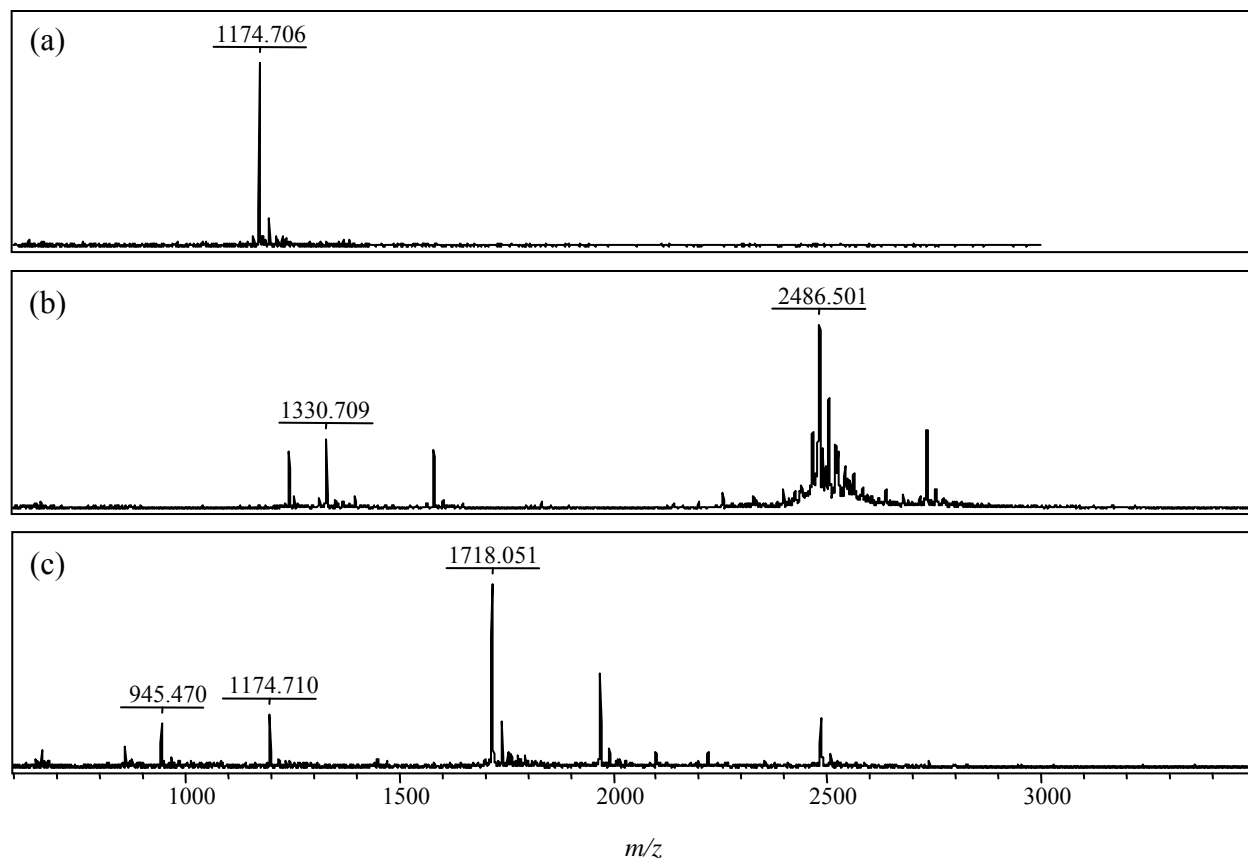
Figure 3-11. MALDI-TOF spectra of (a) Peptide I, (b) the cross-linking mixture of Peptide I and DSS, and (c) trypsin digest of the reaction mixture.

In-Source Collision-Induced Dissociation (ISCID) and Double Activation Analysis.

In the previous reports [16-18], 2-stage multiplexed fragmentation with PIR strategy was performed. This method was achieved by the parallel alternating scans which are acquired at either low or high collision energy in Hexapole #2. The information about PIR labeled peptides was obtained at low collision energy. The report ions and modified peptides were generated by high collision energy in Hexapole #2. Thus, fragmentation of modified peptides can not be performed. To perform 3-stage multiplexed fragmentation, the modified peptides and report ions need to be generated by ISCID [41]. 3-Stage multiplexed fragmentation approach consists of MS, ISCID and double activation scans. In the ISCID stage, the potential difference between capillary exit and skimmer activates MS cleavable bonds to release reporter ions and modified peptides. The double activation is achieved using ISCID and high collision energy in Hexapole #2 to fragment the modified peptides from cross-linked products.

To characterize the dissociation of MS cleavable bonds by ISCID, two PIR labeled model peptides were used. The first sample is PIR I labeled Peptide I, and the other is PIR II labeled Peptide II. Then, MS spectra of these samples were obtained under various potential differences between capillary exit and skimmer. The representative spectra are shown in Figure 3-12. The optimal conditions of MS and ISICD scans were determined using the peak intensities of inter-cross-linked products, and modified peptides. Normalized peak intensities of the species were shown in Figure 3-13. The intensities of inter-cross-linked products (charge state, 4+) present in two different samples are maximal when the potential difference between capillary exit and skimmer is 100V. The maximum intensities of modified peptides (charge state, 1+) are obtained at the potential difference, 240V. These potential differences are used to obtain the optimal spectra of MS and ISCID scans.
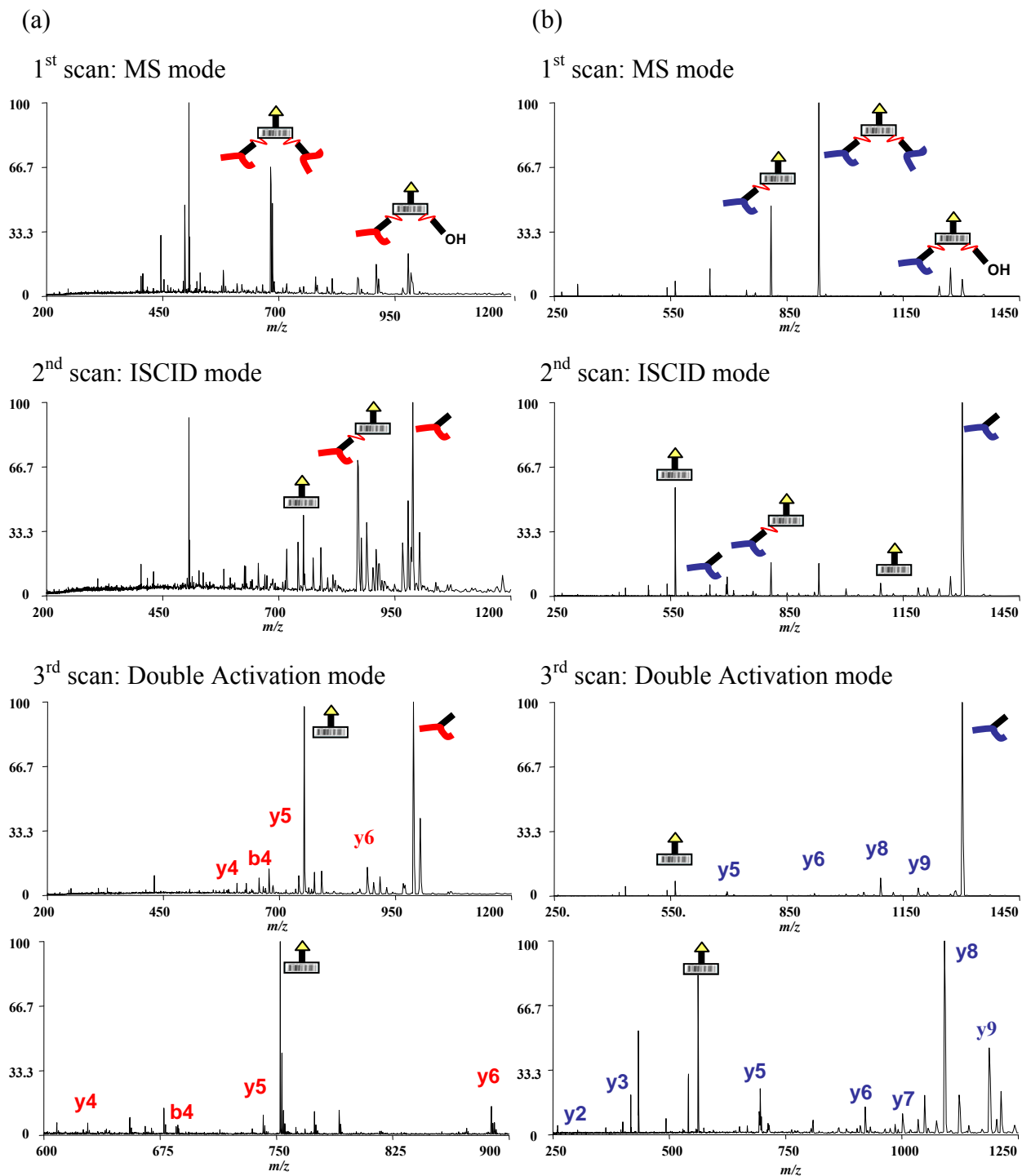
Figure 3-12. 3-Stage multiplexed fragmentation spectra of (a) PIR I labeled Peptide I, (b) PIR II labeled Peptide II. The bottom spectra show the fragment of modified peptides using changing *m/z* range.

37
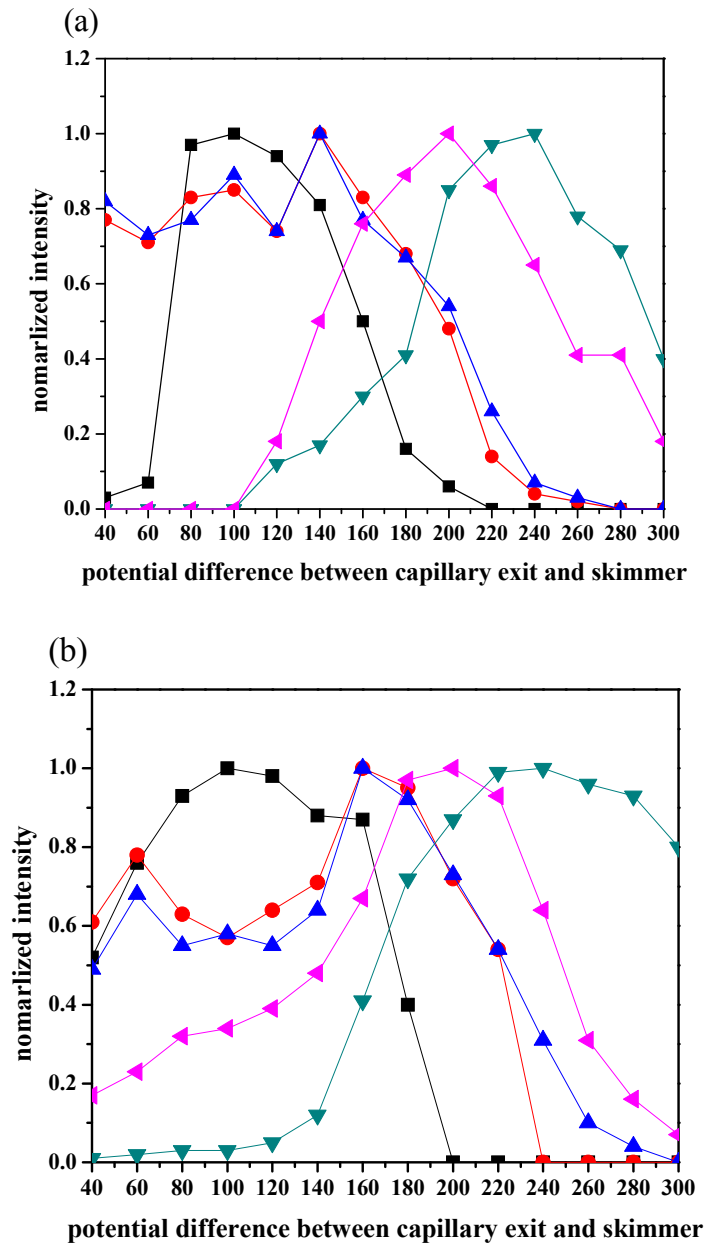
Figure 3-13. Potential difference between capillary exit and skimmer vs. normalized peak intensities present in (a) PIR I labeled peptide I, (b) PIR II labeled peptide II. (■) inter-cross-linked product (charge state, 4+), (●) inter-cross-linked product (charge state, 3+), (▲) dead-end (charge state, 2+) , (▼) modified peptide (charge state, 1+), (►) reporter ion (charge state, 1+ for PIR I, 2+ for PIR II).

The optimal condition for double activation scan mode was determined by taking MS spectra using PIR I labeled peptide I. Figure 3-14 shows the normalized intensities of modified peptide, and y ions. The intensities of these species are maximal when the applied voltage to Hexapole #2 is 30V. Thus, this voltage is used for double activation experiments.
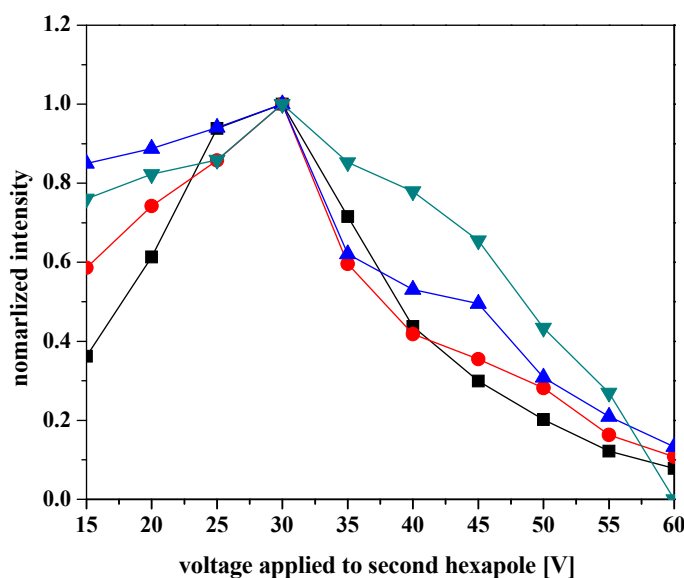
Figure 3-14. Double activation scan experiments using PIR I labeled model peptide I. The applied voltage to the second hexapole for double activation is increasing when the potential difference between capillary exit and Skimmer is 240V. (■) modified peptide, (●) y6 ion, (▲) y5 ion, (▼) y4 ion or modified peptides.

LC/3-Stage Multiplexed Fragmentation Experiments

The voltages applied to each segments of instrument need to be changed automatically to perform LC/3-stage multiplexed fragmentation experiments. Alternated scan experiment was designed to demonstrate the feasibility of LC/3-stage multiplexed fragmentation experiments. The alternative scan experiments consist of direct infusion as a sample introduced method and automatic operation of instrument using Bruker XMASS software to change the applied voltages. Total number of scans for this experiment is 16. In other words, 16 spectra were obtained during this experiment using PIR I labeled Peptide I. Figure 3-15 shows the representative spectra obtained using the alternated scan experiment. The scan number of this experiment determines the scan mode using following relationships.

Scan mode	Scan number	for alternated experiments
MS	3n+1,	1, 4, 7, 10, 13, 16
ISCID	3n+2	2, 5, 8, 11, 14
Double activation	3n+3 [3(n+1), or 3m]	3, 6, 9, 12, 15

Extracted ion chromatograms (EICs) of inter-cross-linked product (charge state, 4+), reporter ion (charge state, 1+), modified peptide (charge state, 1+), and y6 ion (charge state, 1+) of modified peptide are shown in Figure 3-16. EICs of these species were compared to verify the feasibility and the efficiency of alternated scan experiments. As shown in Figure 3-16, EICs of the species indicate that peak intensities are changed by scan numbers. The intensity of inter-cross-linked product is maximal at scan number, 3n+1 and sequentially decreased at the scan number 3n+2, 3n+3. When the scan number is 3n+2, the peaks of reporter ion and modified peptide appear. The intensities of these peaks are maximal at the scan number 3n+3. This EIC indicate that the mass cleavable bonds of cross-linked products are not fully fragmented. But the

intensities of reporter ion and modified peptide are sufficient to identify the species. EIC of y6 ion of modified peptide shows the opposite result to EIC of inter-cross-linked product.

As a result of this experiment, inter-cross-linked products are detected at MS scan and release reporter ions and modified peptides during ISCID scan mode. Then y ions of modified peptides are generated by double activation. Thus, 3-stage multiplexed fragmentation experiments can be carried out using instrument operation software, Bruker XMASS software.
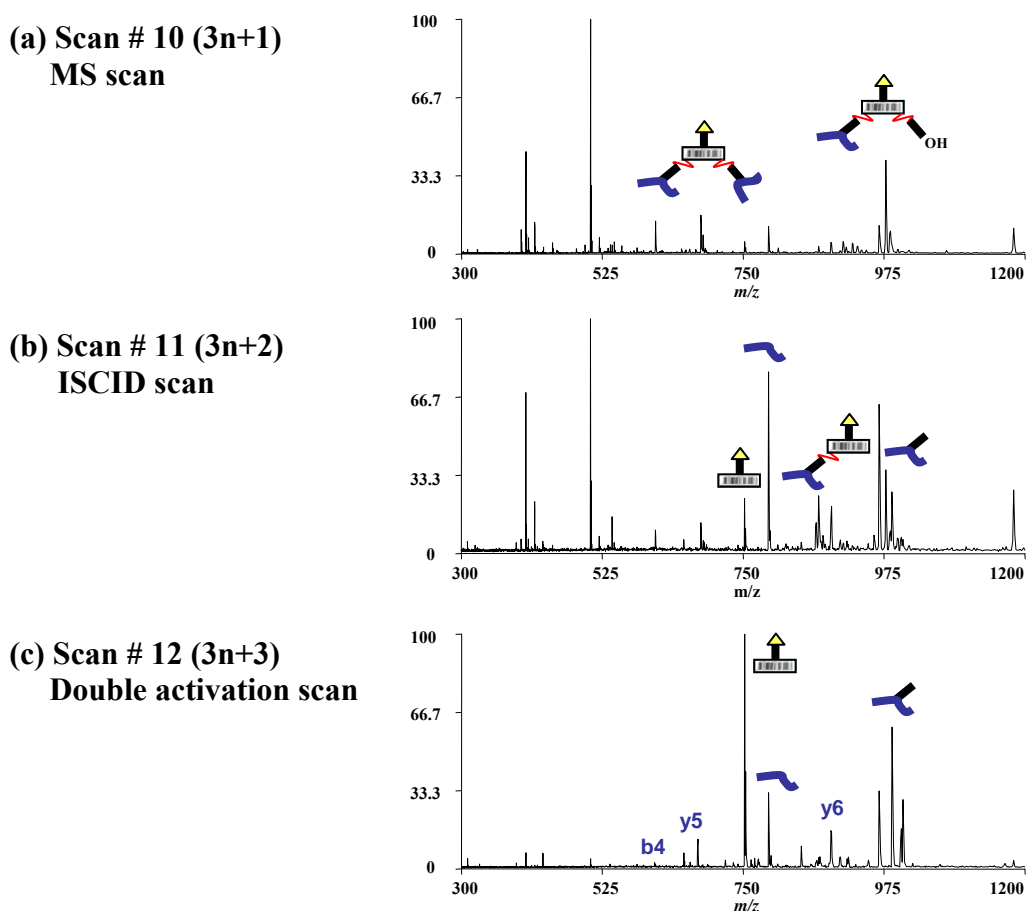
Figure 3-15. MS, ISCID and double activation spectra of PIR I labeled Peptide I using the alternated scan experiment.
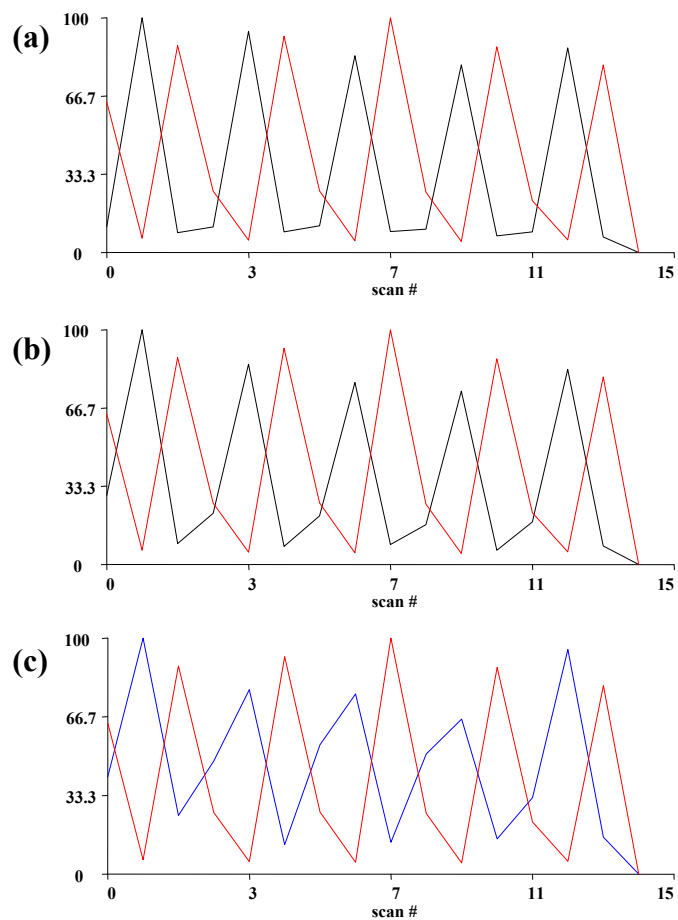
Figure 3-16. Normalized extracted ion chromatograms (EICs) of inter-cross-linked product (charge state, 4+), reporter ion (charge state, 1+), modified peptide (charge state, 1+), and y6 ion (charge state, 1+) obtained using alternative scan experiments with PIR I labeled Peptide I. The red line represents EIC of inter-cross-linked. The black lines represent EIC of (a) reporter ion, and (b) modified peptide. The blue line shows EIC of (c) y6 ion.

LC/3-stage multiplexed fragmentation experiment was carried out using PIR II labeled Peptide II. Total ion chromatogram (TIC) and EICs of inter-cross-linked product, reporter ion, modified peptide, and y ions are shown in Figure 3-17. TIC shows 3 peaks, but EICs of inter-cross-linked product have one peak at the scan number 1200~1250. Also, EICs of other species show the biggest peak at the scan number 1200~1350. To verify the relationship between inter-cross-linked product and other species, normalized EICs of these species were compared and the results are shown in Figure 3-18. The intensity of inter-cross-linked product is maximal at MS scan and sequentially decreased at ISCID and double activation scans. The peaks of reporter ion and modified peptide appear during ISCID scan. The intensities of these peaks are maximal at double activation scan. This EIC indicate that the mass cleavable bonds of cross-linked products are not fully fragmented. The intensities of y ions are increased when that of inter-cross-linked products is decreased. These results are good agreement with alternated scan experiments. Thus, LC/3-stage multiplexed fragmentation experiment with PIR strategy is feasible.
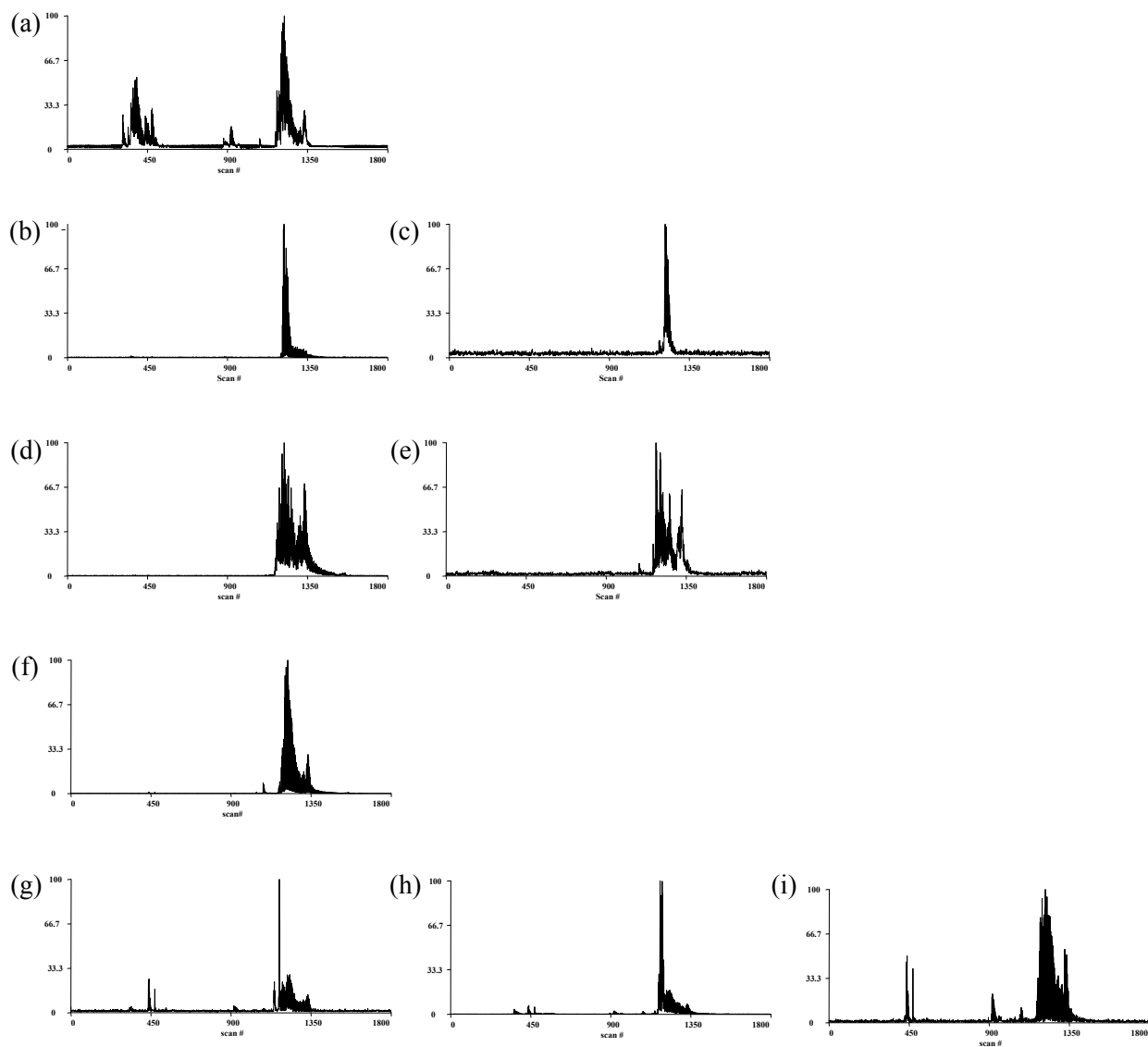
Figure 3-17. TIC and EICs obtained using PIR II labeled Peptide II from LC/3-stage multiplexed fragmentation experiment. (a) TIC, m/z range from 200 to 1800, EIC of (b) inter-cross-linked product (charge state, 4+), (c) inter-cross-linked product (charge state, 3+), (d) reporter ion (charge state, 2+), (e) reporter ion (charge state, 1+), (f) modified peptide (charge state, 1+), (g) y6 ion (charge state, 1+), (h) y8 ion(charge state, 1+) and (i) y9 ion (charge state, 1+).

44
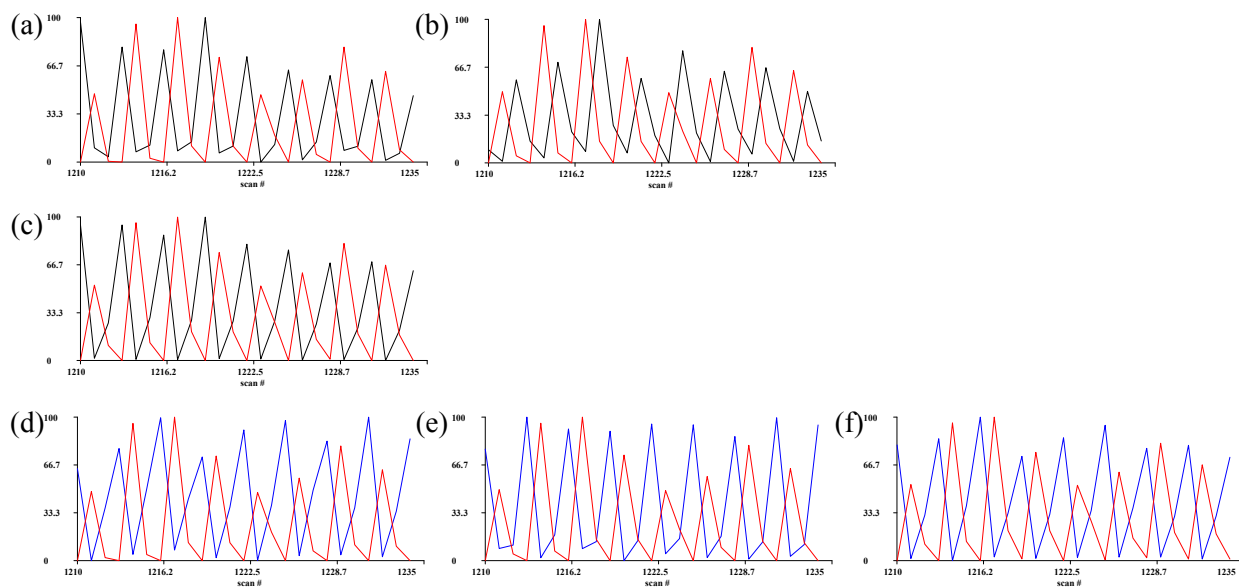
Figure 3-18. Normalized extracted ion chromatograms (EICs) obtained using PIR II labeled Peptide II at scan number 1200~1250. The red line represents EIC of inter-cross-linked product (charge state, 4+), the black lines represent EIC (a) reporter ion (charge state, 1+), (b) reporter ion (charge state, 2+), (c) modified peptide (charge state, 1+). The blue line shows (d) y6 ion (charge state, 1+), (e) y8 ion (charge state, 1+), and (f) y9 ion (charge state, 1+).

False Discovery Analysis of Multiplexed Fragmentation

The major challenge of multiplexed fragmentation method is the unambiguous identification of modified peptides. The identification of modified peptides is accomplished using the mass of this species and its fragment. If the sample contains several peptides having a same monoisotopic mass within the certain mass measurement tolerance, the identification of a certain peptide is ambiguous. In this report, the false discovery analysis was performed using the genome data of *Escherichia coli K12* used for multiplexed fragmentation experiments. The genome data of *Escherichia coli K12* contains 4,126 proteins, obtained using ORF. These protein sequences were loaded to X-links software and theoretical tryptic peptides in the mass rage from 500 to 5000 Da were calculated and listed. All possible tryptic peptides are 285, 227 without any restriction. The tryptic peptides having at least one Lys residue in sequences are 118,340. The mass accuracy of FT-ICR MS is generally known less than 5ppm. For this reason, the mass measurement tolerances used for the false discovery analysis are ±1, ±2, ±5, ±10, ±15, and ±25 part per million (ppm). The result of the false discovery is shown in Figure 3-18. Generally, the number of candidate peptides is smaller with lower mass measurement tolerance.

LC/3-stage multiplexed fragmentation was developed to identify PIR labeled sample, thus the result of restricted database is practically important. The average of candidate peptides from restricted database is 2.744 at ± 5 ppm as a mass measurement tolerance. When FT-ICR MS is used for multiplexed fragmentation experiments, less than 3 candidate peptides can be identified using only mass of modified peptides. As shown in Figure 3-19, the unambiguous identification of modified peptides using their masses is required the high mass measurement accuracy.

46
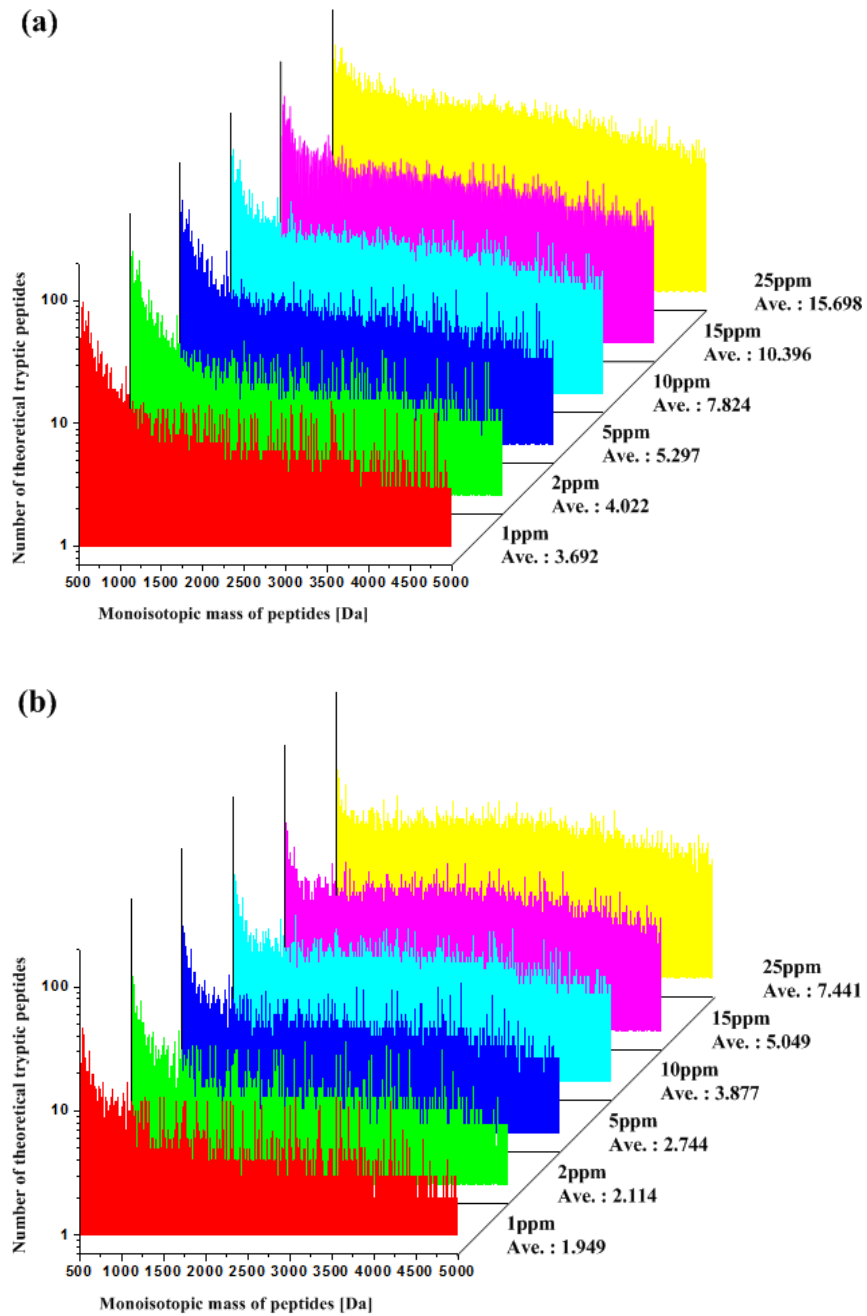
Figure 3-19. Number of theoretical tryptic peptides vs. monoisotopic mass of peptides as a function of mass measurement tolerances. (a) full database resulting in 285,227 peptides, (b) restricted database with 118,340 peptides The peptides in the restricted database contain an internal Lys.

The feasibility of identification of modified peptides using 3-stage multiplexed fragmentation data is verified by comparing the theoretical MS/MS pattern of candidate peptides with double activation spectra. First, the candidate peptides of each model peptide were search from the restricted database, and summarized in Table 3-1. Theoretical MS/MS patterns of candidate peptides were generated using GPMAW software and showed in Figure 3-20. Then the double activation spectra shown in Figure 3-12 were compared with theoretical MS/MS patterns. Asterisks (*) marked on theoretical MS/MS patterns represent the ions present in real double activation spectra. In the case of PIR I labeled Peptide I, only one theoretical pattern contains 4 fragments present in the double activation spectrum. Another pattern has 2 fragments. The other candidate peptides can not product any fragment present in the real spectrum. As a result of the comparison, the unambiguous identification of PIR labeled peptide is feasible using 3-stage multiplexed fragmentation data.

Table 3-1. Candidate peptides of Peptide I.

Protein	Candidate Peptide	Monoisotopic Mass
Hypothetical Protein ECDH10B_1591	[349]VFKAGLR[355]	789.4861
Predicted Transposase	[349]VFKAGLR[355]	789.4861
Conserved Inner Membrane Protein	[66]FGGKLLR[72]	789.4861
Hypothetical Protein ECDH10B_4318	[40]LFNLKR[46]	789.4861
Peptide I	VFLKGAR	789.4861
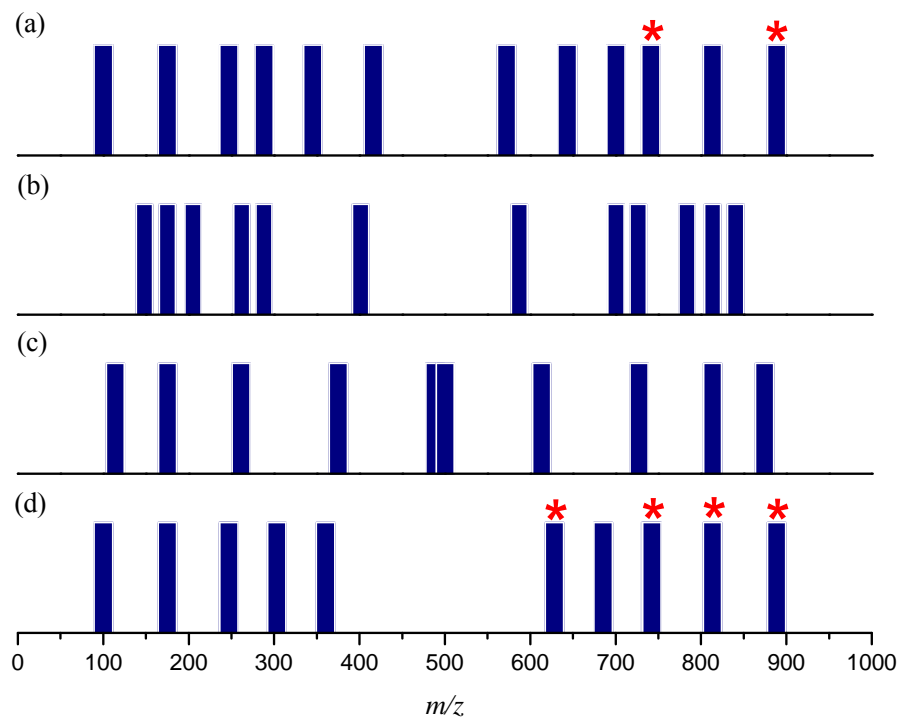
Figure 3-20. Theoretical MS/MS patterns of candidate peptides generated from (a) Hypothetical Protein ECDH10B_1591, and Predicted Transposase, (b) Conserved Inner Membrane Protein, (c) Hypothetical Protein ECDH10B_4318, and (d) Peptide I. Asterisks (*) represent fragments present in the double activation spectrum of PIR I labeled Peptide I.

Conclusion

in vivo identification of protein-protein interaction using PIR coupled with mass spectrometry has been reported [16, 17]. PIRs penetrate the bacteria cell membrane and react with the proteins without breaking cell membrane. In this report, PIR coupled with MS method was applied to human cell lines, such as MCF7 and HeLa. The feasibility of studying protein-protein interaction of human cell lines has been proven. The results showed that *in vivo* human cell line labeling using PIR is feasible for studying protein-protein interactions.

Chemical cross-linking combined with mass spectrometry is a common approach to study protein structure, and protein-protein interactions. This technique is primarily based on data-dependent LC/MS/MS method, however this method allows the identification of relatively abundant proteins. Consequently, this method is inadequate for identifying the protein-protein interactions of low abundant functional proteins. To overcome this problem, novel MS/MS approach, multiplexed fragmentation method has been designed. PIR coupled with multiplexed fragmentation was applied to tryptic peptides. The results established that the identification of PIR labeled peptides is feasible using the LC/3-stage multiplexed fragmentation method.

CHAPTER FOUR

SUPPLEMENTARY DATA

Table 4-1. Protein profile of PIR II labeled MCF7 cell line.

Gene (gi number)	Annotation	Mascot score	MW (KDa)	Subcellular location
gi|10803733	HBV pX associated protein-8; XAP-8	51	136	unknown
gi|1082356	epidermal autoantigen 450K (clone pE450-C/D)	82	73	unknown
gi|11024700	translocase of inner mitochondrial membrane 13	51	11	mitochondrion
gi|113417027	Interleukin-7 receptor alpha chain precursor	49	-	unknown
gi|113427611	KIAA1618	47	183	unknown
gi|114794262	Chain A, Crystal Structure Of The Bb' Fragment Of Erp57	117	28	unknown
gi|117949397	Uncharacterized protein C14orf179	52	24	unknown
gi|119567961	hCG1643231, isoform CRA_b	53	-	unknown
gi|119568113	spectrin repeat containing, nuclear envelope 1	49	-	unknown
gi|119569390	hCG2031476, isoform CRA_d	62	314	unknown
gi|119579660	hCG1748768, isoform CRA_a	73	-	unknown
gi|119580852	X-ray repair complementing defective repair in Chinese hamster cells 6	65	65	unknown
gi|119601287	spectrin, beta, erythrocytic (includes spherocytosis, clinical type I)	50	174	unknown
gi|119602582	epiplakin 1	83	324	unknown
gi|119614803	clathrin, heavy polypeptide (Hc), isoform CRA_c	50	194	unknown
gi|1220311	elongation factor-1 alpha	171	48	mitochondrion
gi|1346343	keratin, type II cytoskeletal 1	157	66	unknown
gi|157831962	Chain A, Human Protein Disulfide Isomerase, Nmr, 40 Structures	106	13	unknown
gi|157881958	Chain A, Solution Structure Of Dsrm Domain In Spermatid Perinuclear Rna-Bind Protein	49	102	unknown
gi|16041796	HNRPU protein	67	80	nucleus
gi|16924240	WAPAL protein	52	46	unknown
gi|177207	4F2 antigen heavy chain	133	119	membrane
gi|178027	alpha-actin	183	42	cytoskeleton

gi|178045	gamma-actin	219	26	cytoskeleton
gi|18031720	keratin protein K6irs	49	-	unknown
gi|18032008	scribble	52	-	membrane
gi|181400	cytokeratin 8	431	54	Cytoplasm
gi|181965	elongation factor 1 alpha	81	24	Cytoplasm
gi|182118	gamma enolase	64	45	membrane
gi|183932	receptor protein kinase	50	112	membrane
gi|186629	keratin 10	89	40	unknown
gi|188492	heat shock-induced protein	348	71	unknown
gi|189998	M2-type pyruvate kinase	365	58	unknown
gi|21753899	hCG2008140	51	20	unknown
gi|23194377	ALFY	49	399	unknown
gi|24485	ubiquitin-activating enzyme E1	133	91	unknown
gi|27368062	class IVb beta tubulin	196	50	unknown
gi|27754056	tubulin, beta 6	83	-	unknown
gi|28317	keratin 10	58	60	unknown
gi|28614	aldolase A	48	40	unknown
gi|292059	MTHSP75	87	74	unknown
gi|30311	cytokeratin 18 (424 AA)	76	47	cytoplasm
gi|30314935	ovarian epithelial carcinoma-related protein	54	30	nucleus
gi|30582781	tubulin, beta, 4	197	90	membrane
gi|306890	chaperonin (HSP60)	80	61	mitochondrion
gi|306891	90kDa heat shock protein	116	83	cytoplasm
gi|307086	keratin-10	66	46	unknown
gi|307129	hepatic lipase precursor	52	-	membrane
gi|307383	RNA helicase A	68	143	cytoplasm
gi|31092	eukaryotic translation elongation	213	50	cytoplasm
gi|31170	enolase 3	62	47	membrane
gi|31542947	chaperonin	210	61	mitochondrion
gi|31645	glyceraldehyde-3-phosphate dehydrogenase	242	36	cytoplasm
gi|32189394	ATP synthase, H+ transporting, mitochondrial F1 complex, beta subunit precursor	295	57	Mitochondrion
gi|3327212	KIAA0699 protein	57	96	golgi apparatus
gi|33286420	pyruvate kinase 3 isoform 2	58	58	unknown

gi|33350932	dynein, cytoplasmic, heavy polypeptide 1	55	535	cytoplasm
gi|338695	beta-tubulin	249	50	unknown
gi|33987931	HSP90AB1 protein	47	20	cytoplasm
gi|34039	keratin 19	107	44	unknown
gi|34098676	Protein bicaudal D homolog 1 (Bic-D 1)	77	111	unknown
gi|34365095	hypothetical protein	52	122	unknown
gi|34740335	tubulin, alpha 1B	230	50	unknown
gi|348239	[Human mRNA, complete cds.], gene product	56	54	nucleus
gi|35655	P4HB protein	136	57	endoplasmic reticulum
gi|35959	tubulin 5-beta	106	50	unknown
gi|37492	alpha-tubulin	136	51	unknown
gi|386758	GRP78 precursor	148	72	endoplasmic reticulum
gi|39930469	brix domain containing 1	51	36	nucleus
gi|40889577	Chain A, Human Topoisomerase I (Topo70) Double Mutant K532rY723F	48	70	unknown
gi|409875	microtubule-associated protein 2	48	-	unknown
gi|41584442	fatty acid synthase	192	276	cytoplasm
gi|42415492	chromosome 14 open reading frame 106	51	130	nucleus
gi|4501885	beta actin	259	42	cytoskeleton
gi|4502249	development- and differentiation-enhancing factor 2	54	113	golgi apparatus
gi|4503481	eukaryotic translation elongation factor 1 gamma	107	50	unknown
gi|4503483	eukaryotic translation elongation factor 2	209	96	cytoplasm
gi|4503571	enolase 1	246	47	membrane
gi|4506675	ribophorin I precursor	75	68	endoplasmic reticulum
gi|4507677	tumor rejection antigen (gp96) 1	162	93	endoplasmic reticulum
gi|4507729	tubulin, beta 2	197	50	unknown
gi|453155	keratin 9	54	62	unknown
gi|4757756	annexin A2 isoform 2	59	39	extracellular
gi|4758304	protein disulfide isomerase-associated 4	71	73	endoplasmic reticulum
gi|4758876	poly(A) binding protein, nuclear 1	52	33	cytoplasm
gi|48146635	PPP1R15A	50	74	unknown

gi|5123454	heat shock 70kDa protein 1A	319	70	unknown
gi|51571895	TPA: TPA_exp: centrosome protein Cep290	47	-	unknown
gi|5174735	tubulin, beta, 2	178	50	unknown
gi|53791219	filamin A	108	280	cytoplasm
gi|547754	keratin, type II cytoskeletal 2 epidermal	154	66	unknown
gi|55956937	meiosis defective 1	48	100	unknown
gi|5729877	heat shock 70kDa protein 8 isoform 1	374	71	cytoplasm
gi|58531795	G protein-coupled receptor 112	49	335	membrane
gi|62421162	actin-like protein	64	11	cytoskeleton
gi|62897441	keratin 8 variant	469	54	cytoplasm
gi|63055057	hypothetical protein LOC345651	130	42	unknown
gi|6648067	Malate dehydrogenase, mitochondrial precursor	72	36	mitochondrion
gi|68533057	APC variant protein	55	314	unknown
gi|70980549	programmed cell death 11	52		nucleus
gi|7141322	p37 TRAP/SMCC/PC2 subunit	50	32	unknown
gi|729422	Alpha-enolase, lung specific	102	50	cytoplasm
gi|799177	100 kDa coactivator	46	108	cytoplasm
gi|80475848	KRT73 protein	48		unknown
gi|860986	protein disulfide isomerase	129	57	endoplasmic reticulum
gi|862457	enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase alpha-subunit of trifunctional protein	80	84	mitochondrion
gi|89062428	U5 snRNP-specific protein, 200 kDa,	48	-	unknown
gi|8923894	thrombospondin type I domain-containing 1 isoform 1	48	96	unknown
gi|915392	fatty acid synthase	49	276	cytoplasm

Table 4-2. Protein profile of PIR II labeled HeLa cell line.

Gene (gi number)	Annotation	Mascot score	MW (KDa)	Subcellular location
gi|10716563	calnexin precursor	57	68	Endoplasmic reticulum
gi|109096516	PREDICTED: alpha tubulin isoform 2	180	38	Unknown
gi|110171869	decay-accelerating factor splicing variant 5	57	60	Unknown
gi|11493459	PRO2619	62	59	Secreted
gi|115206	RecName: Full=C-1-tetrahydrofolate synthase, cytoplasmic; Short=C1-THF synthase; Includes: RecName:	105	102	Cytoplasm
gi|119395750	keratin 1	74	66	Unknown
gi|119581085	keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris), isoform CRA_b	439	64	Unknown
gi|119585758	filamin B, beta (actin binding protein 278), isoform CRA_c	60	266	Unknown
gi|119592244	myosin, heavy polypeptide 14, isoform CRA_a	86	245	Unknown
gi|119592247	myosin, heavy polypeptide 14, isoform CRA_d	82	243	Unknown
gi|119597533	heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A), isoform CRA_b	47	63	Unknown
gi|119620391	chaperonin containing TCP1, subunit 4 (delta), isoform CRA_b	59	47	Unknown
gi|119626071	albumin, isoform CRA_h	79	71	Unknown
gi|119626083	albumin, isoform CRA_t	79	60	Unknown
gi|119628690	hCG31107, isoform CRA_b	53	20	Unknown
gi|1220311	elongation factor-1 alpha	158	48	Nucleus
gi|122920512	Chain A, Human Serum Albumin Complexed With Myristate And Aspirin	136	68	Unknown
gi|12653033	MYH10 protein	57	55	Unknown
gi|12667788	myosin, heavy polypeptide 9, non-muscle	516	228	Unknown
gi|1297274	beta-tubulin	234	51	Membrane
gi|1304128	polyubiquitin	51	68	Unknown
gi|134105063	Chain B, Crystal Structure And Solution Nmr Studies Of Lys48-Linked Tetraubiquitin At Neutral Ph	51	-	Unknown
gi|1346343	RecName: Full=Keratin, type II cytoskeletal 1; AltName: Full=Cytokeratin-1; Short=CK-1; AltName: Fu	373	66	cytoskeletal
gi|145309046	POTE-2 alpha-actin	458	123	Unknown

gi|155030190	mediator complex subunit 8 isoform 3	57	33	Unknown
gi|157830361	Chain A, Human Serum Albumin In A Complex With Myristic Acid And Tri- Iodobenzoic Acid	77	69	Unknown
gi|158258947	unnamed protein product	58	-	Unknown
gi|158260521	unnamed protein product	78	166	Unknown
gi|16041796	HNRPU protein	67	80	Nucleus
gi|177207	4F2 antigen heavy chain	117	58	Membrane
gi|178045	gamma-actin	508	26	Cytoplasm
gi|178345	alloalbumin Venezia	58	-	Secreted
gi|178685	lymphocyte activation antigen	75	58	Membrane
gi|178775	proapolipoprotein	60	29	Secreted
gi|181463	decay-accelerating factor precursor	50	42	Membrane
gi|182710	fibronectin receptor alpha-subunit precursor	54	114	Membrane
gi|189036	nonmuscle myosin heavy chain (NMHC)	52	146	Unknown
gi|19743813	integrin beta 1 isoform 1A precursor	69	92	Membrane
gi|20146101	EMMPRIN	56	29	Membrane
gi|21040386	HSPA9 protein	56	74	Mitochondrion
gi|2104553	Myosin heavy chain (MHY11) (5'partial)	115	215	Melanosome
gi|21961605	Keratin 10	56	59	Unknown
gi|220141	VLA-3 alpha subunit	47	115	Membrane
gi|223217	antigen H,HLA histocompatibility	59	32	Unknown
gi|23307793	serum albumin	79	71	Secreted
gi|2627129	polyubiquitin	51	68	Unknown
gi|27368062	class IVb beta tubulin	192	50	Unknown
gi|27692693	ALB protein	62	49	Secreted
gi|27754056	tubulin, beta 6	177	50	Unknown
gi|28243	unnamed protein product	47	283	Cytoplasm
gi|28317	unnamed protein product	200	60	Unknown
gi|28336	mutant beta-actin (beta-actin)	508	42	Cytoplasm
gi|28590	unnamed protein product	128	71	Secreted
gi|306890	chaperonin (HSP60)	114	61	Mitochondrion
gi|307086	keratin-10	79	46	Unknown
gi|31074631	keratin 1b	151	62	Unknown
gi|31438	unnamed protein product	141	116	Membrane

gi|31542947	chaperonin	438	61	Mitochondrion
gi|32015	alpha-tubulin	110	51	Unknown
gi|32189394	mitochondrial ATP synthase beta subunit precursor	60	57	Mitochondrion
gi|33286420	pyruvate kinase, muscle isoform M1	196	59	Unknown
gi|338827	cytosolic thyroid hormone-binding protein (EC 2.7.1.40)	56	58	Unknown
gi|33987931	HSP90AB1 protein	52	40	Cytoplasm
gi|37267	transketolase	51	58	Unknown
gi|37433	unnamed protein product	58	85	Membrane
gi|37492	alpha-tubulin	201	51	Unknown
gi|386854	type II keratin subunit protein	84	53	Unknown
gi|39645240	HNRPU protein	94	79	Nucleus
gi|40788275	KIAA0477 protein	51	131	Unknown
gi|40788339	KIAA0723 protein	60	96	Nucleus
gi|416178	desmoglein 2	120	123	Membrane
gi|42543212	Chain A, Decay Accelerating Factor (Cd55): The Structure Of An Intact Human Complement Regulator.	50	29	Unknown
gi|42543225	Chain A, Decay Accelerating Factor (Cd55): The Structure Of An Intact Human Complement Regulator.	50	29	Unknown
gi|4467833	cell adhesion molecule L1	61	92	Membrane
gi|4501881	actin, alpha 1, skeletal muscle	319	42	Cytoplasm
gi|4501887	actin, gamma 1 propeptide	553	42	Cytoplasm
gi|4503443	endothelin converting enzyme 1 isoform 1	84	87	Membrane
gi|4503483	eukaryotic translation elongation factor 2	70	96	Cytoplasm
gi|4505467	5' nucleotidase, ecto	50	64	Membrane
gi|4507729	tubulin, beta 2	250	50	Unknown
gi|4507943	exportin 1	102	124	Cytoplasm
gi|453155	keratin 9	111	62	Unknown
gi|4557761	mutS homolog 2	52	105	Nucleus
gi|4757810	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit precursor	134	60	Unknown
gi|48734966	Eukaryotic translation elongation factor 1 alpha 1	242	50	Cytoplasm
gi|5031753	heterogeneous nuclear ribonucleoprotein H1	56	49	Nucleus
gi|51476390	hypothetical protein	66	71	Secreted
gi|5174735	tubulin, beta, 2	196	50	Unknown

gi|521205	apolipoprotein C-III	68	10	Secreted
gi|535182	endothelin-converting-enzyme 1	63	86	Membrane
gi|53791219	filamin A	79	280	Cytoplasm
gi|547754	RecName: Full=Keratin, type II cytoskeletal 2 epidermal; AltName: Full=Cytokeratin-2e; Short=CK 2e;	160	66	cytoskeletal
gi|55665826	phosphodiesterase 4D interacting protein	51	255	Unknown
gi|55669910	Chain A, Crystal Structure Of The Ga Module Complexed With Human Serum Albumin	60	67	Unknown
gi|55956899	keratin 9	78	62	Unknown
gi|561722	monocarboxylate transporter 1	84	55	Membrane
gi|57014043	lamin A/C transcript variant 1	81	74	Unknown
gi|5729877	heat shock 70kDa protein 8 isoform 1	46	66	Cytoplasm
gi|61656603	Heat shock protein HSP 90-alpha 2	59	98	Cytoplasm
gi|62420916	actin-like protein	86	12	Cytoplasm
gi|62420995	actin-like protein	83	11	Cytoplasm
gi|62421162	actin-like protein	130	12	Cytoplasm
gi|62896589	eukaryotic translation elongation factor 1 alpha 1 variant	111	50	Cytoplasm
gi|62897413	pyruvate kinase 3 isoform 1 variant	289	58	Unknown
gi|62897525	eukaryotic translation elongation factor 1 alpha 1 variant	99	50	Cytoplasm
gi|62897639	tubulin, beta, 4 variant	234	51	Membrane
gi|62988624	unknown	76	97	Mitochondrion
gi|63055057	actin, beta-like 2	314	42	Unknown
gi|641958	non-muscle myosin B	83	230	Unknown
gi|641958	non-muscle myosin B	63	230	Unknown
gi|6424942	ALG-2 interacting protein 1	70	96	Cytoplasm
gi|6650826	PRO2044	138	30	Unknown
gi|67464392	Chain A, Structure Of Human Muscle Pyruvate Kinase (Pkm2)	58	60	Unknown
gi|7106439	tubulin, beta 5	259	50	Unknown
gi|71151982	RecName: Full=Myosin-14; AltName: Full=Myosin heavy chain 14; AltName: Full=Myosin heavy chain, non	82	229	Unknown
gi|7161776	cytokeratin	51	-	Unknown
gi|71979932	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5	97	56	Membrane
gi|7339522	CD1E	50	11	Membrane

gi|73535278	Chain A, Human Pyruvate Kinase M2	289	63	Unknown
gi|762885	Plakoglobin	62	82	Cell junction
gi|77702086	heat shock protein 60	62	61	Mitochondrion
gi|7770217	PRO2675	74	-	Unknown
gi|7959791	PRO1708	66	-	Secreted
gi|88942898	PREDICTED: similar to beta-actin	111	-	Unknown
gi|89036703	PREDICTED: similar to tubulin, beta 5	63	12	Unknown
gi|9082289	chaperone protein HSP90 beta	114	73	Cytoplasm
gi|951338	CAS	53	111	Unknown

Bibliography

1. PHIZICKY, E. M.; FIELDS, S., Protein-Protein Interactions: Methods for Detection and Analysis. *Microbiological Reviews* **1995**, 59(1) 01, 94-123.

2. Golemis, E.; Adams, P. A., *Protein-Protein Interactions: a molecular cloning manual.* 2nd Ed., CSHL Press, **2005**.

3. Arkin, M., Protein-protein interactions and cancer: small molecules going in for the kill. *Current Opinion in Chemical Biology*, **2005**, 9(3), 317-324.

4. Strittmattet, W.J.; Burke1, J.R.; DeSerrano, V.S.; Huang, D.Y.; Matthew, W.; Saunders1, A.M.; Scott1, B.L.; Vance1, J.M.; Weisgraber, K.H.; Roses, A.D., Protein:Protein Interactions in Alzheimer's Disease and the CAG Triplet Repeat Diseases. *Cold Spring Harbor Symposia on Quantitative Biology*, **1996** 61, 597-605.

5. Cisse, M.; Mucke,. L., Alzheimer's disease: A prion protein connection. *Nature*, **2009**, 457, 1090-1091.

6. Fields S,; Song O., A novel genetic system to detect protein-protein interactions. *Nature*, **1989**, 340, 245-246.

7. Johnsson, N.; Varshavsky, A., Split ubiquitin as a sensor of protein interactions in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, **1994**, 91, 10340-10344.

8. Fernandez, S.M.; Berlin, R.D., Cell surface distribution of lectin receptors determined by resonance energy transfer. *Nature*, **1976**, 264, 411-415.

9. Rigaut, G.; Shevchenko, a,; Rutz, B.; Wilm, M.; Mann, M,; Séraphin B., A generic protein purification method for protein complex characterization and proteome exploration, *Nature biotechnology*, **1999**, 17 1030-1032.

10. Zhu, H.; Klemic, J. F.; Chang, S.; Bertone, P.; Casamayor, A.; Klemic, K. G.; Smith, D.; Gerstein, M.; Reed M. A.; Snyder, M., Analysis of yeast protein kinases using protein chips, *Nature Genetics*, **200**0, 26, 283-289.

11. Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Tom Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, m.; Snyder, M., Global analysis of protein activities using proteome chips. *Science*, **2001**, 293, 2101–2105.

12. Kaelin W. G.; Pallas, D. C.; DeCaprio,J. A.; Kaye, F. J.; Livingston, D. M., Identification of cellular proteins that can interact specifically with the T/ElA-binding region of the retinoblastoma gene product, *Cell*, **1991**, 64, 521-532.

13. Sinz, A., Chemical cross-lining and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Reviews*, **200**6, 25, 663– 682.

14. Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E., Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Analytical Chemistry* **2005**, 77(1), 311-318.

15. Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R., Identification of cross-linked peptides from large sequence databases. *Nature Methods,* **2008**, 5(4), 315-318.

16. Tang, X.; Yi, W.; Munske, G. R.; Adhikari, D. P.; Zakharova, N. L.; Bruce, J. E., Profiling the membrane proteome of *Shewanella oneidensis MR-1* with new affinity labeling probes. *Journal of Proteome Research,* **2007**, 6(2), 724-734.

17. Zhang, H.; Tang, X.; Munske, G. R.; Zakharova, N.; Yang, L.; Zheng, C.; Wolff, M. A.; Tolic, N.; Anderson, G. A.; Shi, L.; Marshall, M. J.; Fredrickson, J. K.; Bruce, J. E., In vivo identification of the outer membrane protein OmcA-MtrC interaction network in *Shewanella oneidensis MR-1* cells using novel hydrophobic chemical cross-linkers. *Journal of Proteome Research* **2008**, 7(4), 1712-1720.

18. Anderson, G. A.; Tolic, N.; Tang, X.; Zheng, C.; Bruce, J. E., Informatics strategies for large-scale novel cross-linking analysis. *J Proteome Research* **2007,** 6(9), 3412-3421.

19. Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G., Chemical cross-linking and mass spectrometry for protein structural modeling. *Journal of Molecular Biology*, **2003**, 331(2), 303-313.

20. Borch, J.; Jorgensen, T. J.; Roepstorff, P., Mass spectrometric analysis of protein interactions. *Current Opinion of Chemical Biology*, **2005**, 9(5), 509-516.

21. Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O., Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Analytical Chemistry*, **2001**, 73(9), 1927-1934.

22. Taverner, T.; Hall, N. E.; O'Hair, R. A.; Simpson, R. J., Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *Journal of Biological Chemistry*, **2002**, 277(48), 46487-4692.

23. Chen, X.; Chen, Y. H.; Anderson, V. E., Protein cross-links: universal isolation and characterization by isotopic derivatization and electrospray ionization mass spectrometry. *Analytical Biochemistry*, **1999**, 273(2), 192-203.

24. Itoh, Y.; Cai, K.; Khorana, H. G., Mapping of contact sites in complex formation between light-activated rhodopsin and transducin by covalent crosslinking: use of a chemically preactivated reagent. *Proceedings of the National Academy of Sciences of the United States of America*, **2001**, 98(9), 4883-4887.

25. Alley, S. C.; Trakselis, M. A.; Mayer, M. U.; Ishmael, F. T.; Jones, A. D.; Benkovic, S. J., Building a replisome solution structure by elucidation of protein-protein interactions in the bacteriophage T4 DNA polymerase holoenzyme. *Journal of Biological Chemistry*, **2001**, 276(42), 39340-39349.

26. Bennett, K. L.; Kussmann, M.; Bjork, P.; Godzwon, M.; Mikkelsen, M.; Sorensen, P.; Roepstorff, P., Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular protein contacts. *Protein Science*, **2000**, 9(8), 1503-1518.

27. Back, J. W.; Sanz, M. A.; De Jong, L.; De Koning, L. J.; Nijtmans, L. G.; De Koster, C. G.; Grivell, L. A.; Van Der Spek, H.; Muijsers, A. O., A structure for the yeast prohibitin complex: Structure prediction and evidence from chemical crosslinking and mass spectrometry. *Protein Science*, **2002**, 11(10), 2471-8.

28. Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D., *Molecular biology of the Celll*. 3rd Ed., Garland publishing, INC., **1994**.

29. Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G., Mass spectrometry of peptides and proteins. *Methods*, **2005**, 35(3), 211–222

30. Lim, H. K.; Stellingweif, S.; Sisenwine, S.; Chan, K. W.; Rapid drug metabolite profiling using fast liquid chromatography, automated multiple-stage mass spectrometry and receptor-binding. *Journal of Chromatography A*, **1999**, 831, 227–241.

31. Croker, C. G.; Pearcy, J. O.;Stahl, D. C.; Moore, R. E.; Keen, D. A.; Lee, T. D., An Expert Virtual Instrument Approach to the Automated, Data-Dependent MS/MS and LC/MS/MS Analysis of Proteins. *Journal of Biomolecular Techniques*, **2000**, 11(3), 135-141

32. Masselon, C.; Anderson, G. A.; Harkewicz, R.; Bruce, J. E.; Pasa-Tolic, L.; Smith, R. D., Accurate Mass Multiplexed Tandem Mass Spectrometry for High-Throughput Polypeptide Identification from Mixtures. *Analytical Chemistry*, **2000**, 72, 1918-1924.

33. Wilson, J.; Vachet, R. W., Multiplexed MS/MS in a Quadrupole Ion Trap Mass Spectrometer. *Analytical Chemistry*, **2004**, 76, 7346-7353

34. Wu, S.; Kaiser, N. K.; Meng, D.; Anderson,G. A.; Zhang, K.; Bruce*, J. E., Increased Protein Identification Capabilities through Novel Tandem MS Calibration Strategies. *Journal of Proteome Research*, **2005**, 4(4), 1434–1441.

35. Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson J. K., UPLC/MS^E; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Communications in Mass Spectrometry*, **2006**, 20, 1989–1994.

36. Bateman1, K. P.; Castro-Perez, J.; Wrona, M.; Shockcor, J. P.; Yu, K.; Oballa R.; Nicoll-Griffith, D. A., MS^E with mass defect filtering for in vitro and *in vivo* metabolite identification. *Rapid Communications in Mass Spectrometry*, **2007**, 21, 1485–1496.

37. Chowdhury, S. M.; Munske, G. R.; Tang, X.; Bruce, J. E., Collisionally Activated Dissociation and Electron Capture Dissociation of Several Mass Spectrometry-Identifiable Chemical Cross-Linkers. *Analytical Chemistry*, **2006**, 78(24), 8183–8193.

38. Soderblom, E. J.; Goshe, M. B., Collision-Induced Dissociative Chemical Cross-Linking Reagents and Methodology: Applications to Protein Structural Characterization Using Tandem Mass Spectrometry Analysis, *Analytical Chemistry*, **2006**, 78, 8059-8068.

39. Angerer,L.; Davidson, N.; Attardi, G., An Electron Microscope Study of the Relative Positions of the 4s and Ribosomal RNA Genes in HeLa Cell Mitochondrial DNA. *Cell*. **1976**, 9(1), 81-90.

40. Rink, H., Solid-Phase Synthesis of Protected Peptide Fragments using a Trialkoxydiphenyl-Methylester Resin. *Tetrahedron Lettters*, **1987**, 28(33), 3787-3790.

41. Soderblom, E. J.; Goshe, M. B., Collision-Induced Dissociative Chemical Cross-Linking Reagents and Methodology: Applications to Protein Structural Characterization Using Tandem Mass Spectrometry Analysis, *Analytical Chemistry*, **2006**, 78, 8059-8068.

42. Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A., Complementary Sequence Preferences of Electron-Capture Dissociation and Vibrational Excitation in Fragmentation of Polypeptide Polycations. *Angewandte Chemie International Edition*, **2006**, 45, 5301 –5303.

43. Kam, N. W. S,; Jessop, T. C.; Wender, P. A.; Dai, H., Nanotube Molecular Transporters: Internalization of Carbon Nanotube−Protein Conjugates into Mammalian Cells. *Journal of American Chemistry Society*, **2004**, 126(22), 6850-6851.

44. Hardouin, J.; Canelle, L.; Vlieghe, C.; Lasserre, J.; Caron, M.; Joubert-Caron, R., Proteomic Analysis of the MCF7 Breast Cancer Cell Line, *Cancer Genomics and Proteomics*, **2006**, 3, 355-368.

45. Sarvaiya, H. A.;Yoon, J. H.; Lazar3, I. M., Proteome profile of the MCF7 cancer cell line: a mass spectrometric evaluation. *Rapid Communications in Mass Spectrometry*, **2006**, 20, 3039–3055.

46. Fountoulakis, M.; Tsangaris, G.; Oh, J.; Maris, A.; Lubec, G., Protein profile of the HeLa cell. *Journal of Chromatography A*, **2004**, 1038, 247–265.