ABSTRACT

Title of Dissertation: SOLUTION ISOELECTRIC FOCUSING AND

ITSAPPLICATION IN COMPARATIVE PROTEOMIC

STUDIES OF NUCLEAR PROTEINS

Yanming An, Doctor of Philosophy, 2005

Dissertation Directed by: Professor Catherine Fenselau

Department of Chemistry and Biochemistry

In proteomic research, experimental and computational approaches are combined to provide global analysis of the entire proteomes of cells and tissues. The identification and quantification of multiple proteins, which constitute a specific biological system, are important for understanding complex problems in biology. The coupling of highly efficient separations and mass spectrometry instrumentation is evolving rapidly and is being widely applied to problems ranging from biological function to drug development. Development of rapid and high-resolution separation technology is an important field in proteomics. In this study, a solution isoelectric focusing apparatus was modified and built into a two-dimensional separation method for peptides. Newly commercialized isoelectric membranes, which carry immobilized ampholytes, were integrated to establish the pH boundaries in this apparatus. High-performance liquid chromatography was employed as the second dimension, integrated with mass spectrometry. An insoluble nuclear protein fraction was used for

optimization and evaluation of this method. The insoluble nuclear proteins were recovered from the nuclei of human MCF-7 human cancer cells and cleaved enzymatically. The resulting peptides were analyzed by the two-dimensional separation method, which coupled solution isoelectric focusing with reversed-phase liquid chromatography interfaced with mass spectrometry. A total of 281 peptides corresponding to 167 proteins were identified by this experiment. The high sample capacity and concentration effect of isoelectric focusing make it possible to detect relatively low abundance proteins in a complex mixture. This two-dimensional separation method dramatically improves peptide detection and identification compared with a single dimension LC-MS analysis. This method has been demonstrated to provide efficient and reproducible separation of both protein and peptides.

The two-dimensional separation method was combined with proteolytic isotopic labeling for comparative analysis of protein expression in different cells. Abundances of nuclear proteins from three different drug resistant MCF-7 cancer cell lines were compared to those from the drug susceptible parent cell line using this combined strategy. The abundances of 19 proteins were found to be significantly changed. Their functions are considered in relation to potential mechanisms of in drug resistance.

SOLUTION ISOELECTRIC FOCUSING AND ITS APPLICATION IN

COMPARATIVE PROTEOMIC STUDIES OF NUCLEAR PROTEINS

by

Yanming An

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:
Dr. Catherine Fenselau, Chair
Dr. Elisabeth Gantt
Dr. Douglas Julin
Dr. Sang Bok Lee
Dr. Peter Gutierrez

Dedication

This dissertation is dedicated to my parents

Liancheng An and Baoxia Wei

who brought me into the wonderful world

Acknowledgements

I would like to thank my advisor Dr. Catherine Fenselau for accepting me and directing my research. She provides me a relaxed space to develop my skills and capabilities. I have grown from a student to an independent scientist under her direction in research and personality. The period working in her lab will become one of the most wonderful memories in my career. I would also thank all of my lab colleagues for their support and friendship.

I would like to thank Dr. Cheng Lee who introduced me to the field of proteomics. I learned a lot in his lab.

I would like to thank Dr. Peter Guterieez for his help in operation of the mass spectrometry instrument and database searching. I could not have finished my work without his support and instruction.

I would like to thank Dr. Zongming Fu. Dr. Fu developed the nuclei isolation method and left all his samples to me. I couldn't have done my work without his generosity.

I would like to thank Dr. Nathan Edwards for his advice in the data processing and bioinformatics.

I would like to thank all of my friends here.

Table of Contents

List of Tables

List of Figures

Chapter 1: Overview

Proteins are the working molecules in most biological processes and directly displaying and studying proteins, so called proteomics, provides an attractive way to understand complex problems in biology. Unlike the genome, the proteome is not a fixed feature of an organism. Protein technologies required to separate proteins, identify them and study their modifications are not straightforward. A huge amount of effort has been invested in developing robust techniques for proteome analysis.

The present project aims to develop and validate a two-dimensional protein/peptide separation strategy based on free solution isoelectric focusing and reversed-phase liquid chromatography. Our specific goals include: (1) develop and optimize free solution isoelectric focusing method for peptide separation; (2) apply this technique to nuclear proteins from MCF-7 human cancer cells; (3) integrate the two-dimensional separation method with ^{18}O labeling for comparative studies of differences in protein abundances between drug resistant and drug susceptible cells (4) based on the functions of proteins whose abundances are changed, consider possible mechanisms of drug resistance.

Isoelectric focusing (IEF) is widely used in gel based protein separation and also in capillary electrophoresis (CE). Because of the media involved in gel and CE, the loading capacity is limited and proteins with relatively low abundances are hard to detect. Moreover, they both require complex and expensive instruments to perform. The pH range of these two methods is usually 3-10, which limits the separation of proteins/and peptides with extreme acidic or basic isoelectric points.

The device used in our lab for solution isoelectric focusing is simple, easy to assemble and the experimental details will be discussed later. It can separate samples ranging from 0.5mg to 3mg without significant loss. Our experiments show that this fractionation step is compatible with down-stream analyses such as two-dimensional gel electrophoresis, reversed-phase liquid chromatography and mass spectrometry. This separation strategy is an off-line procedure, so we don't have to employ any instrumental modification to integrate the two steps together. Each fraction is collected and processed individually when the isoelectric focusing is completed. Thus it is convenient to analyze several fractions with mass spectrometry without repeating the whole process. However, the separation may take a relatively longer time and it needs monitoring.

In addition, isotope labeling is implemented to analyze the variation of protein abundances in drug resistance. ^{18}O labeling is an *in vitro* post-digestion labeling procedure. Previous work in our laboratory and elsewhere has proven that these labels are stable and suitable to a wide range of sample types, including cultured cells, microorganisms and tissues[1-3].

Chapter 2 introduces the protein technologies which are widely used now, including mass spectrometry, separation methods and quantitative strategies. Chemotherapy and drug resistance are briefly introduced followed with a description of the model organisms.

Chapter 3 presents the initial study in our laboratory on the performance of the solution isoelectric focusing device for protein separation. The resolution, sample recovery and compatibility were investigated for several devices.

Chapter 4 presents the results of nuclear peptide separation by the two-dimensional strategy and comparative proteome analysis of four cell lines. Some interesting results in separation and quantification are discussed in detail.

Chapter 5 presents conclusion for all of the work, and a future perspective.

Chapter 2: Introduction

2.1 Proteome and proteomics

The proteome encompasses the entire complement of proteins expressed in a certain type of cell or tissue at a single time. The term proteome has been widely used since it first appeared in late 1994 at the Siena 2-D electrophoresis meeting[4]. Proteomics is defined as the systematic analysis of proteins expressed in a cell or tissue at a given time or under certain environmental conditions. The success of sequencing the entire human genome along with the genomes of many other species, has led to an understanding of the complexity of organism at the level of information content. However, DNA stores the biological information of an organism rather than building a living unit. An eventual understanding from gene to biological function still relies on the study of proteins, the gene products. The identification and quantification of the multiple proteins that constitute and control a particular process are important for understanding the regulation of biological systems.

While the genome may be considered to be static, the proteome is highly dynamic. The protein complement may change with the stage of development, health or the environmental conditions. The same protein can be post-translational modified in many ways, such as phosphorylation, glycosylation and ubiquititation, which are not accessible from gene sequences[5-7]. In addition, mRNA levels do not necessarily correspond to the abundance of proteins expressed in biological tissues[8-10]. Considering the great complexity and the huge dynamic range of protein expression (at least 5 orders of magnitude), proteome research requires robust protein

4

fractionation and separation techniques. Moreover, with the appearance of soft ionization methods, mass spectrometry plays an important role in protein analysis.

2.2 Mass spectrometry

A mass spectrometer is an instrument that measures rapidly moving ions on the basis of their mass-to-charge ratio, m/z. There are several types of mass spectrometers available from instrument manufacturers. All types of mass spectrometers consist of the components shown in Figure 2-1. The inlet is an interface through which sample is introduce into the ion source. The sample molecules are evaporated and converted to ions at the ion source with any of several ionization techniques. The gaseous ions are separated in the mass analyzer according to their mass to charge ratios. The detector senses the arrival of ion streams and produces an electrical signal, which is processed, recorded and displayed by a computer. In addition, mass spectrometers require a vacuum system to maintain a low pressure through the ion pathway.

Mass spectrometry is widely used as an analytical tool with the ability to provide information about (1) elemental composition of molecules; (2) chemical structure of organic, inorganic and biological compounds; (3) qualitative and quantitative compositions of components in complex mixtures; and (4) isotopic ratios of atoms in sample. However, its application to biological molecules, such as proteins and DNA, was limited historically by the ionization technology. Since the inventions of two ionization techniques, named electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), mass spectrometry is experiencing a rapid growth in biological applications. The major advantages of mass spectrometry include high

sensitivity, speed of analysis and the large amount of information generated in experiments.

Electrospray ionization (ESI) was first described by Fenn and co-workers in 1985[11]. In ESI, an acidic solution containing peptides is sprayed through a small-diameter needle. A high positive voltage is applied on the needle, forming an electric field between the needle and the mass spectrometer. Droplets of the solution are positively charged at the acidic condition and move towards the detector. During the movement, droplets evaporate and split into a population of small droplets with the help of gas flow and heat. This process is repeated until it yields desolvated positively charged analyte ions. One characteristic of ESI is the production of multiply charged ions. Under acidic condition, the basic sites in the analyte molecules tend to be protonated, including the N-terminal amino acid, and the basic side chains of lysine, arginine and histidine residues. As a result, analytes with a mass in excess of the mass range of the instrument are still detectable, since the measurement is based on the mass to charge ratio of the ions. A significant feature of ESI is its compatibility with high-performance liquid chromatography (HPLC). Working as a detector, mass spectrometry can not only provide the elution profile but also characterize the effluents, which conventional optical detectors cannot do. Because the standard HPLC column (1mm i.d.) has a relatively high flow rate (50μm/min), a small-scale capillary column (75μm i.d.) was employed using a lower flow rate (2μL/min). Nanospray is an additional ionization technique, which is based on the same principle as ESI, operating on a significantly smaller needle diameter and flow rate (0.02μL/min). Nanospray is more sensitive than ESI and it is more suitable for

limited sample volumes. In a word, electrospray (and nanospray) is a "soft" ionization technique, which has been successfully used on a variety of mass spectrometers.

Matrix-assisted laser desorption/ionization (MALDI) was introduced by Karas and Hillenkamp[12] the same year as a related publication by Tanaka[13]. For this method, the sample is combined with a solution of an organic "matrix" compound, which has strong UV absorbance. As the solvent dries, the sample molecules are co-crystalized into the matrix. As the sample target is exposed to a pulsed UV-laser light, the matrix is vaporized and carries sample ions into gas phase. As with ESI, the ionization proceeds by protonation under the acidic conditions. In MALDI ionization, singly charged ions are predominant. One character of MALDI is the tolerance of a variety of contaminating species, including some buffers, which usually disrupt ESI. In practice, MALDI is mostly combined with time-of-flight (TOF) mass analyzer providing high sensitivity and resolution in proteome experiments.

Mass analyzers in mass spectrometers separate and determine the m/z ratio of analyte ions. There are two classes of ions seen in mass spectrometry experiments: molecular ions, providing molecular weight information for an analyte, and fragment ions, from which structure information can be derived.

Two types of mass analyzers that are commonly used in mass spectrometers are the ion trap (IT) and the time-of-flight (TOF). In an ion trap mass analyzer, rf-voltages are applied to produce a trapping region, where ions of all m/z oscillate. Subsequently, an m/z dependent rf-voltage is applied to eject ions of increasing mass to the detector. Ion trap has extremely good sensitivity because all of the ions from the ion source can be effectively trapped and transported to the detector. For mass

analysis in time-of-flight, an ion is given a certain amount of kinetic energy by acceleration in an electric field. Then the ion flies into a field-free region with a velocity inversely proportional to its m/z value. The time it consumes to travel through the length of the region is used to calculate the m/z value. A significant character of time-of-flight is the high mass (or m/z) range.

.
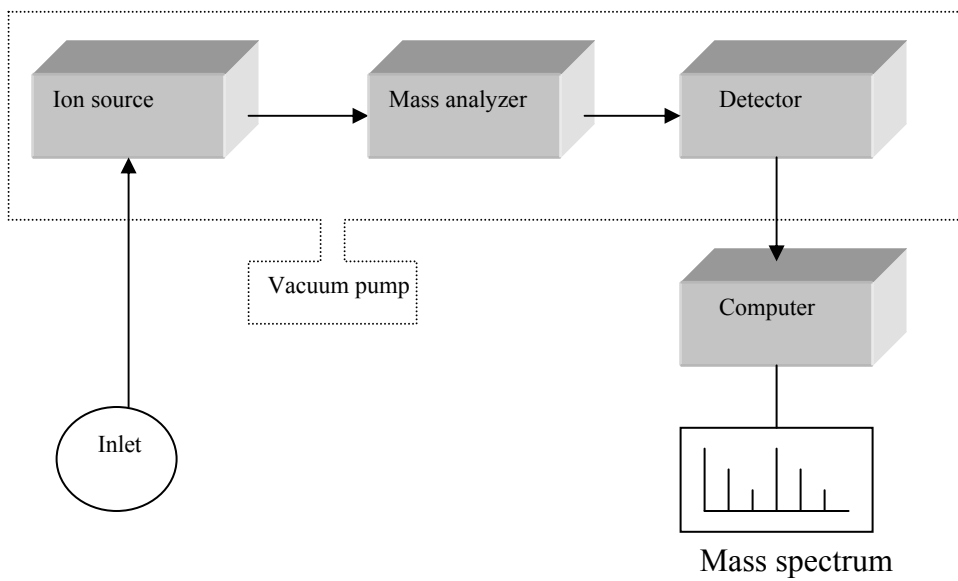
Figure 2-1 Scheme of a mass spectrometer

Tandem mass spectrometry (MS/MS) refers to two stages of mass analysis carried out in series in one experiment. After the first mass analysis, ions with a specific m/z value are isolated for collision-induced dissociation (CID). The activation is usually carried out in a collision cell, where each selected ion undergoes a number of collisions with gas molecules contained in the region which induce breakage of covalent bonds to produce ionic and neutral species. The resulting ionic products are referred to as "product ions" and analyzed in the second mass analyzer. The doubly (or triply) protonated peptide ions produced in ESI are predominantly fragmented by low energy collisions in ion trap and quadrupole time-of-flight mass spectrometers. The product ions are defined in two sets of ions: a-, b- and c- ions all contain the N-terminus of the peptide, and b- ions are dominant; x-, y- and z- ions contain the C-terminus of the peptide, and y- ions are major series (see Figure 2-2). The mass difference between two conjunct b-series or y-series ions reflects the mass of an amino acid residue. Complete interpretation of product ion spectra deduces the entire or partial amino acid sequence of a peptide. Figure 2-2 shows the fragmentation reactions of the peptide induced by low energy collisions.
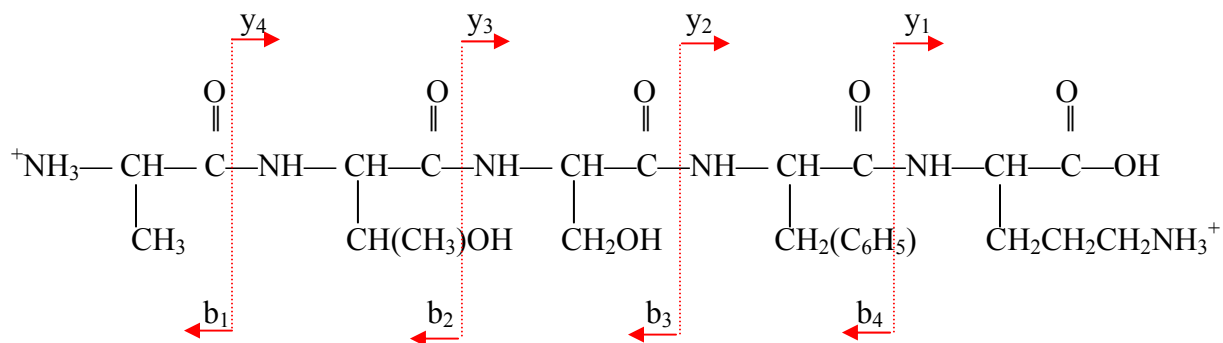
Figure 2-2 Scheme of peptide fragmentation by collision induced dissociation (CID)

2.3 Protein identification

Protein identification is based on the concept that different proteins could be distinguished not only by different functions but also by differences in amino acid sequences. Initially, proteins have been sequenced and identified by *de novo* sequencing, most frequently by Edman degradation of the proteins or their peptide fragments[14]. These partial sequences are used to assemble the complete protein sequence. But this technique suffers from low sensitivity, slowness and the complexity of cellular proteins. Following the accomplishment of the human genome and the genomes of other species, the expanded sequence databases made it possible to identify proteins by using short peptide sequence information. This is done by correlating information extracted from mass spectrometric data of a protein or peptide with sequences in databases. Correlation of experiment data with databases also depends on the development of novel search algorithms. Protein identification has been dramatically enhanced by convergence of the complete sequence databases, mass spectrometry techniques and various search algorithms.

Two mass spectrometry based protein identification methods are widely used: peptide mass mapping and microsequencing. The principle behind peptide mass mapping is that, after proteolysis with a specific protease, proteins of different amino acid sequences will generate different groups of peptides, whose masses make up unique mass fingerprints. Therefore, if a specific protein is contained in a sequence database, it can be identified by matching experimental masses with peptide mass predicted within the database. Typically, the method includes the following sequence of steps:(i) the protein sample is digested into peptides using sequence-specific

cleavage reagents such as trypsin so that the termini of peptides can be fixed for the search; (ii) peptide masses are measured in a mass spectrometer; (iii) proteins in the database are digested following the same proteolytic rules applied in the experiments to generate a theoretical peptide mass list; (iv) an algorithm is used to compare the experimental masses with the theoretical masses for each protein in the database and to assign a score for each hit which ranks the quality of the match. Peptide mass mapping is most popular for identification of proteins purified by 2D-PAGE, where proteins are usually purified and where approximate mass and isoelectric point information can also be obtained to aid identification. This method is not suitable for identification of protein in mixture.

The second strategy, microsequencing is widely used for large-scale protein identification. The amino acid sequence of a peptide combined with its mass is more constraining than its mass alone for protein identification. Peptide sequence information is commonly obtained by tandem mass spectrometry. Because of the complicacy of CID spectra, algorithms have been developed to use uninterpreted fragment ion pattern and the mass of the parent ion as input for sequence database searching. The algorithm first creates a list of peptides whose masses are equal to the observed parent ion mass by searching the database. Then it calculates the masses of the band y- fragment ions and generates a theoretical CID spectrum for each candidate peptide. The algorithm compares the experimental spectrum with the theoretical spectra and generates a score for each peptide which reflects the confidence of the match. Each protein is identified by finding as many derived peptides as possible.

2.4 Separation of protein and peptide mixtures

Although mass spectrometry is a powerful tool for proteomics analysis, it is impossible to get rapid and accurate identification and quantification of proteins without also involving various separation techniques. A key aspect of proteomics is the development of multiple separation methods to resolve the individual components of complex protein and peptide mixtures before mass spectrometry analysis. Figure 2-3 shows the separation methods most commonly employed in proteomics studies. Gel electrophoresis has been well established, is easy to carry out, compatible with mass spectrometry and has high resolution power. Several in-solution methods are listed, which are often combined together to provide rapid and high-throughput analyses for proteomics.
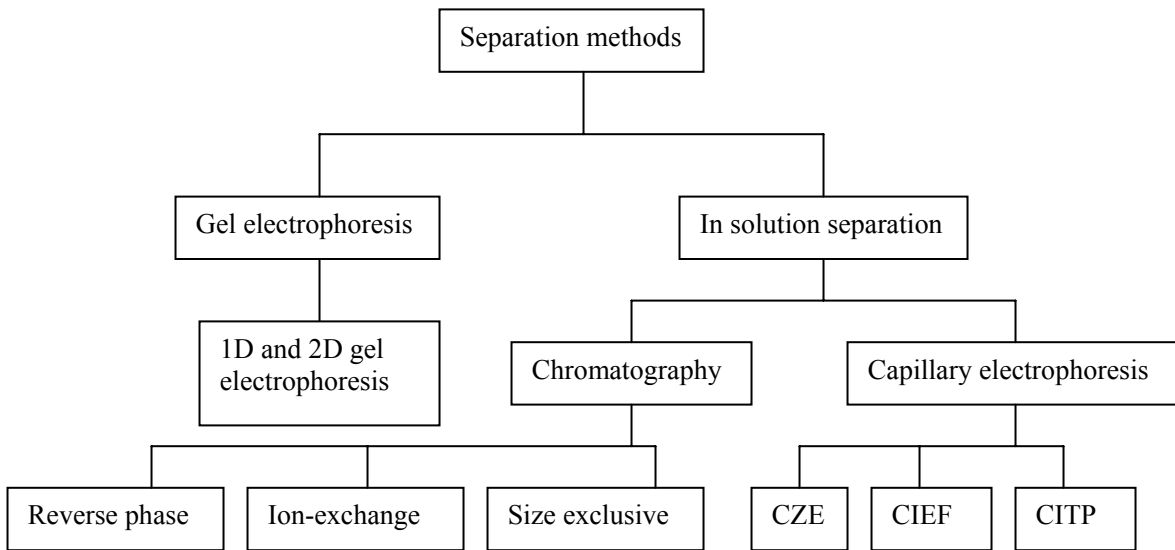
Figure 2-3 Classification of fractionation methods used in proteomics

2.4.1 Two-dimensional gel electrophoresis

Electrophoresis is based on the migration of charged proteins in an electric field. Two-dimensional gel electrophoresis is the most widely used method for protein separation, which combines two types of electrophoresis: isoelectric focusing (IEF) and polyacrylamide gel electrophoresis (PAGE). IEF is a procedure used to separate proteins according to their isoelectric points (pI). This is the characteristic pH value at which the net charge of a protein is zero. A pH gradient is established by the distribution of ampholytes, low molecular weight organic acids and bases, across the gel. After the protein mixture is applied, each molecule migrates in the electric field until it reaches the pH matching its pI. The introduction of immobilized pH gradient (IPG) gels eliminated the poor reproducibility associated with ampholytes[15-19]. In an IPG gel, the pH gradient is covalently bonded to the acrylamide gel matrix, resulting in a stable gradient at all the pH values. The precast IPG gels are commercially available in a variety of narrow and broad pH ranges with the resolution as high as 0.01 pH unit.

The second dimension is called SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). In this step, the proteins are separated based on their molecular weight (MW). SDS-PAGE is carried out in gels made of cross-linked polyacrylamide. SDS is a detergent that bonds to proteins, so that the proteins are denatured and the intrinsic charges of proteins eliminated. The number of SDS molecules, and thus the number of negative charges associated with the protein, will depend on the length of the denatured protein. When the electrical field is applied,

smaller proteins migrate more rapidly in the gel matrix and larger molecules move slowly.

To perform 2D-PAGE, a protein mixture is dissolved in rehydration buffer, which contains nonionic detergent, denaturing and reducing reagents to increase the sample solubility during IEF. An IPG gel strip is soaked in the sample solution for several hours to suck the proteins into gel matrix, which is called rehydration. Then a high voltage is applied across the gel strip and proteins are focused at their pI. The IPG gel strip containing focused proteins is equilibrated in SDS buffer and sealed with a SDS-PAGE gel piece. In the second dimension, the proteins migrate perpendicularly to the IPG gel strip under the electric field. To visualize the spots, the gel can be stained by various methods, including silver staining, Coomassie brilliant blue staining and fluorescent staining. Protein spots can be excised for in-gel digestion. The resulting peptides are extracted and subjected to mass spectrometry for protein identification.

Two-dimensional PAGE which combines IEF and SDS-PAGE was introduced in 1975 by O'Farrell[20] and Klose[21] independently. Two-dimensional PAGE transited from a descriptive to an analytical technique in the 1990s. With the improvement of techniques, including IPG strips and mass spectrometry, 2-D PAGE has been widely used in proteomic studies[22-27]. However, it is not sufficient for proteome analysis because of some fundamental limitations. Usually, 2D-PAGE is able to separate proteins with the pI range of 3.0-10.0 and MW range of 20-200 kDa. Those proteins with pI and MW values exceeding these ranges are hard to detect. This problem could be partly resolved by extending the available pH gradients[19, 28]. In addition, it has been demonstrated that the limited dynamic range (max. 10^4) of 2D-PAGE has

hampered the ability to identify low-abundance proteins, which are usually 5 to 6 orders of magnitude lower than the most abundant proteins in complex biological samples[29]. This means that enrichment or prefractionation methods should be used to access the less abundant proteins. Moreover, more hydrophobic proteins, such as membrane proteins, tend to precipitate during IEF due to the zero net charge. Membrane proteins should be solublized in lipid bilayers and not in water. Detergents are used to mimic a lipid-like environment in aqueous solutions. This problem also links to the initial extraction and solubilization of membrane proteins and seems not to be resolved by a single solution.

2.4.2 Liquid-phase separation

Because of the limitations of 2D-PAGE method outlined above, several strategies have been proposed for more comprehensive proteome analysis based on liquid phase separation. Generally, protein/peptide mixtures extracted from cells or tissues are separated with high performance liquid chromatography (HPLC) or capillary electrophoresis (CE) at first, then the resulted fractions are subjected to mass spectrometry on-line or off-line.

2.4.2.1 High Performance Liquid Chromatography (HPLC)

Chromatography is a powerful separation technique applied in all branches of science. Because of its high resolving power, reproducibility and compatibility with electrospray mass spectrometry, the applications of high performance liquid chromatography have grown extensively in the past decade to meet the increasing needs of scientists for characterize complex mixtures.

In a chromatographic separation the sample is carried by a mobile phase, which could be a gas or a liquid. Then the mobile phase is driven through a stationary phase, which is fixed in a column or on a solid surface. The two phases are selected so that the components in the sample can distribute between the two phases to different degrees. The components, which stick strongly to the stationary phase, move slowly with the flow, while those components interacting weakly with the stationary phase travel rapidly. The difference of mobility results in separation of the sample components, which can be detected and analyzed qualitatively and quantitatively.

2.4.2.1.1 Reverse-Phase Liquid Chromatography (RPLC)

Reverse-phase liquid chromatography has become the most widely used of all the types of chromatographic procedures. In RPLC, the stationary phase is nonpolar, often a hydrocarbon (such as C4 or C18 chains), covalently bonded to the support particles of the packing. The mobile phase is relatively polar, usually a mixture of water and organic solvent. A reverse-phase separation is started from an aqueous solution containing low concentration of such solvents as methanol, acetonitrile, or tetrahydrofuran and the most polar component is eluted out first. As the organic component of the mobile phase is increased, the polarity is decreased and the less polar molecules are flushed out subsequently. The performance of the column is affected by the length of hydrocarbon chains: longer chains are more retentive. In practice, C4 columns are used for protein separation while peptides are fractionated with C18 columns. Currently, most of the LC separations in proteomics are performed in RPLC mode because of its compatibility with mass spectrometry.

2.4.2.1.2 Ion-exchange chromatography

Ion-exchange chromatography was first developed in the mid-1970s[30] and includes cation-exchange and anion-exchange chromatography, according to the stationary phase of the column. The active sites for cation-exchange resins are usually the sulfonic acid group-$SO_3^-H^+$ and the carboxylic acid group-COO^-H^+. The former is a strong acid and the later is a weak acid. Anion-exchangers contain tertiary amino groups $R-N(CH_3)_3^+OH^-$, a strong base, or primary amine groups $R-NH_3^+OH^-$, a weak base. The mobile phase in ion-exchange chromatography consists of an aqueous buffer of various concentrations of ionic compounds such as Na^+Cl^-. Ion-exchange processes are based on the exchange equilibrium between the ions on the resin surface and ions of the same sign in solution. In most applications of ion-exchange chromatography, elution is carried out by a salt solution with concentration increasing in steps. The analyte ions with the weakest static electric interaction with the resin are elution at lower salt concentration and ions strongly interacting with the resin are eluted at higher salt concentrations.

Ion-exchange chromatography has been used for protein and peptide separations for a long time and is often used as the first dimension in a 2-D LC separation. For example, Opiteck et al. described a 2-D orthogonal LC-MS analysis system for the separation of protein mixtures[31]. This system used cation-exchange chromatography and reversed-phase chromatography coupled by an eight-port valve. The RPLC effluent was detected by a UV detector and an electrospray mass spectrometer.

An innovative approach termed "multidimensional protein identification technology" (MudPIT) was introduced by Yates et. al. [32-33]. In this approach, a

protein mixture was digested and peptides were separated by a biphasic microcapillary column which is shown in Figure 2-4. The biphasic column was packed with strong cation-exchange and reversed-phase packing materials and connected to a microcross. The effluent from the column was sprayed into the mass spectrometer directly. Since introduced, MudPIT has been used for profiling a wide range of proteomes with little modification[35] and extended for quantitative analysis [36-38].
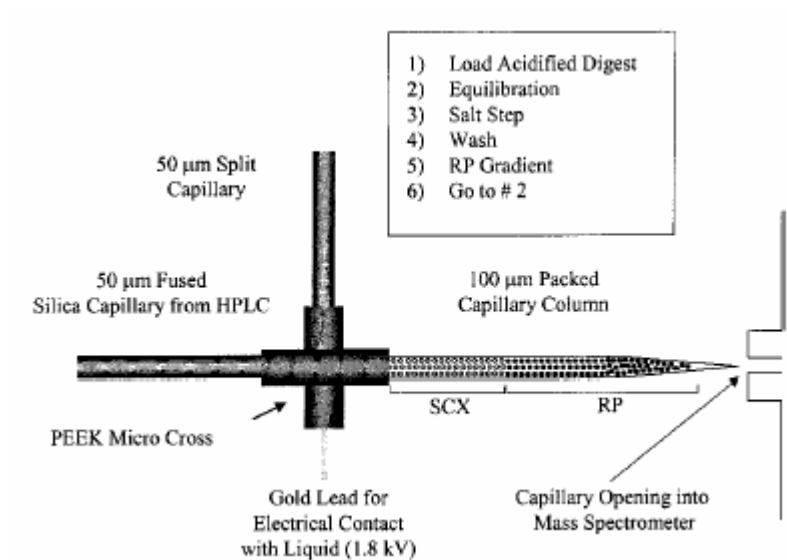
Figure 2-4. Multidimensional protein identification technology (MudPIT) electrospray interface including a biphasic microcapillary column packed with strong cation-exchange and reversed-phase packing material connected to a microcross[34].

2.4.2.1.3 Size-exclusion chromatography

Size-exclusion chromatography (SEC), which is also called gel filtration chromatography, is a powerful technique particularly useful for high-molecular-weight species. The particles packed in a size-exclusion column contain a network of uniform pores in which molecules are trapped and removed from the flow of mobile phase. The retention time of analytes depends on the size and shape of molecules. Molecules which are larger than the pore size of the stationary phase are excluded and eluted first, while molecules with significantly smaller size than the pore size (such as solvent and salt molecules) will penetrate throughout the pore maze and travel for the longest time. The intermediate-size molecules between the two extremes are eluted according to their diameters. Size-exclusion chromatography differs from other procedures in that no chemical or physical interaction between the analytes and the stationary phase is involved.

Size-exclusion chromatography is usually used as the first dimension in 2-D LC separation with high reproducibility and relatively short analysis time[39]. SEC is not that popular in proteome analysis because of its limited loading capacity and resolving power.

2.4.2.2 Capillary electrophoresis (CE)

Electrophoresis is a method to separate charged species based on their differential migration rates in a buffer across which has been applied a dc electric field. Electrophoresis separations are performed in two formats: one is called gel electrophoresis (mentioned above), and the other is capillary electrophoresis (CE). Capillary electrophoresis has become an important separation tool used by chemists

and life scientists since it was developed in 1970s[40]. CE has demonstrated the potential yielding high-speed, high-resolution separations of bio-polymers as well as small pharmaceutical reagents analyzed in the miniscule amounts. Sample size for CE is in the nanoliter range. Thus the sensitivity of this method is much higher than other separation methods on a mass basis. Additionally, the separated species are eluted from one end of the capillary, so quantitative detectors, such as UV absorption, mass spectrometry, can be used instead of the staining techniques of gel electrophoresis. Capillary electrophoresis is performed in several modes: capillary zone electrophoresis (CZE), capillary isoelectric focusing (CIEF) and capillary isotachophoresis (CITP).

2.4.2.2.1 Capillary zone electrophoresis (CZE)

CZE is the simplest and most universal of the techniques for the separation of analytes varying in size and character. To perform a CZE separation, the capillary is filled with a separation buffer with appropriate pH and the sample is introduced at the inlet. Both ends of the capillary and the electrodes from a high voltage power supply are placed into running buffer reservoirs. The applied potential causes the ionic species in the sample plug to migrate with mobility determined by their charges and masses, and eventually to separate into zones. The resolved zones with buffer regions between them pass a detector where information is collected and stored by a data processing system. Capillary zone electrophoresis provides a rapid separation technique for small ions, small molecule species and biopolymers, which usually takes several minutes to complete.

Capillary zone electrophoresis has been widely used for separation of peptides with high separation efficiency[41-43]. However, the limited sample injection volume makes it difficult to detect minor components in a complex mixture. So preconcentration-methods, such as transient isotachophoresis[44], are often used to satisfy the detection limits.

2.4.2.2.2 Capillary isoelectric focusing (CIEF)

Capillary isoelectric focusing is used to separate amphiprotic species, such as amino acids and proteins that contain a weak acidic and a weak basic group. An amphiprotic compound is capable of accepting and donating a proton in solution. As zwitterions the amphiprotic molecules bear both positive and negative charges in solution. The pH, at which the net charge of the zwitterions is zero, is called the isoelectric point (pI) and is an important physical character of amino acids. A CIEF separation is performed in a buffer mixture varied in pH continuously along the capillary length. This pH gradient is established from a mixture of different ampholytes in aqueous solution. Ampholytes are small compounds containing carboxylic and amino groups. Ampholytes mixtures with different pH ranges are available commercially. To perform a CIEF separation, the sample is dissolved in a buffer containing ampholytes of the desired pH range and transferred into a capillary tube. One end of the capillary is inserted into a basic solution such as ammonium hydroxide, and that also holds the cathode. The other end is in an acidic solution, such as acetic acid, that also holds the anode. When the voltage is applied, the ampholytes negatively charged would migrate towards the anode while the positively charged species would move to the cathode. In migration each species is continuously titrated

and eventually reaches the pH where its net charge is zero and stops moving. This process goes on for all ampholyte species and ultimately a continuous pH gradient is formed throughout the tube. Analyte ions migrate until they reach their pI and are sorted into several narrow bands located at certain pH value. In order to detect the sample bands, the focused zones are mobilized towards the detection point at one end. The mobilization can be accomplished by applying a pressure or vacuum at one end of the capillary, which is called hydraulic mobilization.

Shen et al. studied capillary isoelectric focusing of peptides and about 500-fold concentration factor was gained through focusing process[45]. And CIEF was proven to rapidly measure the isoelectric points and provide high resolution of up to 0.01 pI difference[45]. CIEF has been combined with mass spectrometry through an ESI interface to provide high efficiency, speed separation of protein and peptides[46].

2.4.2.2.3 Capillary isotachophoresis (CITP)

Capillary isotachophoresis, unlike other modes in capillary electrophoresis, employs a discontinuous electrolyte system. In a separation, the sample is injected between two buffers, the leading buffer containing ions of a higher mobility than any analyte ions and a terminating one containing ions with lower mobility than any of the sample ions. For the cation separation the leading buffer is connected to the cathode and the terminating buffer is to the anode. When the potential is applied, analyte ions migrate as in CZE and resolve into several adjacent bands sandwiched between the leading and terminating zones. After the equilibrium is formed, the bands move at the same velocity. The difference between CITP and CZE is that in CITP the

26

analytes bands are immediately adjacent to one another without a buffer band in between (See Figure 2-5).

Capillary isotachophoresis is used as an online preconcentration method prior to CZE, because both of them are performed under the same conditions. Gysler et al. employed a single CITP to concentrate recombinant cytokine fragments before separation and detection with CZE-MS[47].

Moreover, some of these techniques were combined together for multidimensional separation of proteins and peptides. Mohan et al. coupled CIEF with CITP/CZE through a microdialysis junction and demonstrated that this 2-D system provided a peak capacity of up to 1600[48]. Chen et al. integrated CIEF with RPLC by a series of short trap columns to separate yeast cell lysate[49].

Zone electrophoresis

Isoelectric focusing

Continuous pH gradient

Isotachophoresis

Terminating Buffer   Leading buffer
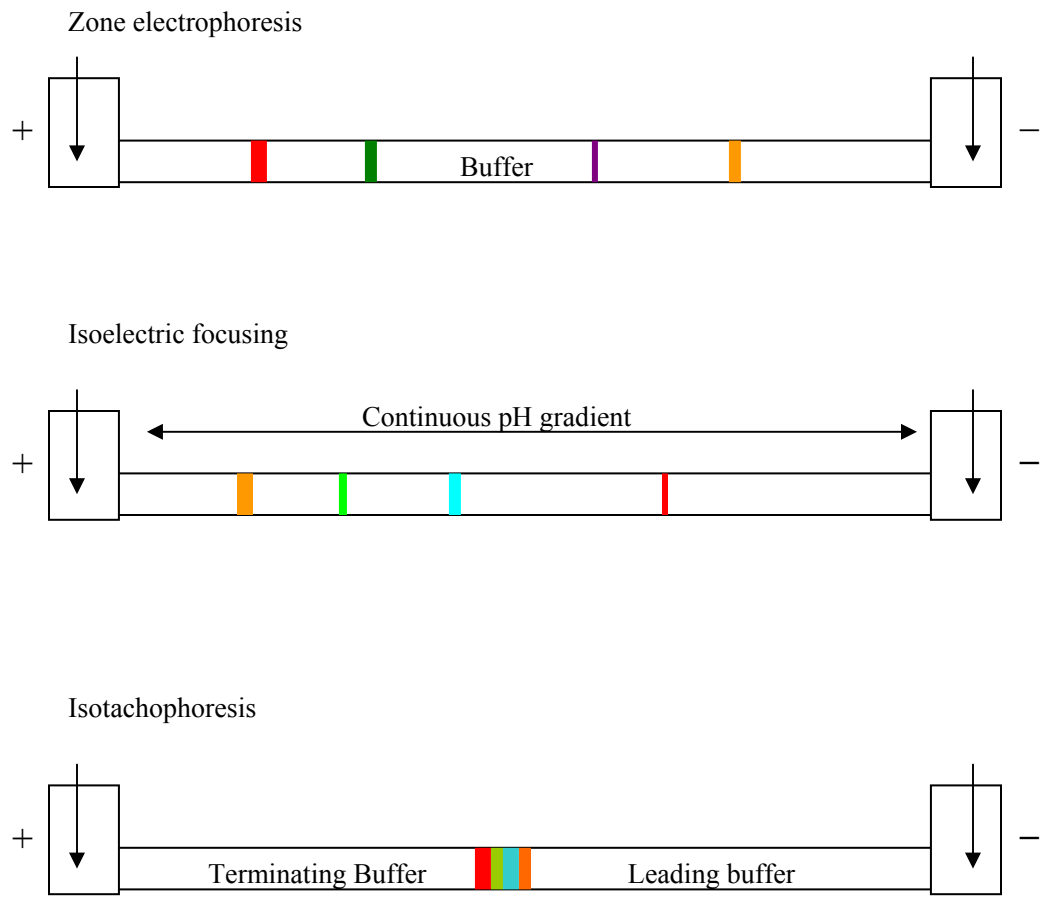
Figure 2-5 Procedures of capillary electrophoresis separations

2.4.2.3 Two-dimensional separations involving gels

Each dimension of 2-D gel electrophoresis has been combined with LC-MS/MS to characterize protein mixtures.

SDS PAGE coupled with LC-MS/MS has been used by many groups to characterize more hydrophobic proteins, such as membrane proteins and mitochondrial proteins. Taylor and colleagues analyzed mitochondrial proteins isolated from human heart cells by SDS PAGE as the first dimension and each band was excised and subjected to gel digestion. The result peptides were extracted from gel piece and further analyzed by RPLC-MS/MS[50]. By using the same method, Rezaul et al. identified 680 proteins which were mitochondrial or mitochondrial associated proteins[51].

Cargile et al. demonstrated the potential use of IPG as the first dimension in shotgun proteomics followed with LC-MS/MS[52-54]. The cytosolic fraction from E coli was digested and loaded on an IPG strip for isoelectric focusing. Then the strip was cut into several pieces and the peptides were extracted from the gel and separated by LC-MS/MS. More 6000 peptides and 1200 proteins were identified from 10 μg starting material[52].

2.4.3 Prefractionation techniques

The success of proteomic research relies on the rapid development of several areas: mass spectrometry, bioinformatics and separation technology. Large-scale analysis of complex mixture has been realized by coupling HPLC separation and MS. Thousands of proteins can be characterized by shotgun proteome technologies. However, cells are extremely complex and may contain hundreds of thousands of

proteins with various physical and chemical properties. And their expression levels may differ by 5-7 orders of magnitude within a cell. The relatively low abundance proteins are usually obscured by more abundant proteins when analyzed by gel or gel-free techniques. Therefore, most of the proteins identified in the experiments are highly abundance proteins. Thus, initial fractionation methods must be employed to get a better understanding of a cell.

2.4.3.1 Subcellular fractionation

Subcellular fractionation is the first step to reduce the sample complexity and can be efficiently combined with various strategies for downstream proteomic analysis. Progress in biological techniques has made it possible to isolate and purify organelles, such as nucleus, mitochondria, ribosome, Golgi apparatus, exosomes, lysosomes and peroxisomes. Subcellular fractionation separates organelles based on their physical or biological properties and usually consists of two major steps: (i) breaking the cellular organization (homogenization), and (ii) isolateing the different population of organelles from the homogenate. The collected cells are homogenized, and centrifuged at a low speed to spin down the pellet, which contains nuclei, cell debris and unbroken cells. Nuclei can be purified by further analysis and the other organelles can be separated from the supernatant by gradient centrifugation where sucrose is commonly used as a medium. Further purification of the isolated organelles is necessary for comprehensive analysis of total organelle proteomes.

As an example, Wu et al. characterized a stacked Golgi fraction using multidimensional protein identification technology[55]. The Golgi fraction was enriched from rat liver using two different sucrose step gradient centrifugations. In this study,

110 proteins identified were known Golgi residents and proteins localized in other organelles were reported functionally interfacing between Golgi, endoplasmic reticulum and cytoskeleton. Jiang and colleagues have constructed a protein database for rat liver by combining subcellular fractionation and 2-D LC-MS/MS[56].

2.4.3.2 Affinity chromatography

Affinity chromatography is a powerful protein fractionation method, which is based upon the specific interaction between the immobilized ligand and the target proteins. Affinity chromatography can be employed to reduce sample complexity before 2D gel or gel-free separation.

Affinity chromatography can be used to remove a specific protein or group of proteins which have high abundance to help detect proteins present at low concentration. A typical example is the removal of albumin and immunoglobulin G (IgG) from serum samples by affinity resins. The antibodies for these proteins are immobilized and cross-linked to the stationary phase of a column. When the sample flows through the column, the target proteins will bind to their antibodies and be retained on the column. Other proteins will come out. The column can be regenerated by changing buffer conditions[35, 57].

On the other hand, affinity chromatography can also help enrich low concentration proteins. In the study of phosphorylated proteins, immobilized metal ion affinity chromatography (IMAC) is often used to selectively enrich phosphoproteins and phosphopeptides. In this technique, metal ions including Fe (III), Al (III) or Ga (III) are chelated to a support. Because of the affinity of the metal ions for the phosphate moiety, phosphopeptides can be bound. The phosphopeptides

can be eluted from the column with high pH or phosphate buffers, the latter usually requiring a further desalting step before MS analysis[58]. As another example, Kaji and colleagues developed a strategy for large-scale identification of N-glycosylated proteins from a complex biological sample by using a lectin column to affinity capture glycopeptides[59].

2.4.3.3 Prefractionation of proteins and peptides based on their electrochemical properties

At present, two major approaches have been developed to fractionate proteins and peptides according to their isoelectric points: chromatofocusing and preparative IEF.

Chromatofocusing is an ion-exchange chromatography technique where the elution is realized by dropping the pH of the elution buffer so that proteins come out in the order of decreasing isoelectric point. Usually, the ion exchanger is equilibrated at a relatively alkaline pH and then the resin is titrated by amphoteric buffers to the lower pH in a linear manner. Components with a particular pI will focus and elute as a single peak during the pH gradient. Lubman and colleagues fractionated proteins using chromatofocusing followed by RPLC[60]. However, like conventional ion-exchange chromatography, this procedure introduces a huge amount of salt which can be a problem for down stream analysis.

There are two types of preparative IEF apparatus available now. The Rotofor has the longest history. The instrument setup is shown in Figure 2-6. It consists of 20 chambers separated by liquid-permeable nylon screens[61]. It uses the same principle as conventional isoelectric focusing utilizing ampholytes to establish the pH gradient throughout the chambers. During the separation, the whole cell rotates to avoid

overheating and gravity effect. After the separation, the fractions are collected simultaneously by 20 needles connected to a vacuum source. Recently, the Rotofor has been used as the first dimension in a 2-D methodology. Lubman et al. combined Rotofor with RPLC to analyze intact proteins from cancer cells[62]. Xiao and colleagues employed Rotofor and RPLC-MS/MS to separation peptides from human serum and resulted in 437 proteins identification[63].

The second system employs isoelectric membranes to realize separation. Isoelectric membranes are fabricated with immobilized ampholytes adjusted to a desired pH. Figure 2-7 shows the scheme for this kind of apparatus. The membranes work as pH boundaries, through which charged protein or peptide molecules are titrated. The analyte keeps moving driven by the electric field until it reaches its pI. Then it stays in the corresponding chamber. This technique was introduced by Righetti et al. [64], and has resulted in the large-scale apparatus-multicompartment electrolyzer (MCE) (sold by ProteomeSystems). Pedersen et al. used an MCE to fractionate a membrane fraction of yeast and identified 780 proteins, including 28% low abundance proteins and 49% membrane or membrane associated proteins[65]. Based on the same principle, Speicher et al. developed a smaller device and applied it in global analysis of human cancer cell proteomes [67-70](IEF Fractionator, Invitrogen Inc).
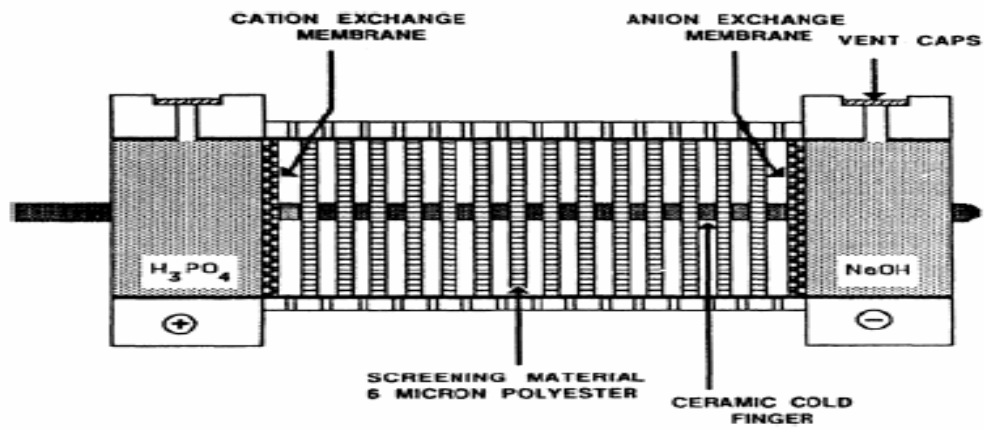
Figure 2-6 Scheme of Rotofor instrument[61]
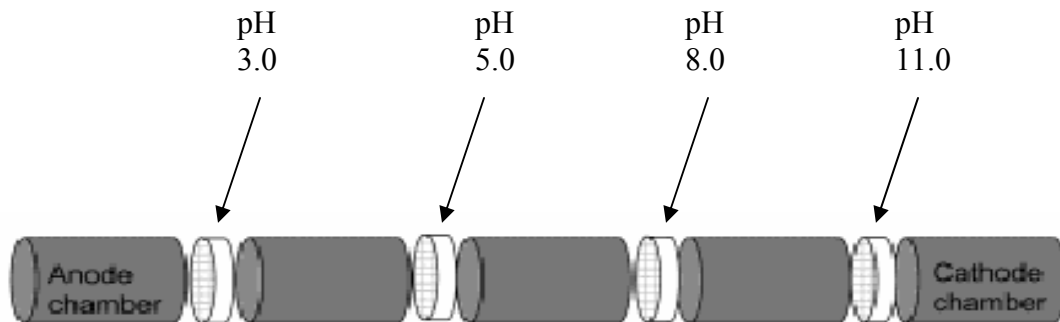
Figure 2-7 Scheme of protein prefractionation using isoelectric membranes[66]

2.5 Quantitative proteomics

Much effort in proteomics is focused on methods to effectively identify proteins. When cells are exposed to certain environments, protein expression levels can be changed by mechanisms involving cell survival. Protein degradation can also be modified. Thus, quantitative protein profiling is important to study protein dynamics involved in cellular responses. A variety of chemical, metabolic and enzymatic stable isotope labeling techniques have been developed to analyze relative protein expression in two or more cell lines. In these methods, one sample is labeled with a light reagent and the other one is labeled with a heavy reagent. The two samples are combined and analyzed by mass spectrometry. The ratio of the two isotopic peaks determined from the mass spectra can be used to calculate the relative protein abundance.

2.5.1 *In vivo* labeling

In this procedure, cells are grown in two separate media, one of which contains amino acids carrying heavy stable isotopes. Thus the protein components in one culture are labeled with heavy isotopes while the other culture is labeled with light isotopes. The two cell harvests are combined, the proteins extracted, and peptides are analyzed by MS based techniques. This method was introduced by Oda et al.[71]: two pools of proteins, one of which was extracted from bacteria grown in ^{15}N media, were combined and separated by gel electrophoresis. Then the proteins of interest were digested and analyzed by MS and the relative quantities were determined by the isotope distribution ratio. More recently, this method was combined with LC-MS/MS for analyzing the whole cell lysate[36, 72-75]. Media containing isotopically labeled

lysine and arginine has been used to study mammalian cells[37-38, 76]. *In vivo* isotopic labeling has been proven to be an effective method for quantitave proteomics. However, this approach cannot be used to label clinical and animal samples.

2.5.2 ***In vitro* labeling**

2.5.2.1 Isotope-codes affinity tag (ICAT) labeling

The isotope-coded affinity tag (ICAT) approach was introduced by Aebersold and coworkers[77]. As shown in Figure 2-8, the ICAT reagent consists of a biotin moiety (affinity tag), an isotopically labeled linker (^{1}H/^{2}H) and a thiol-reactive group (which alkylates cysteinyl residues). The proteins extracted from two separate cell cultures are labeled with ICAT reagents on the cysteine residues before they are combined and digested. Then the peptide mixture is eluted through an avidin affinity column, interacts with the biotin tag to isolate the labeled peptides. Qualitative and quantitative information can be obtained from mass spectra. This approach is limited to quantify proteins containing cysteines.
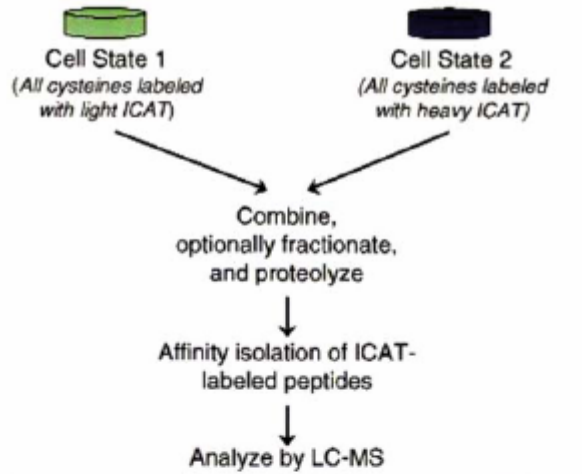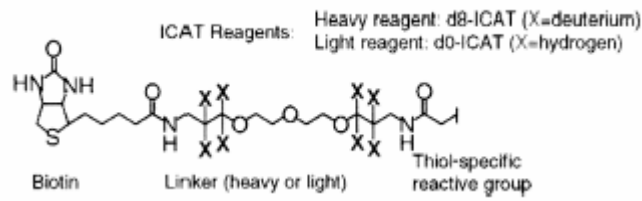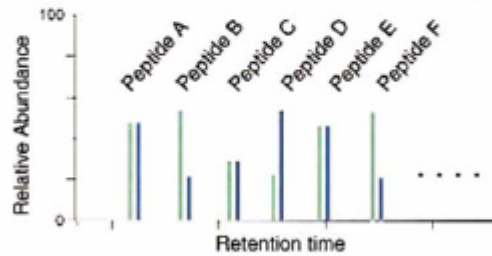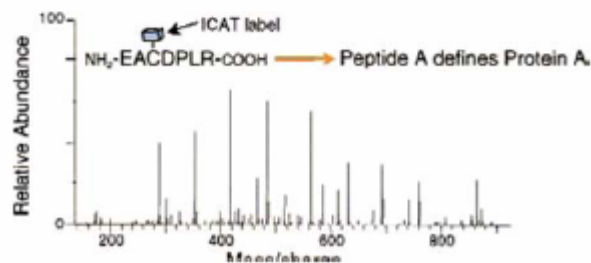
Figure 2-8 Structure of the ICAT reagent and the ICAT strategy for quantitative proteomic analysis[77]

2.5.2.2 Enzyme catalyzed ^{18}O labeling

Enzyme-catalyzed ^{18}O-labeling is a simple and convenient stable isotopic coding strategy, in which proteins are enzymatically digested in normal or ^{18}O water[1-3, 78-80]. The reaction mechanism is shown in Figure 2-9. The enzyme covalently bonds the target residue, e.g. arginine and lysine in the case of trypsin, and that intermediate is then decomposed by $H_2^{18}O$ or $H_2^{16}O$. The enzyme recognizes and re-bonds the C-terminal residues of the peptide and decomposition introduces the second ^{18}O atom. Members of the family of serine proteases (including trypsin, Glu-C endoprotease, Lys-C endoprotease, chymotrypsin) incorporate two atoms of ^{18}O into the C-termini of peptides, resulting in a mass shift of +4 Da for each peptide fragment. It was reported that the covalent bond between oxygen atoms and the carbonyl carbon are resistant to chemical back exchange and stable through liquid chromatography and ionization process[81-82]. Enzymatic back exchange is controlled by removing immobilized trypsin, or chemically inactivating it. The comparative proteomic studies are carried out by mixing two pools of peptides, one of which is labeled with ^{18}O and the other is labeled with ^{16}O, and analyzed by mass spectrometry.
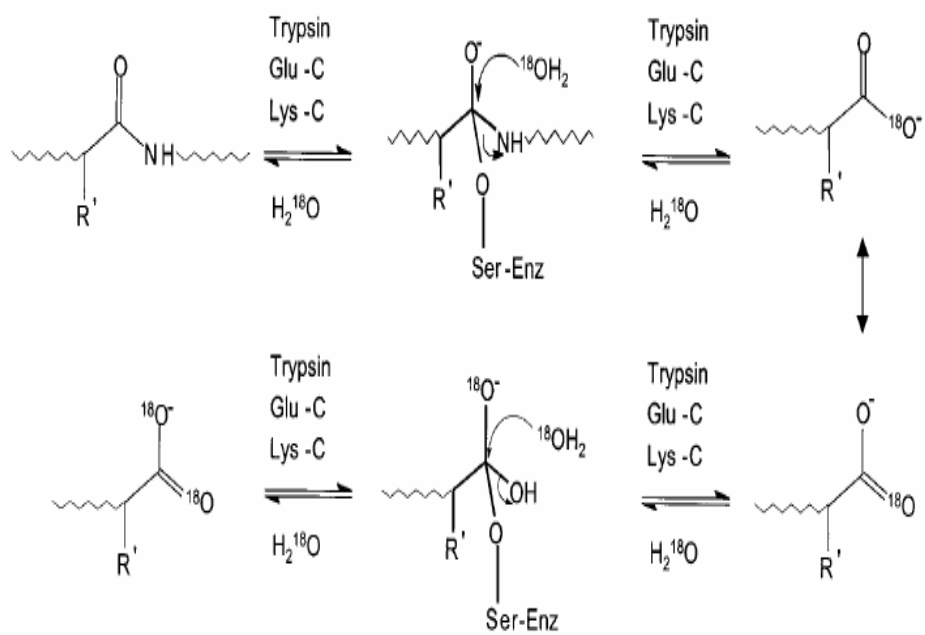
Figure 2-9 Mechanism of enzymatic ^{18}O labeling reaction[2]

The ratios of $^{18}O/^{16}O$ peptides are calculated using equation 1[2]

$$\text{Ratio I} = \frac{I_4 - \dfrac{M_4}{M_0} I_0 - \dfrac{M_2}{M_0}\left[I_2 - \dfrac{M_2}{M_0} I_0\right] + \left[I_2 - \dfrac{M_2}{M_0} I_0\right]}{I_0} \qquad \text{equation 1}$$

This equation takes into account not only the ratio of labeled and unlabeled peak areas, but also contributions from incomplete labeling, and it corrects for contributions from the natural isotopic distribution of the peptide. In this equation, I_0, I_2 and I_4 are the observed peak areas for the monoisotopic peak for the peptides without ^{18}O label, the peak 2 Da higher, and the peak 4 Da higher, respectively. M_0, M_2 and M_4 are the theoretical peak areas for the monoisotopic peak of a peptide with known composition, the peak 2 Da higher and the peak 4Da higher, respectively. The theoretical distribution of the unlabeled and labeled peptides is assumed to be identical.

2.6 Cancer chemotherapy

Chemotherapy is the treatment of cancers with drugs that kill cancer cells. It is the most effective treatment for metastatic tumors. Chemotherapy is used to treat patients whose cancer does not respond to local excision or radiation (about 50% of total cancer cases), and patients with haematological malignancies. The anticancer drugs are often administered as a pill or injected into vein, so they can travel throughout the body in the bloodstream. This makes chemotherapy different from surgery and radiation which are confined or localized to one area or organ. Traveling throughout

the whole body, the drugs can access cancer cells which may have spread to other organs and treat cancers systemically[83].

There are more than 50 chemotherapeutic drugs available now and more are being developed and tested in clinical trials. Many anticancer drugs target DNA in the nuclei of the cancer cells to interfere with the ability of cells to grow or multiply. Based on their actions, the drugs can be classified into the following types: covalent DNA binding drugs, non-covalent DNA binding drugs, antimetabolites, inhibitors of chromatin function, and drugs affecting endocrine function[83].

2.7 Drug resistance

Tumor cells are able to become simultaneously resistant to anticancer drugs, which is called drug resistance. Drug resistance is a major problem in cancer treatment. There are two general classes of resistance to anticancer drugs: intrinsic and acquired resistance. Intrinsic resistance impairs delivery of anticancer drugs to tumor cells resulting from poor absorption of orally administered drugs or increased drug metabolism that lowers the diffusion of drugs from blood to the tumor mass[84, 85]. Acquired resistance arises in the cancer cell itself due to genetic and epigenetic alterations that affect drug sensitivity. When tumor cells develop resistance, they usually become resistant to several drugs, especially to those structurally and functionally related. This phenomenon is called multidrug resistance (MDR) [86].

Drug resistance is a significant obstacle to successful chemotherapy. It is important to understand the mechanisms of drug resistance in order to predict and overcome it by improving chemotherapy. Cellular mechanisms of drug resistance have been studied for three decades. Figure 2-10 illustrates different mechanisms of

41

cellular drug resistance described during these years. Increased drug efflux generally results from the expression of ATP-dependent efflux pumps, which belong to the family of ATP-binding cassettes (ABC) transporters[86-88]. Reduced drug uptake refers to the failure of water-soluble drugs to accumulate in cells without evidence of increased efflux[89]. MDR can also result from activation of detoxifying systems[90], and from defective apoptotic pathways[91-92]. The analysis of drug resistance is still a work in progress.

2.8 Model organisms used in this study

Drug susceptible MCF-7 cancer cells (control cells), have been widely used in laboratory studies since the line was established from the pleural effusion of a patient with metastatic mammary carcinoma in 1973[93]. The three drug resistant MCF-7 cell sublines studied here are stable lines derived from this parental cell line.

The VP-16 or etoposide resistant MCF-7 cancer cell line (MCF-7/VP) was developed in Dr. Ken Cowan's lab at NIH by selection during culture in increasing concentrations etoposide. It is reported to be 28-fold resistant to etoposide[94]. VP-16 is clinically used in the treatment of small cell lung cancer and testicular cancer. It can induce DNA strand breaks in the tumor cells[95-96].

MCF-7 cells resistant to mitoxantrone (MCF-7/MX) were also provided by Dr. Ken Cowan[98]. The line was isolated by serial passage of the parental control MCF-7 cells in stepwise increasing concentration of the anthracenedione mitoxantrone. These cells are about 4000-fold resistant to mitoxantrone[97].

The MCF-7 cell line selected for resistance to adriamycin in the present of verapamil (MCF-7/AdrVp) was provided by Dr. Douglas Ross from University of

Maryland Medical School[98]. It was isolated by selecting the control MCF-7 cells with

incremental increases of adriamycin in the presence of 10 microverapamil. It is 900-

fold resistant to Adr[99].
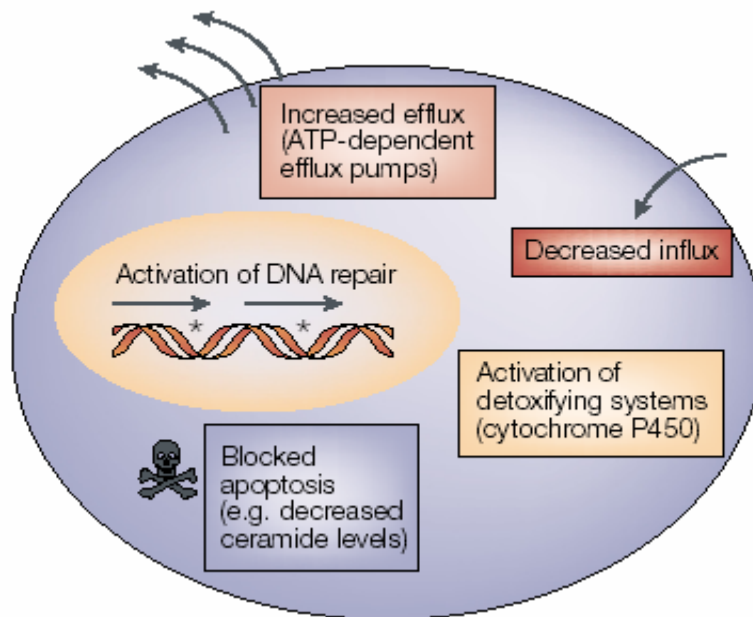
Figure 2-10 Mechanisms of drug resistance in cancer chemotherapy[86]

Chapter 3. Evaluation of Solution Phase Isoelectric Focusing as part of Proteomics Strategies

3.1 Introduction

Solution phase isoelectric focusing (solution IEF) has been introduced by several laboratories[64-70]as a protein prefractionation strategy before 2-D gel electrophoresis. In this technique, a protein mixture can be fractionated into a series of chambers under an electric field. These chambers are bounded by membranes with immobilized ampholytes. In this study, we have optimized the solution isoelectric focusing method for peptide separation and built it into a two-dimensional separation strategy. This chapter describes the evaluation of the performance of this solution IEF device. Chapter 4 presents the integration and application of the two-dimensional method for proteomic analysis.

3.2 Experimental

Materials

The protein isoelectric focusing cell, a Protean II Cell, model 1000/500 power supply, immobilized pH gradient (IPG) strips (17cm, pH 3-10), Protean II ready gels (8-16% Tris-HCl precast gel 110x80x1.0mm, IPG well), microbiospin P6 columns in Tris-HCl buffer and Biosafe Coomassie Stain were purchased from Bio-Rad (Hercules, CA). Nuclei Pure Prep Nuclei Isolation Kit, trypan blue, urea, thiourea, chaps, dithiothreitol (DTT), NH_4HCO_3, NaCl, iodoacetamide, TrismaBase, sodium dodecylsulfate (SDS), glycerol, L-lysine, L-arginine, phosphoric acid were from Sigma Co. (St. Louis, MO). Modified porcine trypsin (sequence grade) was

purchased from Promega (Madison, WI). ZipTip C_{18} was from Millipore (Billerca, MA). The MCE kit is from Proteome Systems (Woburn, MA).

Cell culture and isolation of nuclei and preparation of nuclear proteins

Nuclei pellets were provided by Dr. Zongming Fu in the Fenselau laboratory. Briefly, MCF-7 cancer cells were cultured in MEM (Sigma, St. Louis, MO) with 10% of FBS and 1% penicillin streptomycin. Every 6 months, the drug resistant cell lines were subjected to a reselection cycle of three passages with culture medium containing increased concentration of the appropriate drugs. Cultured MCF-7 cells were harvested at 95% confluence, released with trypsin, centrifuged at 500g, and washed twice with PBS. A nuclei isolation kit (Sigma) was used to isolate and purify MCF-7 nuclei, with slight modifications from the user instructions [98]. The nuclei were suspended in a NaCl buffer (20mM Hepes pH 7.9, 1.5mM $MgCl_2$, 0.5M NaCl, 0.2mM EDTA and 25% glycerol) with a ratio of 4ml solution every 1 gram of nuclei, vortexed thoroughly for 15 seconds every 10 minutes for a total of 40 minutes on ice and centrifuged at 16000g for 10 minutes. The supernatant was immediately transferred to new pre-chilled tubes and snap-frozen in aliquots with liquid nitrogen and stored at -80°C. The pellets were frozen in -80°C for further analysis.

Protein assay

Bio-Rad Protein Assay kit was used to determine protein concentration of the nuclear protein extracts. The Bio-Rad protein Assay is based on the method of Bradford, which is compatible with the chemicals used in protein extraction. This method is a dye-binding assay, in which the color of the dye changes differentially in response to various concentration of protein. The Coomassie Brilliant Blue G-250 dye

binds to primary basic and aromatic amino acid residues, especially arginine, and the absorbance maximum of the dye shift from 465nm to 595nm. The extinction coefficient of a dye-albumin complex solution was found to be constant over a 10-fold concentration range. Thus, Beer's law could be used to accurately measure protein quantity by selecting an appropriate ratio of dye volume to sample concentration. A series of 0, 1.5, 2.5, 5.0, 7.5 and 10μg/ml of bovine serum albumin (BSA) standard solutions were used to make a standard curve. The unknown sample was diluted properly. The standard and unknown solutions were mixed thoroughly with dye solution at a volume ratio of 4: 1. The absorbance was read at 595nm using a Beckman DU 530 life science UV/Vis spectrometer. The protein concentrations of unknown were calculated based on their absorbance and dilution factor[100].

Protein fractionation by solution isoelectric focusing using Amika device

A five-chamber separation device and an electrophoresis tank (Amika Corp. Columbia, MD) were used in the experiments. The device includes five Teflon dialysis chambers (500μl-volume each) connected in tandem plus two terminal Teflon caps (Amika Corp.). The adjacent separation chambers were divided by membranes with the desired pH values (ProteomeSystems, Woburn, MA), and two 3.5-kDa dialysis membranes (MFPI, Seguin, TX) were put at each end of the terminal chambers. Two O-rings (12 mm i.d., Scientific Instrument Services, Ringoes NJ) were used to seal each membrane between two chambers. A sample of MCF-7 nuclear protein (1.6mg) was solubilized in 1.5 ml of IPG rehydration buffer and loaded into the three separation chambers. The two terminal chambers were filled with electrode buffers, 7mM phosphoric acid (anode) and 20 mM lysine/20 mM

arginine (cathode). The assembled device was put into the electrophoresis tank and the two parts of the tank were filled with anode and cathode electrode buffers, respectively. A model 1000/500 power supply (BioRad, Hercules, CA) was used for isoelectrofocusing. Typically, 150 V was used for 1 h (initial ~2.5 mA, final ~0.4mA), 250 V for 1 h (initial ~ 0.7mA, final ~ 0.5 mA), followed by 600 V (initial ~ 1.4mA, final ~ 0.2 mA), and then 1000 V for overnight (final 0.17mA). After fractionation, solutions (~500μl each) were removed from the three chambers and the surfaces of the membranes and the inside walls of the chambers were washed with 200μl of rehydration buffer. The rinses were combined with the sample fractions. The membranes were removed and soaked in 250μl rehydration buffer for 1 h to extract proteins and combined with the appropriate sample fractions. In order to evaluate the separation efficiency, 25% of each fraction was separated on pH 3-10 2D-PAGE [102].

Sample fractionation by multi-compartment electrolyzer (MCE).

ProteomeSystems IsoelectrIQ^{TM} MCE was used in these experiments and all the materials were from Proteome systems (Woburn, MA). Five chambers (5ml each) were assembled according to the instructions and an MCP-7 nuclear extract (8mg) was dissolved in the MCE Sample Solubilizing Solution and loaded into the central chamber. Two chambers at each end were filled with MCE Electrode Solution and other chambers were filled with MCE Chamber Solution. A two-step isoelectrofocusing program was used, 100-1500 V slow ramp for 8 h, and then 1500 V for 8 h. After separation, each fraction was removed and the membranes and inside walls of the chambers were rinsed with chamber buffer. The membranes were extracted with 500μl rehydration buffer for 1 h. The rinses and extracts were

combined with the corresponding chamber fractions. Each fraction was concentrated with a centrifugal filter (5kDa, Millipore, MA) and separated on pH 3-10 2D-PAGE[102].

2-D gel electrophoresis

The protein sample was desalted with a Biospin-6 spin column (Bio-Rad, Hercules, CA), dried by Speed Vac, resuspended in rehydration buffer, which contained 7M urea, 2M thiourea, 2% chaps, 50mM DTT and 1% IPG buffer (Pharmacia, Piscataway, NJ), and incubated at room temperature for 1 hour. For the isoelectric focusing step, linear 11cm IPG strips pH 3-10 (Bio-Rad, CA) were rehydrated in sample solution for 12 hours, and IEF was performed in a Protean IEF Cell (Bio-Rad). When the focusing was completed, the IPG strips were equilibrated in buffers containing SDS for 14 minutes each buffer. The second-dimensional gels were Tris-HCl 8-16% IPG well gels (Bio-Rad). The strip was pushed into the gel well and sealed with agarose solution (5% agarose, 25mM Tris, 192mM glycine and 0.1% SDS). The gel was mounted in the Protean cell II and the reservoirs were filled with SDS running buffer. The electrophoresis was performed at 200 V for 60 minutes. Then the gels were removed and soaked in a fixed solution, which contains acetic acid/methanol/water (5:45:50, v/v/v) for 2 hours, followed by three washes with water. The gels were stained with Bio-Safe colloidal Coomassie Blue G-250 (Bio-Rad) for 2 hours and rinsed with water. The gel image was recorded with a GS-800 calibrated densitometer (Bio-Rad).

In-gel tryptic digestion

Spots of interest were excised from the gels. The excised gel particles were washed with water/acetonitrile 1:1(v/v) twice. After all liquid was removed, acetonitrile was added to cover the gel. Then the gel pieces were rehydrated in 0.1 M NH_4HCO_3 for 5 min, and an equal volume of acetonitrile was added and incubated for 15 min. All the liquid was removed and the gel pieces were dried in a vacuum centrifuge. The gel pieces were incubated with 10 mM dithiotreitol/0.1 M NH_4HCO_3 45min at 56°C and 55 mM iodoacetamide/ 0.1 M NH_4HCO_3 for 30min in the dark, respectively. After the iodoacetamide solution was removed, the gel particles were washed with 0.1 M NH_4HCO_3/acetonitrile and dried in a vacuum centrifuge. The gel particles were rehydrated in a 50 mM NH_4HCO_3 solution of 12.5ng/μl sequencing grade modified trypsin (Promega, Madison, WI) for 45min at 4°C. Then the enzyme supernatant solution was replaced by 50 mM NH_4HCO_3 buffer and the gel pieces were incubated at 37°C overnight. Peptides were extracted from the gel with 25mM NH_4HCO_3 and acetonitrile, followed by 5% formic acid and acetonitrile (1:1, v/v). All the extracts from same gel spot were pooled and dried in a Speed Vac.

The peptides were redissolved with 10 μl 0.1% aqueous TFA and desalted with ZipTip C_{18} pipette tips (Millipore, Bellerica, MA), according to the user instructions. The tip was prewet by aspirating a wetting solution (50% acetonitrile in water) and dispensing the waste, and then the tip was equilibrated with the equilibration solution (0.1% TFA in water). The peptides were bond to the tip by aspirating and dispensing the sample solution for 10 cycles. Next, the tip was washed with wash solution (0.1% TFA in water). Finally, 5μl elution buffer (0.1%TFA, 70% acetonitrile in water) was

aspirated to elute peptides from the tip. The eluted samples could be used directly for MALDI-TOF mass spectrometer analysis.

Mass spectrometry analysis and protein identification

An AXIMA-CFR MALDI-TOF (Kratos, Chestnut Ridge, NY) was used to acquire peptide mass spectra and Mascot search programs were used for protein identification. The instrument was operated in reflectron mode, analyzing positive ions. The laser power was set at 45-50 arbitrary units. The instrument was calibrated with melletin and angiotensin II. One microliter of sample solution was loaded on the plate, covered with 1μl matrix solution (50mM α-cyano-4-hydroxycinnamic acid in 0.1% TFA, 70% acetonitrile) and dried before it was put into the instrument. The spectrum was recorded by accumulating 100 laser shots and used to identify proteins, using the Mascot search program[103] against the SwissProt database (www.expasy.org).

3.3 Results and Discussion

The separation efficiency of isoelectric focusing was investigated by subjecting nuclear protein fractions recovered from each chamber to 2-D gel electrophoresis with Coomassie blue staining. Figure 3-3 is the gel array for the nuclear protein mixture before fractionation separated on a gel with a pH range of 3-10 for the first dimension separation. Figure 3-4 presents the 2-D gel arrays of the material recovered from the three chambers in the Amika isoelectric focusing device. When the set of gels from the fractionation experiment is compared with Figure 3-3, it is apparent that the nuclear extracts were separated into three pools. It should be noted that there are some overlapping spots found in fraction 3-5 and 5-8. Most of these spots represent

high abundance proteins in the nuclear fraction, such as actin, prohibitin. Because of the high abundance, these proteins tend to precipitate easily during isoelectric focusing and stick to the membrane and chamber inner surface. The sample was loaded into both fraction 3-5 and 5-8 chambers before separation. This could cause the overlap. It also appears that protein is most poorly recovered from the most basic chamber (Table 3-1). This is not yet understood. However, the total number of spots in the 2-D gel without prefractionation was 157, compared with 197 spots in the combined set of fractionation gels. The successful fractionation by the device qualifies isoelectric focusing as a potential method for analysis of complex eukaryotic proteomes.

A preparative isoelectric focusing instrument, the multi-compartment electrolyzer (MCE) (Proteome Systems), was also used to separate the nuclear protein and Figure 3-5 shows the 2-D array of each fraction recovered from this instrument. It can be seen that this instrument provides highly effective fractionation for protein mixtures. However, MCE requires large sample volumes (5-50mg proteins) and results in large dilute fractions that need to be concentrated. Only 12% recovery was obtained from the MCE experiment. The large sample loss was caused by sample processing before and after separation. Considering the huge sample requirement and substantial loss, MCE is not suitable for our experiment.

In initial tests of solution isoelectric focusing, gel membranes cast in our lab were used and separation was obtained (data not shown). However, the membranes were fragile and could not survive higher voltage and longer running times. These resulted in protein overlap and poor reproducibility. The robust membranes used in the MCE

apparatus have high mechanical strength and perform better in separation experiments.

The total protein recovery of the smaller device was also explored. Table 3-1 summarizes protein yield from each of the chambers and the overall recovery. It appears that a significant amount of protein dissolved into the membranes or precipitated out and was not recovered.

It was also demonstrated that isoelectric focusing can be part of a proteomics strategy. In these experiments, several spots were excised from the gels shown in Figure 3-4 and analyzed by peptide mass mapping. Protein identifications are summarized in Table 3-2 and peptide MALDI mass spectra for the first and last entries are shown in Figure 3-6, 3-7. In these spectra, the masses labeled on the peaks were used in the database search. The peak at m/z 842 is a trypsin autolysis peptide, which was used as an internal calibration for each spectrum.

3.4 Conclusions

After evaluating isoelectric focusing fractionation of human nuclear proteins, we propose that solution isoelectric focusing can make valuable contributions to the fractionation of protein mixtures and can be involved in strategies designed to identify proteins by mass spectrometry. The new membranes from ProteomeSystems have been proven to be sturdy, durable and provide satisfactory fractionation.

Figure 3-1 Photographs of the solution isoelectric focusing device used in our lab. Left: the electrophoresis tank (Amika Corp.), Teflon chambers and caps (multichamber Teflon dialyzer system, Amika Corp.) and 12-mm-i.d. O-rings (Scientific Instrument Services, Inc., Ringoes, NJ); Right: pH membranes 3-5-6.5-8-11 (ProteomeSystems, Inc., Woburn, MA)
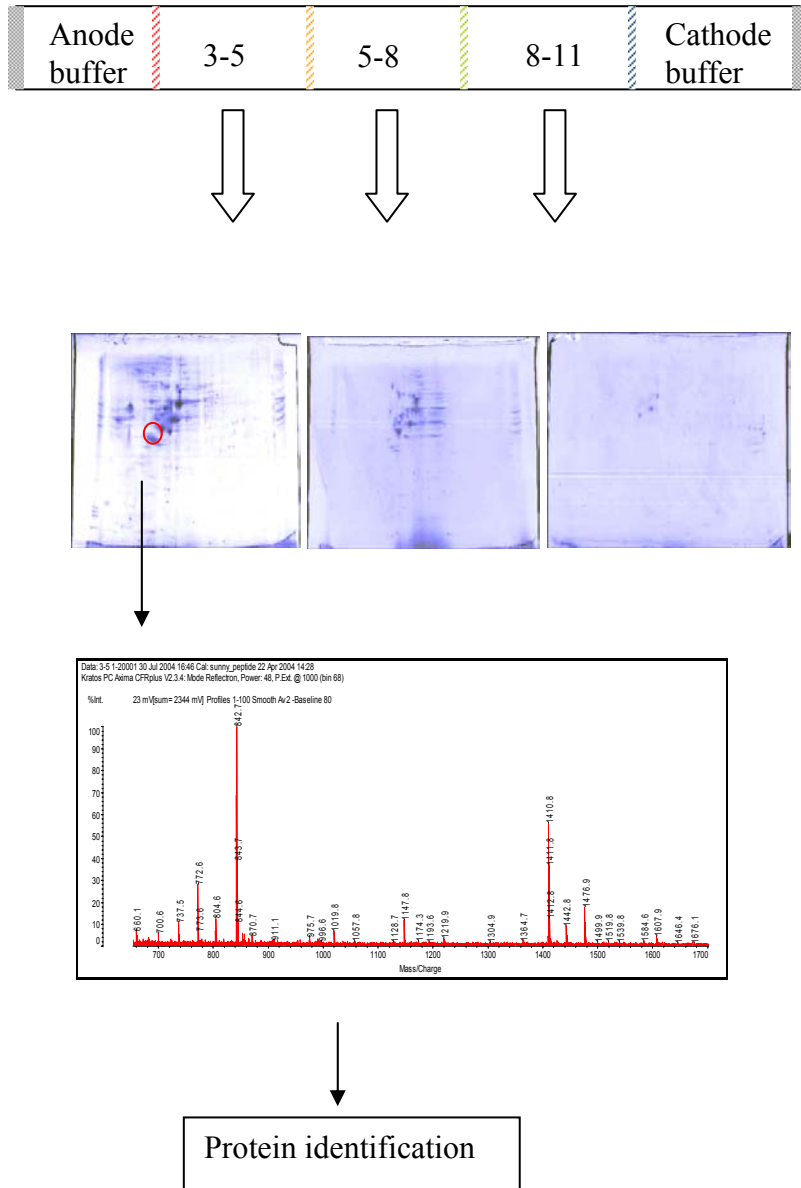
Figure 3-2 Schematic illustration of the experiment procedure.
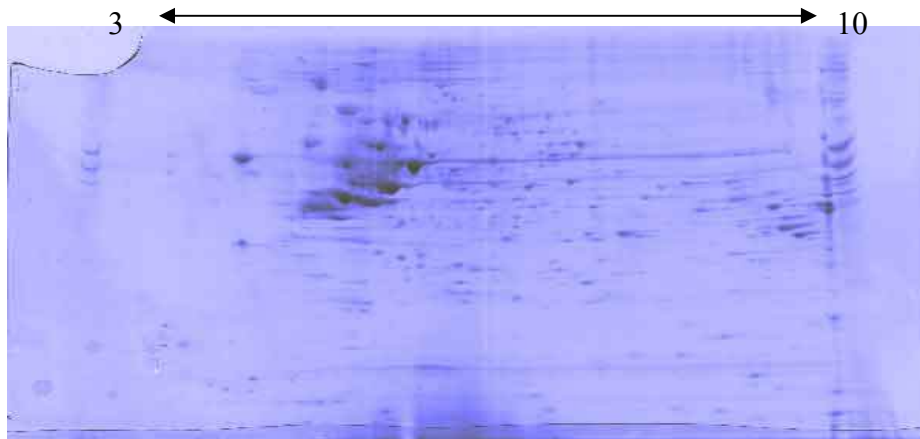
Figure 3-3 2-D gel array of the nuclear protein mixture without solution IEF fractionation.
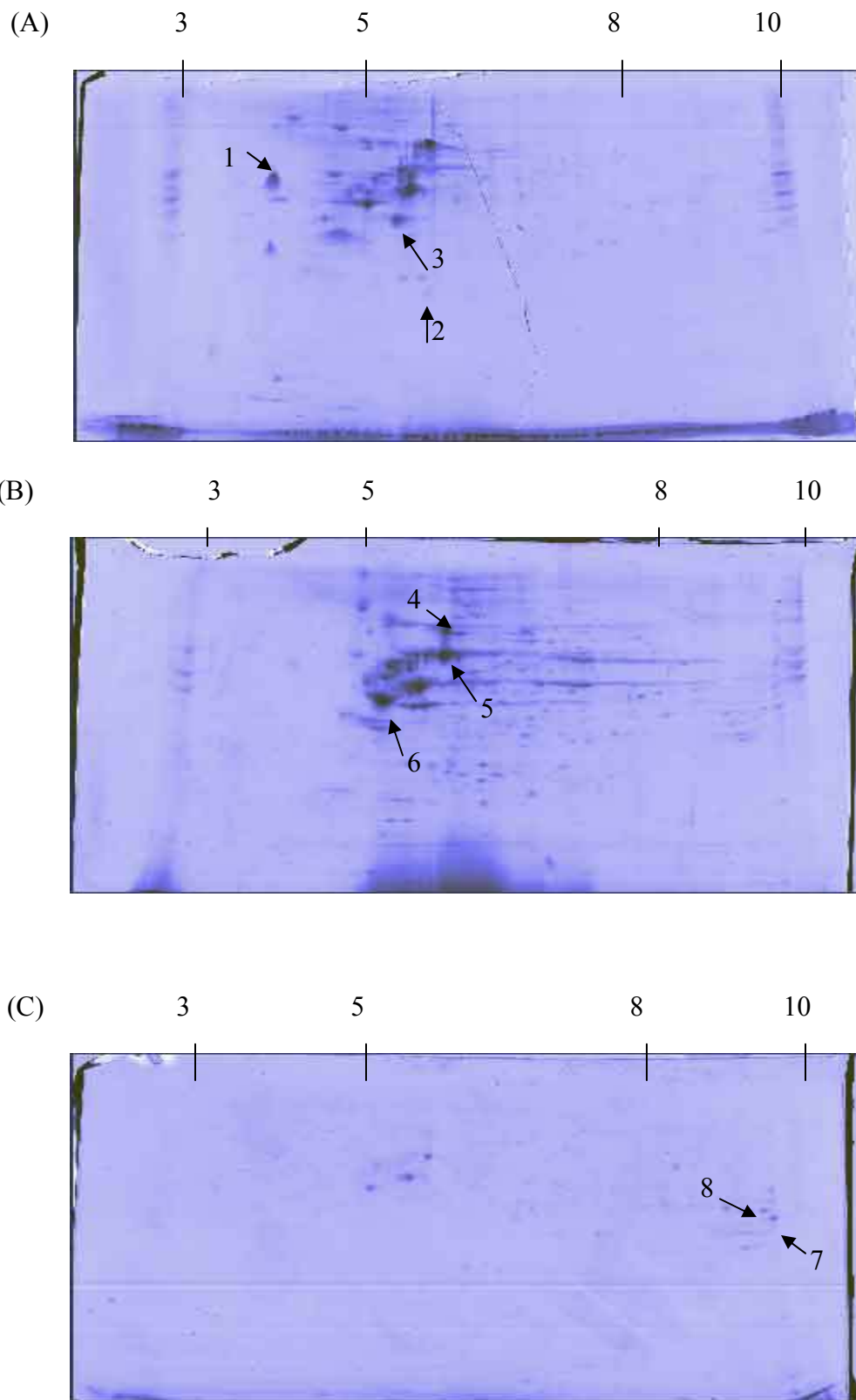
Figure 3-4 2-D gel arrays of nuclear protein mixture from three chambers after fractionation. (A) fraction pH 3-5, (B) fraction pH 5-8, (C) fraction pH 8-11
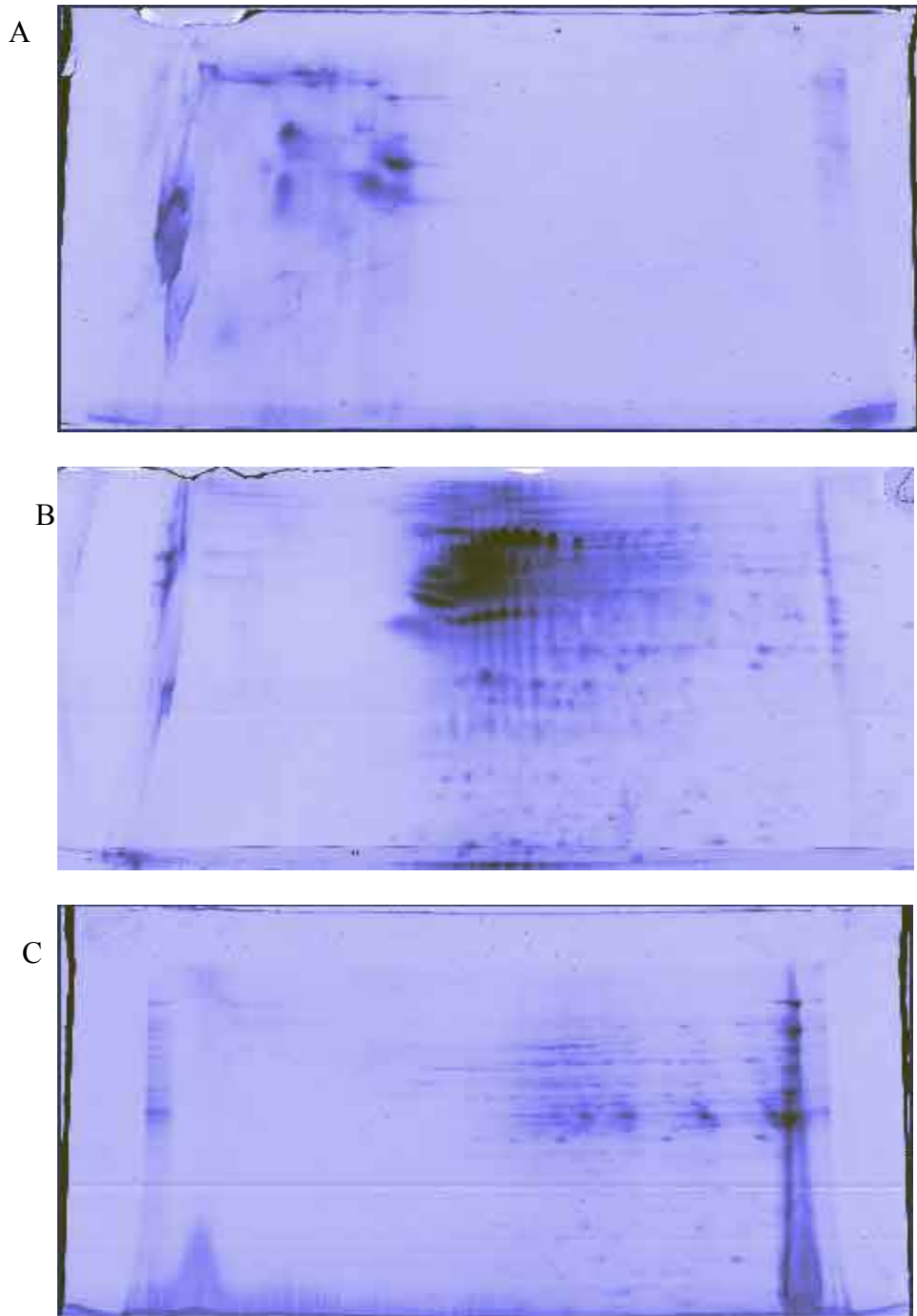
Figure 3-5 2-D gel arrays of nuclear protein mixture fractionated by MCE. (A) fraction pH 3-5; (B) fraction pH 5-8; (C) fraction pH 8-11.

Table 3-1 Protein recoveries of the solution IEF device in the lab

Device	Initial loading	Fraction pH3-5	Fraction pH5-8	Fraction pH8-11	Sample recovery
Amika	1.6mg/1.5ml	304 μg	302 μg	254 μg	54%
	2.0mg/1.5ml	540 ug	620 ug	14 ug	59%
	2.0mg/1.5ml	197μg	635μg	362μg	60%

Table 3-2 Proteins identified from spots excited from the gels in Figure 3-4

Spot NO	Protein ID	pI	MW	Accession NO	Sequence coverage
1	Calreticulin	4.29	48141 Da	P27797	24%
2	Prohibitin	5.57	29804 Da	P35232	37%
3	Actin Beta	5.15	42408 Da	P60709	22%
4	Stress-70 protein	5.87	73625Da	P38646	27%
5	HSP 60	5.70	61016Da	P10809	34%
6	Cytokeratin 19	5.04	44079Da	P08727	62%
7	hnRNP A1	9.27	38715Da	Po9651	41%
8	hnRNP A2/B1	8.97	37430Da	P22626	35%
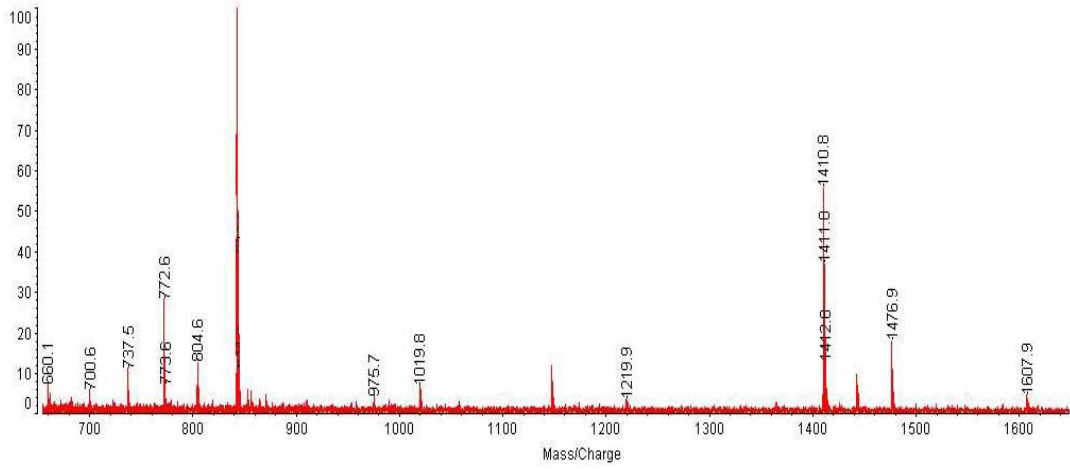
Figure 3-6 Peptide mass map of the material in spot 1 in Figure 3-4 (A)
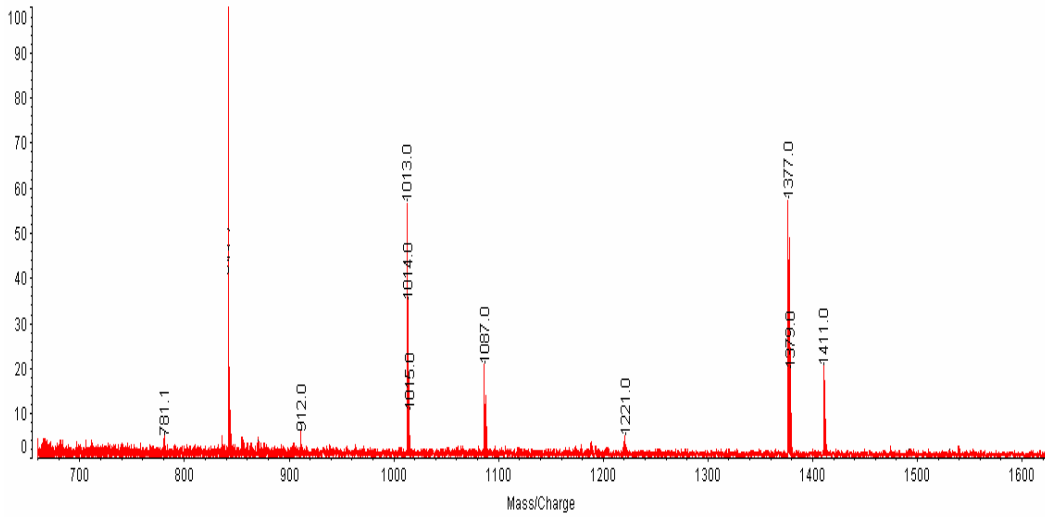
Figure 3-7 Peptide mass map of the material in spot 8 in figure 3-4 (C).

Chapter 4 Introduction of solution isoelectric focusing in high throughput strategies to quantitate nuclear proteins from MCF-7 cancer cells

A two-step labeling protocol developed in our lab introduces ^{18}O isotope labels into peptides, One reason to develop electric focusing as a fractionation method for peptides is to optimize the implementation of ^{18}O labeling in proteomic strategies. It is important to minimize protein manipulation before the introduction of labels so as to minimize differential loss. Fractionation of peptides after they are labeled will increase the accuracy of the quantitative measurement.
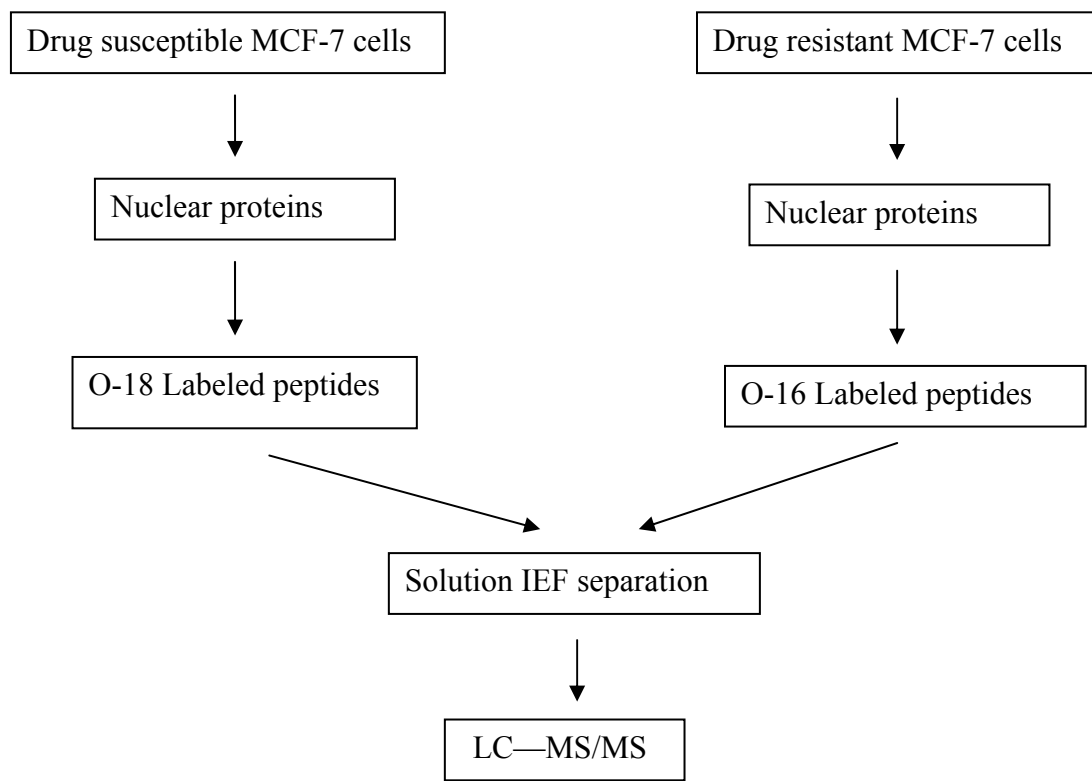
Figure 4-1 Scheme of the comparative proteomic strategy

4.1 Experimental:

Materials:

L-lysine, L-arginine, and phosphoric acid were obtained from Sigma Co. (St. Louise, MO). Modified porcine trypsin (sequence grade) was purchased from Promega (Madison, WI). Dialysis membrane (500 Da cut-off) came from Millipore (Billerca, MA). MCE kit is from Proteome Systems (Woburn, MA). Isotopically enriched $H_2{}^{18}O$, > 95% ^{18}O was purchased from Isotech, Inc. (Miamisburg, OH). Trypsin immobilized on Poros beads was purchased from Applied Biosystems (Foster City, CA). PepClean^{TM} C-18 spin columns came from Pierce (Rockford, IL). Water was purified by a MilliQ system and filtered with 0.22 μm membrane (Millipore, MA).

Nuclear protein extraction and in-solution digestion

The nuclei pellets were obtained from Dr. Zongming Fu in the Fenselau lab. The nuclei isolation procedure was described in Chapter 3. Briefly, the MCF-7 cancer cells were cultured and harvested in the lab. The nuclei were isolated and purified with a nuclei isolation kit (Sigma). The soluble nuclear protein fraction was extracted with a NaCl buffer and analyzed by Dr. Zongming Fu[98]. The insoluble nuclear protein fraction was analyzed in this experiment. The pellets recovered after NaCl buffer extraction were resuspended in the Sample Solubilizing Buffer (ProteomeSystems, Woburn, MA) with the detergent excluded, and held for 30 minutes on ice. This suspension was vortexed vigorously for 60 seconds every 10 minutes. Then the suspension was centrifuged at 16000 g for 10 minutes. The supernatant fraction was immediately transferred to new tubes and stored at -80°C. Protein concentration was

determined with the Bradford protein assay mentioned before. The proteins were reduced with tributylphosphine (TBP) (1:40, v/v) and alkylated with Acrylamide Alkylation Reagent (1:100, v/v) (ProteomeSystems) for 90 minutes at room temperature. Then the solution was diluted 10 times with 50mM NH_4HCO_3 pH 8.0. Proteins were digested with trypsin (Promega) (1:50, w/w) at 37°C overnight.

Proteolytic $H_2^{18}O$ labeling

In the comparative proteomic experiment, only one of the peptide pools is labeled with ^{18}O, but the procedure is applied to both pools in order to ensure experimental homology. Immobilized trypsin was washed with water and added into peptide solutions in a ratio of 1:5 (v:v) . The peptide and immobilized trypsin mixtures were completely dried in a vacuum concentrator. Then one residue was redissolved in 80% $H_2^{18}O$ and 20% acetonitrile, and the other was redissolved in 80% $H_2^{16}O$ and 20% acetonitrile. The solutions were rotated at room temperature for approximately 5 hours on a bench-top rotator. Then the labeled and unlabeled peptide pools were mixed and analyzed.

Peptide separation with solution isoelectric focusing

The same Amika device was used for peptide separation with some changes in its configuration. An additional membrane (pH 6.5) was inserted so that the peptides were divided into 6 fractions (4 separation chambers plus the two terminal chambers). Two 500 Da dialysis membranes were put at each end of the terminal chambers to protect the system from the running buffer outside. The sample was loaded in the chamber with pH 5-6.5. The two terminal chambers were filled with Electrode Buffer and the chambers without sample were filled with Chamber Buffer both of which

were purchased from ProteomeSystem. A shorter program was used than for protein separation: 100 V for 10 min, 200 V for 10 min, and 500 V for 20 min and 1000 V for 100 min. After fractionation, solutions were collected and the inner wall and membranes were rinsed with 200μl chamber buffer. The rinses were combined with the sample fractions.

Peptide separation with conventional reversed-phase liquid chromatography

A Shimazu LC system (Columbia, MD) was used to separate peptides after fractionation by isoelectric focusing. The column was purchased from Phenomenex (Torrance, CA) and packed with C18 particles. An 80 min elution gradient was used: 95% solvent A (0.1% TFA in water) for 5 min, then solvent B (0.1% TFA in acetonitrile) from 5% to 60% in 55min, 60% to 90% for 10 min, hold for 5 min, and 90% A for 10 min. The effluent was monitored by a UV detector and recorded in a computer.

Peptide desalting

Each of the six IEF peptide fractions was desalted with a PepClean C-18 spin column (Pierce) according to the user guide. Each column was wetted with 200 μl activation solution (50% acetonitrile) twice and equilibrated with 200 μl equilibration solution (0.5% TFA in 5% acetonitrile) twice. The peptide fractions were mixed with sample buffer (3:1 v/v, 2%TFA in 20%acetonitrile) and loaded onto the columns. The columns were washed with 200 μl wash solution (0.5%TFA in 5%acetonitrile) three times. The peptides were eluted with 60μl elution buffer (70% acetonitrile). The effluent was dried by vacuum centrifugation and stored at -80°C before LC-MS/MS analysis.

MALDI-TOF and ESI-QqTOF analysis of ^{18}O labeling

Lysozyme was used as a standard protein to exam the ^{18}O labeling efficiency. The resultant peptides were desalted and loaded on a MALDI sample plate. An AXIMA-CFR MALDI-TOF (Kratos Analytical, Chestnut Ridge, NY) was used to acquire peptide mass spectra using the same condition described in the last chapter.

Next a standard protein mixture (bovine serum albumin, chicken egg albumin, myoglubin, lysozyme, ribonuclease A and cytochrome c) was digested and split into two portions. One was labeled with regular water and the other was labeled with $H_2{}^{18}O$ and they were mixed at a ratio of 2:1. The mixture was fractionated by solution IEF and the fraction from the pH 5-6.5 chamber was analyzed LC-MS/MS. An API QSTAR Pulsar Qq-TOF (Applied Biosystems, Foster City, CA) instrument was used to acquire data. The mass spectrometer was operated in a data-dependent mode in which a full MS scan was followed by three MS/MS scans of the most intensive ions automatically selected for collision-induced dissociation (CID).

LC-MS/MS analysis

Capillary reverse-phase liquid chromatography was performed using an LCQ DecaXP (ThermoFinnigan, San Jose, CA) instrument. The reverse-phase column was a 75μm I.D. x 10 cm fused capillary with a 15 μm nanoelectrospray tip and packed with 5μm, 300 Å BioBasic C18 particles (New Objective Inc., Woburn, MA). After injecting 10μl of sample, the peptides were eluted using a linear gradient of 5% solution B (0.1% formic acid in acetonitrile, v/v) to 60% solution B in 55 min. The column was then washed with 98% solution B for 5 min and 100% solution A (0.1% formic acid in water, v/v) for 26 min at a constant flow rate of 275 nl/min. The ion-

trap mass spectrometer was operated in a data-dependent mode in which a full MS scan was followed by three MS/MS scans of the most intensive ions automatically selected for collision-induced dissociation (CID).

Bioinformatics analyses

The MS/MS spectra obtained from LC-MS/MS were searched against the NCBI human database (www.pubmed.org) using SEQUEST[104] (ThermoFinnigan, CA). The peptides concerned to be identified must achieve the criteria: cross-correlation (Xcorr) were higher than 1.9 for singly charged ions, 2.2 for doubly charged ions, 2.5 for triply charged ions, and the delta correlation (ΔCn) were higher than 0.1.

| + | Anode buffer | 3-5 | 5-6.5 | 6.5-8 | 8-11 | Cathode buffer | − |

Desalting with C18 spin column

Separation LC-MS/MS

Protein identification
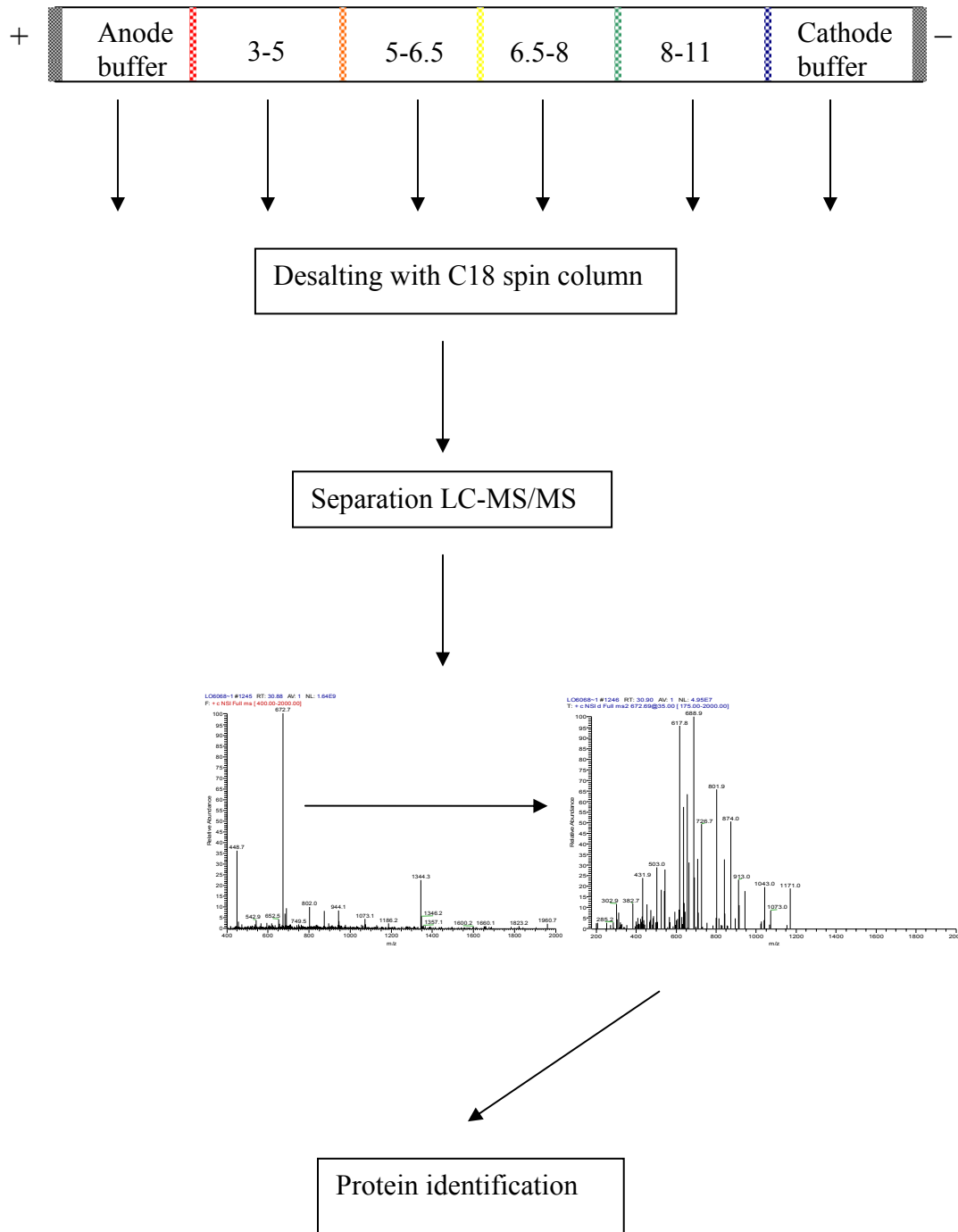
Figure 4-2 Scheme of the two-dimensional solution IEF-LC-MS/MS proteomic strategy

4.2 Results and discussion

4.2.1 Peptide fractionation by solution isoelectric focusing using Amika device

Shotgun proteomics strategies, based on peptide separation, are increasingly used today. Two-dimensional liquid chromatography (for example with SCX and RPLC as the first and second dimension respectively) provides high-throughput analysis of complex peptide samples. Reversed-phase liquid chromatography is always used as the second dimension because of its high resolution capability and compatibility with mass spectrometry. But ion-exchange chromatography as the first dimension has limited separation capacity which results in a lot of overlapping between fractions. Solution isoelectric focusing has been proven to successfully fractionate proteins by several groups including us. In this experiment solution IEF was applied to peptide separation using the small volume device (Amika) with a little modification.

Because each protein is digested into several peptides, the peptide mixture is much more complex than the initial protein mixture. Consequently, in the peptide separation, a total of 6 fractions were collected.

Detergents in solutions can improve protein solubilization, but most of them impede mass spectrometry and stick on RPLC column. To optimize the subsequent HPLC separation and mass spectrometry analysis, a chamber buffer without any detergent was used. No sample precipitation during IEF was observed. IPG buffer was added to chamber buffer to increase the conductivity and to help maintain the pH gradient across the device. Because the initial concentration of the IPG buffer (0.5%,

v/v) was very low and dilution occurred during the experiment, the ampholytes in the IPG buffer were eluted earlier and did not show any interference with peptides.

A shorter running program, which took about 140 min, was used for peptide separation. Initially, the program used for protein separation was used. It resulted in no significant separation of the peptide mixture (data not shown). This is because of the difference in diffusion behavior between proteins and peptides. The separation in IEF is driven by the electric field and neutralized by molecular diffusion. The velocity of diffusion and electrical mobility are both inversely proportional to the molecular weight. Because the proteins are much heavier than the peptides (more than 20 times higher on average), the movement of protein during separation is much slower than that of peptides and it takes much longer for proteins to focus and diffuse than peptides. A long running time helps proteins to separate but hurts peptide separation. Before focusing is completed, the electric force is dominant. But after the isoelectric focusing finished, molecules are driven by diffusion. So after peptides were focused rapidly, diffusion occurs during the rest of the time. Peptide fractionation improved greatly when a short running time was used. Figure 4-3 shows the chromatogram of each fraction. Each fraction was eluted through the same column with the same solvent profile. The variation in the chromatograms indicates the varying composition of each fraction and indicates that peptides were successfully fractionated during solution isoelectric focusing.

Peptide identification using tandem mass spectrometry

The mass spectrometer was programmed to select the three most intensive peptides from the full MS scan for tandem MS analysis. The MS/MS spectra were searched

against a human protein database. The peptides, which passed the criteria mentioned in the Experimental Section, were considered to be identified. Figure 4-4 shows a sample MS/MS spectrum of a doubly charged peptide. The peptide was eluted at 35.50 min in fraction 2 and the database search proposed the sequence MIAGQVLDINLAAEP.

All peptide pI values were calculated using the pI/MW calculator in www.expasy.org and Table 4-1 lists the sequence, pI, Xcorr evaluation, and fraction number for all identified peptides.

Figure 4-5 shows the total number of unique peptides identified in each fraction. Fraction 3, ranging from pH 5 to pH 6.5, has the highest number of peptides identified followed by fraction 2 pH 3-5. The sum of the numbers in the Figure 4-5 is higher than the total number of peptides identified, which means there are peptides re-identified between fractions. Figure 4-6 shows the number of chambers in which a peptide was observed. About 80% of the total peptides appeared in only one fraction. The other 20% appeared in adjacent chambers with close pH ranges. This indicates that most of the peptides were separated according to their pI values.

Figure 4-7 shows the pI range of peptides in each fraction. The discontinue lines indicate the pH range of each fraction defined by the membranes, fraction 2 and 5 fit exactly in the range, most of the peptides in fraction 3 were also within the range. The pI values of peptides in fractions 4 and 6 were lower than the defined ranges, while the peptides in fraction 1 were more basic. The pI overlaps between each fraction were minimal except for the last two fractions, which mean the peptides were fractionated according to their pI. The overlap between fraction 5 and 6 was probably

caused by intrinsic properties of peptides. The proteins were digested into peptides by trypsin which cut at the C-termini of lysine or arginine. Lysine and arginine are two basic amino acids with pI of 9.74 and 10.76, respectively. A peptide with pI higher than pH 11 should have multiple lysines and arginines. Most of the peptides are fully tryptic peptides with one lysine or arginine at the C-termini. The peptides with extreme basic pI are rare. The pH of the cathode buffer was close to 11. The pH difference between fraction 5 and 6 is small. Several peptides occupied both fractions. The IPG buffer added is another factor that affects the pH gradient in the device. The components in the IPG buffer are ampholytes from pH 3 to 10. These ampholytes may affect the pH range of each chamber. Finally, the pI distribution of the peptides themselves may also play a role in the separation. It is obvious that there is few peptides appeared in the range of pH 7 to 8. Bundy and colleagues also observed this phenomenon as well, when they studied yeast peptide pI distribution using IPG strip[53].

The reproducibility of solution isoelectric focusing was also investigated. Three nuclear peptide mixtures were fractionated with solution IEF separately and the same fractions from different runs were characterized by HPLC separation. Figure 4-8 shows three chromatograms of fraction 2 obtained using a UV detector. Most of the peaks appeared in all of the three runs. The similarity of the elution profiles indicates the components of the same fractions from the three different solution IEF separations were same.

After isoelectric focusing, each fraction was desalted, and analyzed by LC-MS/MS. Each fraction was analyzed three times. It is interesting that the peptides identified in

the first trial didn't necessary appear in the following repeat runs. Meanwhile, new

peptides, which were not identified in the first run, were observed in the second or

third trial. Figure 4-8 summarized this result. The blue bars indicate the number of

peptides identified in each fraction during the first LC-MS/MS analysis; the purple

bars are the sum of the peptides in the first and second LC-MS/MS analyses; and the

yellow bars are the total peptides identified in each fraction after three LC-MS/MS

analyses. More peptides could be identified when the same sample was analyzed

several times by LC-MS/MS. This observation agrees with a recent report of Yates[105].

The mass spectrometer was set to select the three most abundant ions in each full

mass scan for analysis. Unlike a UV detector, the mass spectrometer can not monitor

the LC effluent continuously although the scan time is very short. So each MS scan

only catches the components eluted at a certain time. As for the LC separation, each

analysis would not be exactly the same. The MS scan may catch new ions which are

lost in previous trial and lose another part of ions. Table 4-2 shows that the number of

peptides identified increased with increasing injection times[105].
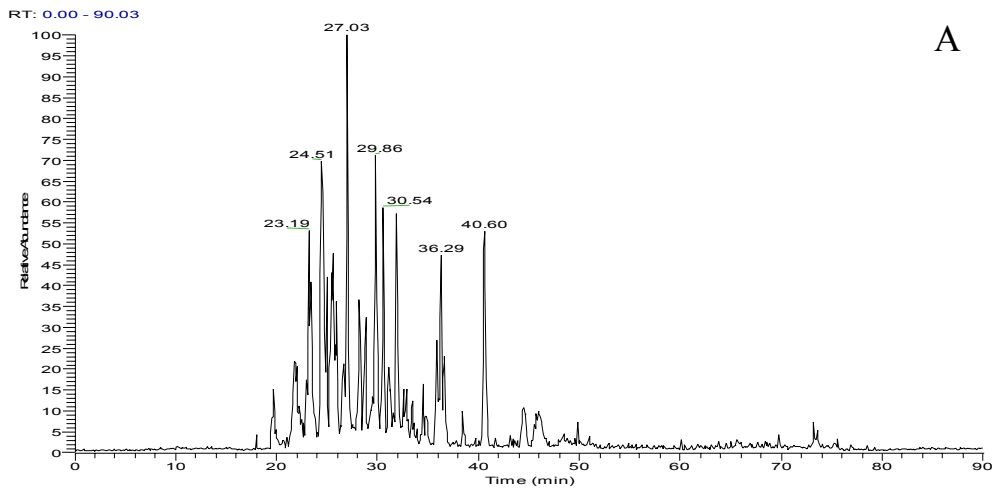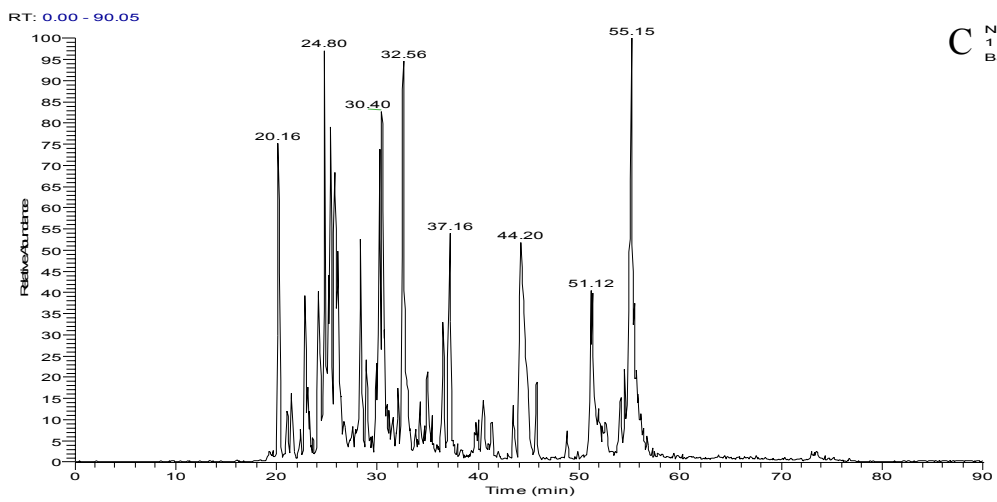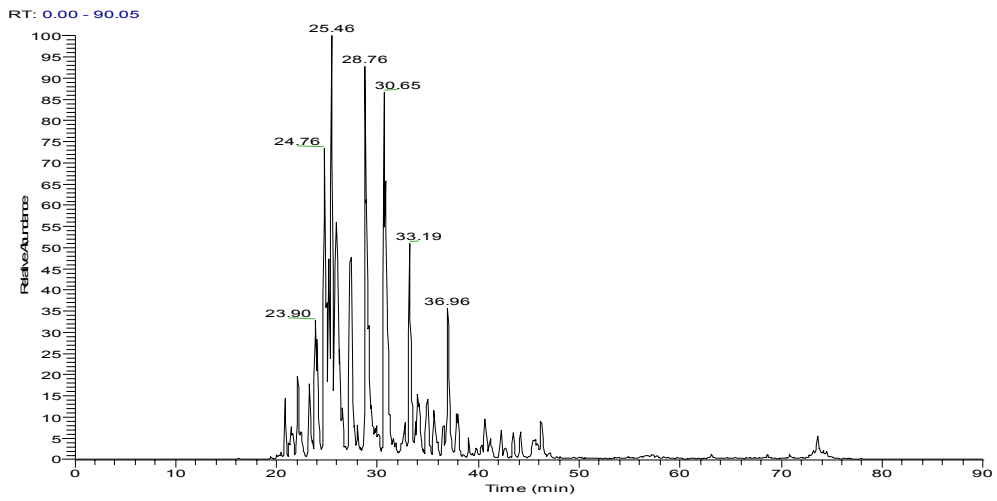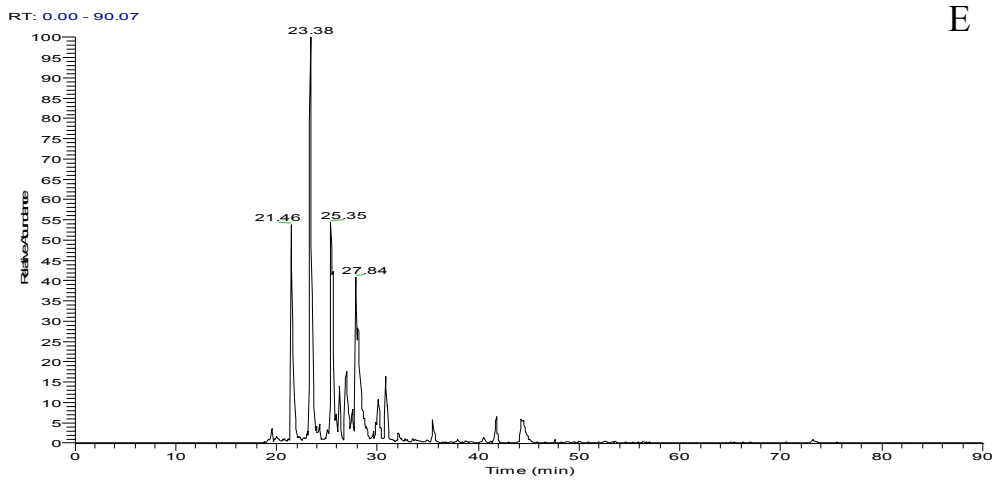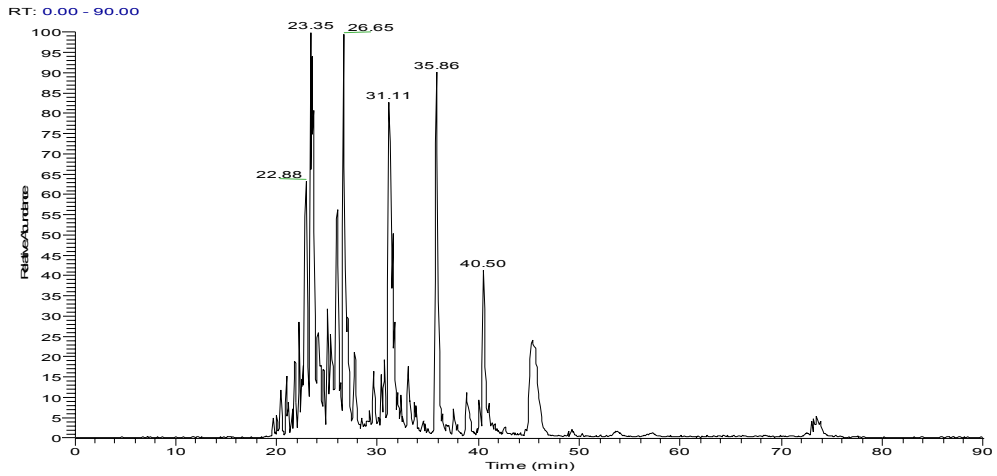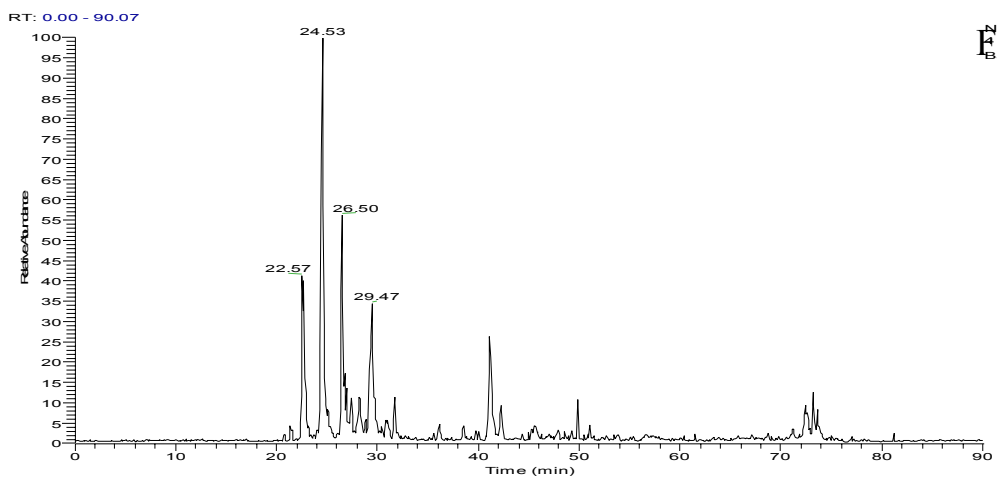
Figure 4-3 Chromatographs of nuclear peptides from each solution IEF fraction. A:
pH < 3; B: pH 3-5; C: pH 5-6.5; D: pH 6.5-8; E: pH 8-11; F: pH >11
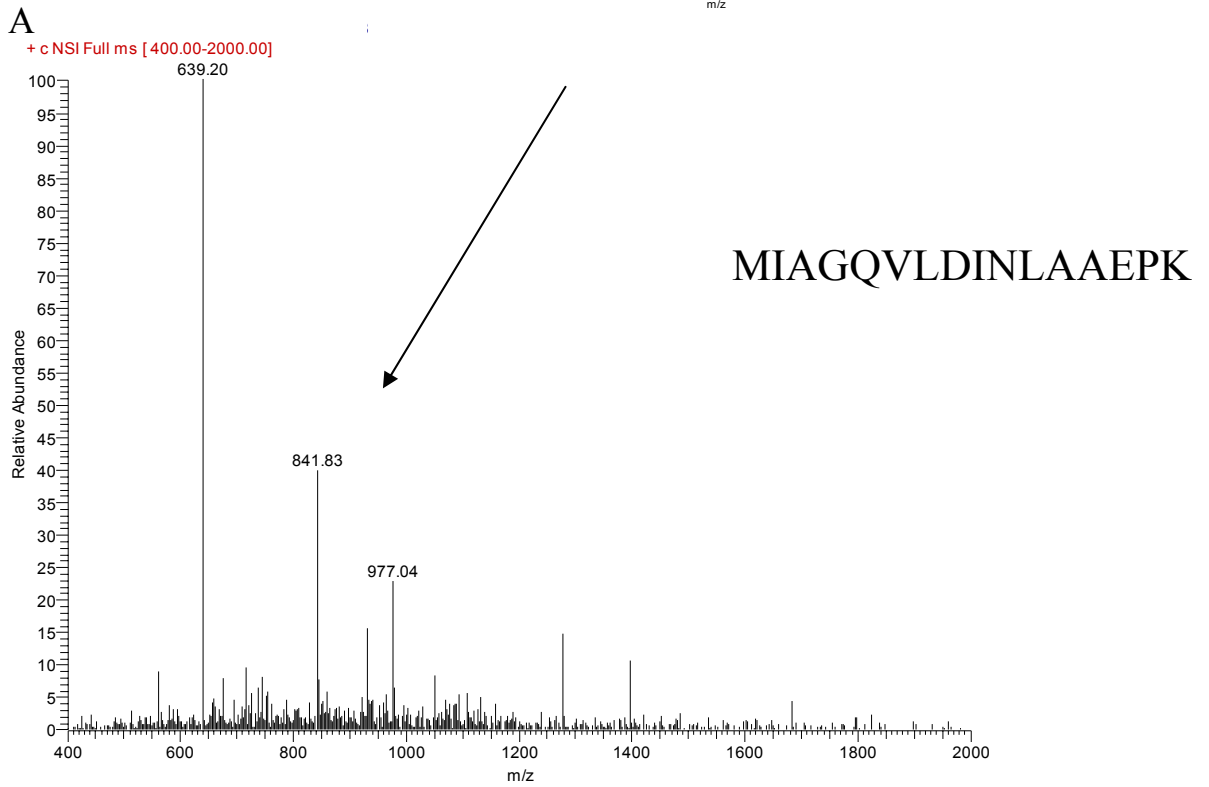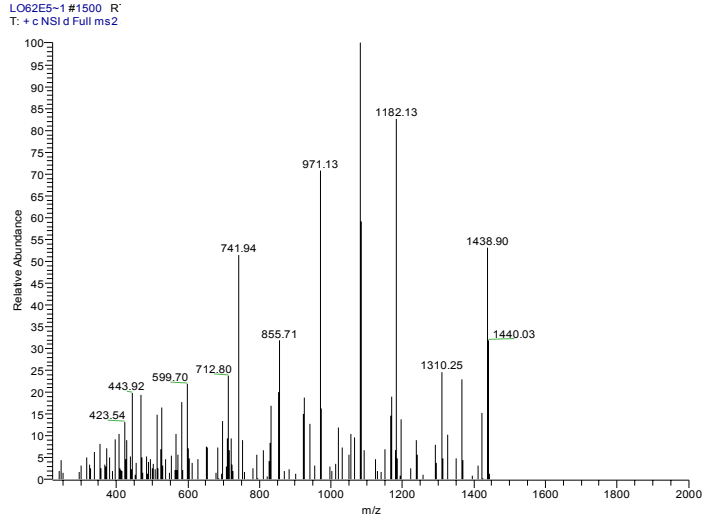
Figure 4-4 Mass spectra of a peptide in fraction 2. A: MS spectrum of the effluent at 35.43 min, the three most intensive ions are labeled; B: MS/MS spectrum of the ion with m/z at 841.83.

Table 4-1 List of nuclear peptides identified by the two-dimensional strategy

sequence	charge	pI	Xcorr	fraction
MESEGGADDSAEEGDLLDDDNEDRGD	3	3.42	3.34	1
GDAEKPEEELEEDDDEELDETLSER	3	3.59	3.81	1
LLDPEDVDVPQPDEK	2	3.66	3.51	1
LLEDGEDFNLGDALDSSNSMQTIQK	2,3	3.66	6.07	1,2
EEDGSLSLDGADSTGVVAK	2	3.77	3.3	1
LPPNTNDEVDEDPTGNK	2	3.77	3.59	1
LEDLLQDAQDEK	2	3.77	2.38	1
FMQDPMEVFVDDETK	2	3.77	3.43	1
ELDALDANDELTPLGR	2	3.77	2.84	1
GSSEQAESDNMDVPPEDDSKEGAGEQK	3	3.8	2.51	1
SANAEDAQEFSDVER	2	3.83	3.58	1
QSGEAFVELGSEDDVK	2	3.83	2.93	1
VDIEGPDVNIEGPEGK	2	3.83	2.86	1
GVEEEEDGEMRE	2	3.88	3.73	1
EDLPAENGETKTEESPASDEAGEK	3	3.89	3.54	1
PMEELSEEDR	2	3.91	2.2	1,2
GDEELDSLIK	2	3.92	3.22	1,6
GFGFDFNSEEDAK	2	3.92	3.32	2
LDLDLTADSQPPVFK	2	3.93	3	1
TVLDPVTGDLSDTR	2	3.93	2.71	1
DDGTGQLLLPLSDAR	2	3.93	2.77	1,6
S@EFLLTDYFEEDPNSAMDKER	3	3.95	2.7	2
ELEQIFCQFDS@KLEAADEGSGDVKY	3	3.95	2.84	3
LEAELGNMQGLVEDFK	2	4	5.1	1,2,3
AALEDTLAETEAR	1,2	4	4.09	1,2,6
DSALETLQGQLEEK	2	4	3.2	1
GLSEDTTEETLK	2	4	2.86	2
IDTIEIITDR	2	4.03	2.35	1
SLDMDSIIAEVK	2	4.03	4.08	1,2,3
TVQSLEIDLDSMR	2,3	4.03	4.79	1,2,3
GQVGGQVSVEVDSAPGTDLAK	2	4.03	2.35	2
EGQGEGETQEAAAATAAAR	2	4.09	2.92	2
RLLEDGEDFNLGDALGSSNSMQTIQK	3	4.11	3.23	1
DIQVGAQDGVLESGVMLGDREAVR	3	4.11	2.62	3
VDSLLENLEK	2	4.14	2.93	2
CPALEELDLTACR	2	4.14	2.32	1
ASLEAAIADAEQR	2,3	4.14	4.79	1,2,3

LEGLTDEINFLR	2	4.14	4.12	1,2,3,6
TGAIDVPVGEELLGR	2	4.15	3.99	2
IEVIEIMTDR	2	4.14	3.15	2
ILSISADIETIGEILK	2	4.14	3.42	2
NTSEQDQPMGGWEMIR	2	4.14	2.61	2
EILSARLT@QALDNYEGFSLA	3	4.14	2.94	4
LSAAVEVGDAAEVK	2	4.14	2.55	4
VLVLEMFSGGDAAALER	3	4.14	2.9	5
LQEKEDLQELNDR	2	4.18	4.24	2
VVDALGNAIDGK	2	4.21	3.27	2
VIDDTNITR	2	4.21	2.66	1,2
IVDDCGGAFTMGVIGGGVFQAIKGF	3	4.21	2.5	2
LLDAQLATGGIVDPR	2	4.21	2.8	2
PVGGLS@SSDTMDYR	2	4.21	2.32	6
WKDSDEADLVLAK	2	4.23	3.74	2
TEMENEFVLIK	2	4.25	4.1	1,2
ALEAANGELEVK	2	4.25	4.41	1,2,3
SENGLEFTSSGSANTETTK	2	4.25	4.3	1,2
MEEESGAPGVPSGNGAPGPK	2	4.25	3.21	1,2
ITESEEVVSR	2	4.25	3.12	2
LFIGGLSFETTEESLR	2	4.25	3.62	2,3
TGTAEMSSILEER	2	4.25	3.8	2
STGEAFVQFASQEIAEK	2	4.25	3.74	2
IFVGGLNPEATEEK	2	4.25	2.97	2
VELAEICAKS@ERYIGTEGGGMDQS	3	4.25	2.83	2
KLNFNGEGEPEELMVDNWR	3	4.25	3.22	3
VDNDENEHQLSLR	2	4.31	2.69	2
FDIEMSMRGDIFER	2	4.32	2.53	2
AAVAGEDGRMIAGQVLDINLAAEPK	3	4.32	2.63	3
MSGWADERGGEGDGR	3	4.32	2.61	6
ADGAAKEGAGAAAAAGPDGAPEAR	3	4.32	2.71	4
LQLETEIEALKEELLFMK	2,3	4.33	5.57	2,3
TEGYAAFQEDSS@GDEAESPSKMKR	3	4.36	2.6	2
AMGIMNSFVNDIFER	2	4.37	4.71	1,2,3,4
DAEAWFTSR	2	4.37	3.17	1,2,3
LVSESSDVLPK	2	4.37	3.34	2,3
GLQAQIASSGLTVEVDAPK	2	4.37	4.94	2,3
ETMQSLNDR	2	4.37	2.26	2
MIAGQVLDINLAAEPK	2,3	4.37	5.03	2,3

IIGATDSSGELMFLMK	2	4.37	3.83	2
VLQDMGLPTGAEGR	2	4.37	2.2	2
NTTNDLVTAEAQVTPK	3	4.37	3.54	6
LKLEAELGNMQGLVEDFK	3	4.41	3.26	2,3
DFVAEPMGEKPVGSLAGIGEVLGK	3	4.41	4.11	2
GMDKAQGSRPPDQACTGDPELPER	2	4.44	2.56	4
GEQVTNGREAGAELLTEVNR	3	4.49	2.5	3
LQAEIEGLK	2	4.53	2.76	1,2,3
LSELEAALQR	2	4.53	4.24	1,2,3,4
TLQGLEIELQSQLSMK	2	4.53	4.82	2,3
TALINSTGEEVAMR	2	4.53	3.41	2,3
SLETENAGLR	2	4.53	2.36	2
AGTLTVEELGATLTSLLAQAQAQAR	3	4.53	5.03	2
LFIGGLSFETTNESLR	2	4.53	4.07	2
RTVQSLEIDLDSMR	2	4.56	2.79	3
LETSVQGGGGLAMNDRAAAAGSLDR	3	4.56	2.59	4
NHEEEVKGLQAQIASSGLTVEVDAPK	3	4.57	5.59	2,3
EAAVSASDILQESAIHSPGTVEKEAK	3	4.57	3.68	2,3,4
LKSFPEDPQHLGEWGHLDPAEENLK	3	4.63	2.69	3
TLEGELHDLR	2	4.65	2.25	3
T@VFPLADVSRIEEYLK	3	4.68	3.1	2,3
QKVDSLLENLEK	2	4.68	2.35	3
LQTMKEELDFQK	2	4.68	2.88	3
DGLVISCGPDSCGPDSCSEWKPGSLQRFQNK	3	4.68	2.95	6
AALQEATFDPQEVRK	2	4.68	2.33	6
NHEEEISTLR	2	4.75	2.64	3
VAPEEHPVLLTEAPLNPK	3	4.75	2.97	3
TEMENEFVLIKK	2,3	4.78	2.82	3
LAADDFRTKFETEQALR	3	4.78	3.46	3
QS@LMLMATSNEGCKATYEQGVEK	3	4.79	2.53	3
ELNETFKEAQR	2	4.79	2.38	4
LTKEDEQQQALQDIASRCTANDLKCIIR	3	4.86	3.43	3
LKLEAELGNMQGLVEDFKNK	3	4.87	4.26	3
RFDTEEEFKK	2	4.87	2.4	4
RPAGECPITMSDLEAKPSTEHLGDKI	3	4.9	2.75	2
TEELNREVAGHTEQLQMSR	3	4.9	2.8	3,4
FGKGY@FLEIKLKDWIENLEVDR	3	4.94	3.33	2
HSGPNSADSANDGFVR	3	5.21	3.25	3
MGPLGLDHMASSIER	2.3	5.3	3.32	3

FGAQLAHIQALISGIEAQLGDVR	2,3	5.32	4.65	3,4
VHIEIGPDGR	2	5.32	2.49	3
EVAGHTEQLQMSR	2	5.4	3.47	3,4
NSNLVGAAHEELQQSR	2	5.4	3.72	3
LVLEVAQHLGESTVR	2	5.4	3.73	3
TGIVDISILTTGMSATSR	2	5.5	2.27	3
TIAQGNLSNTDVQAAK	2	5.5	3.75	3
TIDQGNILTLFY@LIIR	3	5.5	2.66	5
HRDFVAEPMGEKPVGSLAGIGEVLGK	2	5.53	3.62	3
DASRRTGDEKGVEAIPEGSHIK	3	5.56	2.75	3
MGLAMGGGGGASFDR	2	5.59	2.61	3
MGPAMGPALGAGIER	2	5.75	2.8	3,4
MGAGMGFGLER	2	5.75	2.33	3
SCMLQADSEKLR	2	5.79	2.68	1
TKFETEQALR	3	5.81	3.41	4
DSGGQTSAGCPSGWLGTR	2	5.83	2.59	3
DFLAGGVAAISK	2	5.84	4.07	3
LTMQNLNDR	2	5.84	2.61	3,4
IVLQIDNAR	2	5.84	2.5	3,4
IDSLSAQLSQLQK	2	5.84	2.74	3
NFILDQTNVSAAAQR	2	5.84	4.38	3
NQVALNPQNTVFDAK	2	5.84	3.11	3
LNDFASTVR	2	5.84	2.38	3
LTFDSSFSPNTGK	2	5.84	2.81	3,4
DTNGSQFFITTVK	2	5.84	2.2	3
LTLSALIDGK	2	5.84	2.83	3
NDGAAILAAVSSIAQK	2	5.84	2.3	3
LLFNDVQTLK	2	5.84	2.22	4
AQIFANTVDNAR	2	5.88	2.81	3,4
AFITNIPFDVK	2	5.88	2.83	4
APILIATDVASR	2	5.88	2.2	4
MVSEAETLKSPTQR	3	5.9	2.51	6
KVIDDTNITR	2	5.96	2.43	4
VLALPEPSPAAPTLR	2	5.97	2.28	4
VFLENVIR	2	5.97	2.95	4
VQAQVIQETIVPK	2	5.97	3.58	4
CVALESQLMK	2	5.99	2.76	1,2,3
LLEAQIATGGVIDPVHSHR	3	5.99	3.84	4
GGMGSGGLATGIAGGLAGMGGIQNEK	3	6	3.06	3,4,5

LSSEMNTSTVNSAR	2	6	3.35	3,4
ILGATIENSR	2	6	2.59	4
IESLSSQLSNLQK	2	6	3.44	4
ITPENLPQILLQLK	2	6	2.51	4
LLLPGELAK	2	6	2.37	4
LFIGGLNVQTSESGLR	2	6	4.35	4
APVPASELLASGVLSR	2	6.05	3.35	4
AGVNFSEFTGVWK	2	6.05	2.55	4
QMY@CVFNRNEDACR	2	6.06	2.36	2
RILQPMLDSSCSETPK	3	6.06	2.79	3
RVLDELTLAR	2	6.07	2.83	4
KSDVEAIFSK	2	6.07	2.57	4
LKDLEALLNSK	2	6.07	2.71	4
VDIKRTVAAPSVFIFPPSDAAELSR	3	6.09	3.02	2
EITALAPSTMK	2	6.1	2.56	3,4
EQGVLSFWR	2	6.1	2.5	3
VETGVLKPGMVVTFAPVNVTTEVK	3	6.11	2.63	4
KDILCDVTLIVERK	3	6.11	3.05	5
RQLETLGQEK	2	6.14	2.48	4
RY@RFEGEGDIQR	2	6.18	2.23	1
VLDAS@WYSPGTREARKEY@LER	3	6.23	2.64	2
EKEKAQLAAEALK	2	6.33	2.42	3
SQIHDIVLVGGSTR	2	6.46	3.09	4
SKWQMENSNLDLSHFK	3	6.47	2.52	3,4
SLESLHSFVAAATK	2	6.47	3.29	4
VEQKLHLSGVQAVTCDR	3	6.71	2.62	5
DSKLCSVLTQDFCMLFNNKHEK	3	6.74	2.8	3
HMIDGRWCDCKLPNSKQSQDEPLR	3	6.74	2.68	3
CMALAQLLVEQNFPAIAIHR	3	6.74	3.2	4
LESGMQNMSIHTK	2	6.75	3.8	4,5
LAQALHEMR	2	6.75	2.39	4
Y@LEHLVIDKR	2	6.75	2.61	4
QQALT@EFEAYKHR	2	6.76	2.56	4
AGQVVTIWAAGATHSPPTDLVWK	3	6.79	4.7	4
AQIHDLVLVGGSTR	2	6.79	3.18	4
ESTLHLVLR	2	6.85	2.36	4
DCHLAQVPSHTVVAR	2	6.91	2.43	4
MEGQRCSLQAGPGQTTK	2	6.98	2.74	4
TVMIDVCTTCRCMVQVGVISGFK	3	7.54	3.05	2

TCVSNCTASQFVCK	2	7.64	2.31	5
T@KLRGLYTTAKADAEAECNILR	3	7.85	2.67	4
TPLMKALQCER	2	7.89	2.2	5
S@GGGYGS@GCGGGGGSYGGSGR	2	7.92	2.36	5
STCAINNTLIAFFILTTIK	3	7.94	2.78	2
FLRSMSSACLQCKVLICDSSDHQSR	3	7.97	2.5	6
KQSEMQMKAGVTCEVCMNVVQK	3	8.03	3.06	3
VLAAACLGAALLLLCAAPR	3	8.04	2.55	2
RCELCPHK	2	8.07	2.67	1
EEKSGAIMCENCMTTNQKKALK	3	8.11	2.88	3
Y@GVFCSGNEAVSHY@KLLLQQNK	3	8.16	2.91	5
Y@GVFCSGHNEAVSHY@KLLLQQNK	3	8.17	2.95	6
SKPTTLKPIILNEIVDAHKEK	3	8.18	2.6	6
Y@VTITNCSPLVVK	2	8.2	3.27	1
LYACEVTHQGLISPVTKS@FNR	3	8.21	2.51	3
SEKLWSMEKMK	3	8.22	2.59	4
RCLQVSETMETLRTSR	3	8.25	3.04	3
LARQAEMLTCR	2	8.25	2.43	4
TSFFQALGITTK	2	8.41	2.72	5
TPSIQPSLLPHAAPFAK	3	8.44	2.5	5
KLILAQKLSELAVEK	2	8.5	2.47	2
IKAEPDKIEAFRASLSK	2	8.5	2.68	3
QTARLWAHVY@AGAPVSS@PEYTK	3	8.5	2.67	3
LKGDDLQAIKK	3	8.5	2.58	5
KVASMMESKDVHK	2	8.51	2.25	4
EKNPDMVAGEKRK	2	8.59	2.34	3
FQAKAEANSLKLEVR	2	8.59	2.55	5,6
NRIIYLRPMQQVDT@LTLEQK	3	8.59	2.59	5
KDFQHLISSPLK	2	8.6	2.66	4
VCICAGPVMSKSCLLELARSGK	3	8.68	2.58	4
EQNPY@VVQSIISLIMGMKFFR	3	8.69	2.63	2
VTIAQGGVLPNIQAVLLPK	2	8.72	3.97	5,6
VFIGNLNTLVVK	2	8.72	3.79	5,6
VAVFFGGLSIK	2	8.72	2.49	5
VSVFFGGLSIK	2	8.72	2.52	5
VFIGNLNTAIVK	2	8.72	3.56	5
WSLLQQQK	2	8.75	3.13	5,6
LGGIGQFLAK	2	8.75	2.39	6
IRIDSLSAQLSQLQK	3	8.75	2.73	5

HWPFMVVNDAGRPK	3	8.75	5.01	5
LISLFQAMK	2	8.75	2.44	5
LLSPVVPQISAPQSNK	2	8.75	2.43	5
LAMLTPNSPK	2	8.75	2.41	5
LLIHQSLAGGIIGVK	3	8.76	3.62	6
HQGVMVGMGQK	2	8.76	3.05	5
QTASVTLQAIAAQNAAVQAVNAHSNILK	3	8.76	3.48	5
TITKMCEQALGKGCGADSKK	3	8.8	2.53	1
AGQMVTVWAAGAGVAHSPPSTLVWK	3	8.8	3.46	4
APGSGLALLPLK	2	8.8	2.92	5
ATIAGGGVIPHIHK	2,3	8.81	2.94	5,6
IMAYSRGQTDMCRCSK	3	8.9	3.37	2
QKTACGAPSGICLQVK	2	8.9	2.99	6
QANSIKVSVSSGGDCIRT@YKPEIK	3	9.11	3.08	6
ICFKYYHGISGALRAT@T@PCITVK	3	9.31	2.57	2
MEAVWHVWRGMASTVR	3	9.37	2.69	5
TTNFAGILSQGLR	2	9.41	2.77	5
SGAQASSTPLSPTR	2	9.47	3.35	5,6
GFGFVLFK	2	9.47	2.46	5
IQNAPNSHSAQHVKMGY@GAWY@LK	3	9.53	2.75	2
KFCS@SSCITAYKQKSAK	3	9.6	2.89	2
IVAERPGTNSTGPAPMAPPR	2	9.6	3.22	5
KPGAS@VKVSCRASGY@TFTSYDINWVR	3	9.63	2.78	5
MEKEVGKKINVR	2	9.7	2.56	3
SEMVKGSGIVWKVVMK	2	9.7	2.49	3
FPKGDVISVEKTVKR	3	9.7	3.48	5,6
VAVVTGSTSGIGFAIAR	2	9.72	2.74	5
VMVQPINLIFR	2	9.72	2.32	5
QSSATSSFGGLGGGSVR	2,3	9.75	3.8	5,6
FGPGVAFR	2	9.75	2.89	5,6
ILVATNLFGR	2	9.75	2.56	5
GNFGGSFAGSFGGAGGHAPGVAR	2,3	9.76	4.83	5
HLQLAIR	2	9.76	2.46	5,6
AGLQFPVGR	2	9.79	2.67	5,6
ASASGSGAQVGGPISSGSSASSVTVTR	3	9.79	3.64	5,6
AGGPTTPLSPTR	2	9.79	2.67	5
KDNMRLGLSLATNPK	3	9.99	3.13	3
DPPGRKGDSWLGGTTLRGVTAGPSK	3	9.99	3.02	5
PSAAGINLMIGSTR	2	10.18	3.46	5

RINFLSNNQCENIRR	2	10.26	3.02	2
Y@LPGRGDPKRAQAAHGR	2	10.29	2.37	5
GAVGGGGWAGGCRLR	3	10.35	2.53	3
RLVPRYRLQLLGIACMVICT@R	3	10.72	2.63	3
SSGPTSLFAVTVAPPGAR	2	10.9	2.41	5
LCFLLGRLSIRKVK	2	11.01	2.51	5
MKKSGVLFLLGIILLVLIGVQGTPVVR	3	11.17	2.78	1
AAGFLRSNKIAALFMK	2	11.17	2.51	3
NQGSGAGRGKAAILK	2	11.17	2.6	4
KLAKNKR	2	11.26	2.37	2
QIAQITRELRR	2	11.7	2.42	2
QVGVQIRGLASLQGLPHR	2	12	2.79	5,6
IFAVSRKHAHAINNLR	2,3	12.01	3.08	2
SGKVAAIVVKRPR	2	12.02	2.25	4
LFRLVAASRHLILKK	2	12.02	2.24	4
RVSGSRSNVFMR	3	12.3	2.6	4

@: phosphorylation added by the algorithm
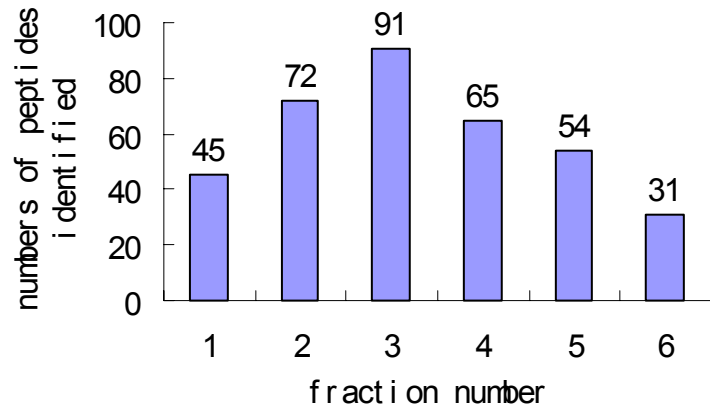
Figure 4-5 Number of peptides identified in each fraction
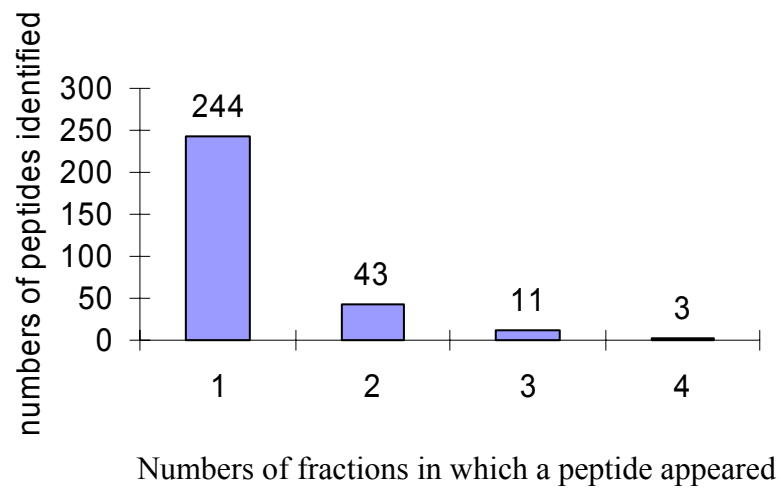
Figure 4-6 Peptide distributions
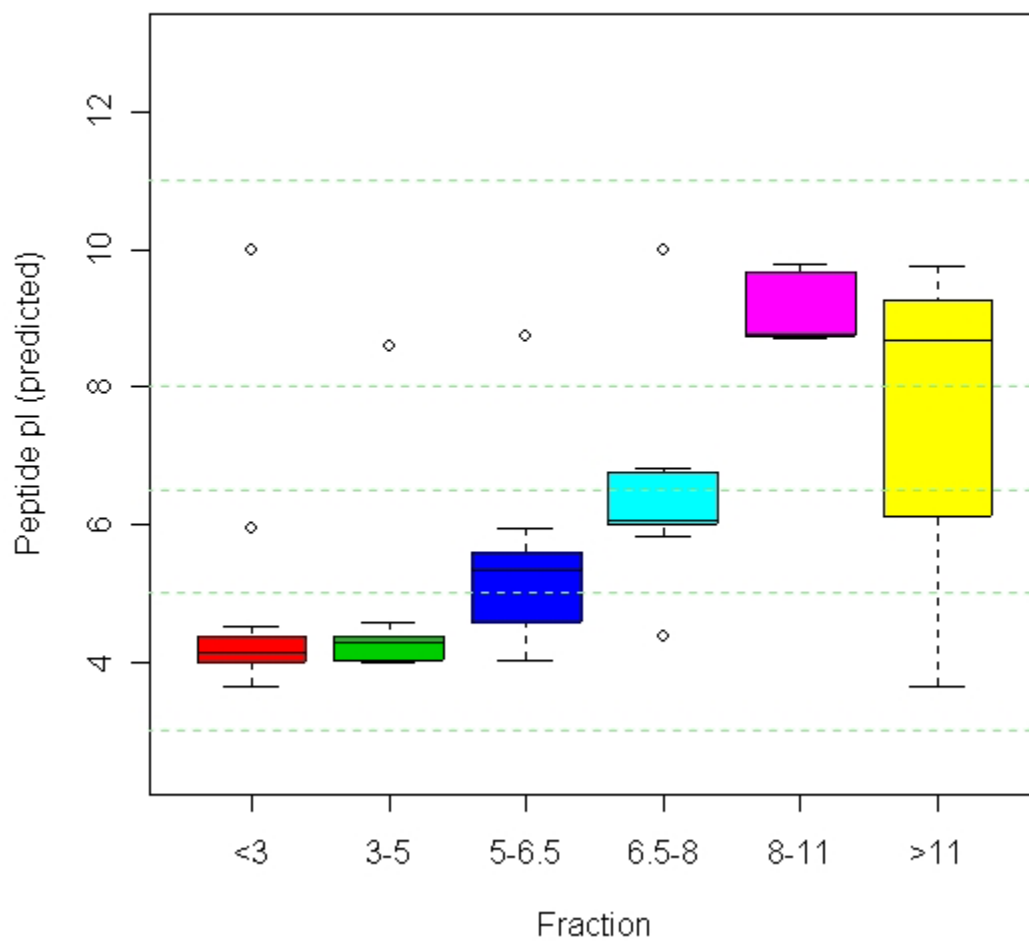
Figure 4-7 Range and average pI of peptides eluted in each fraction
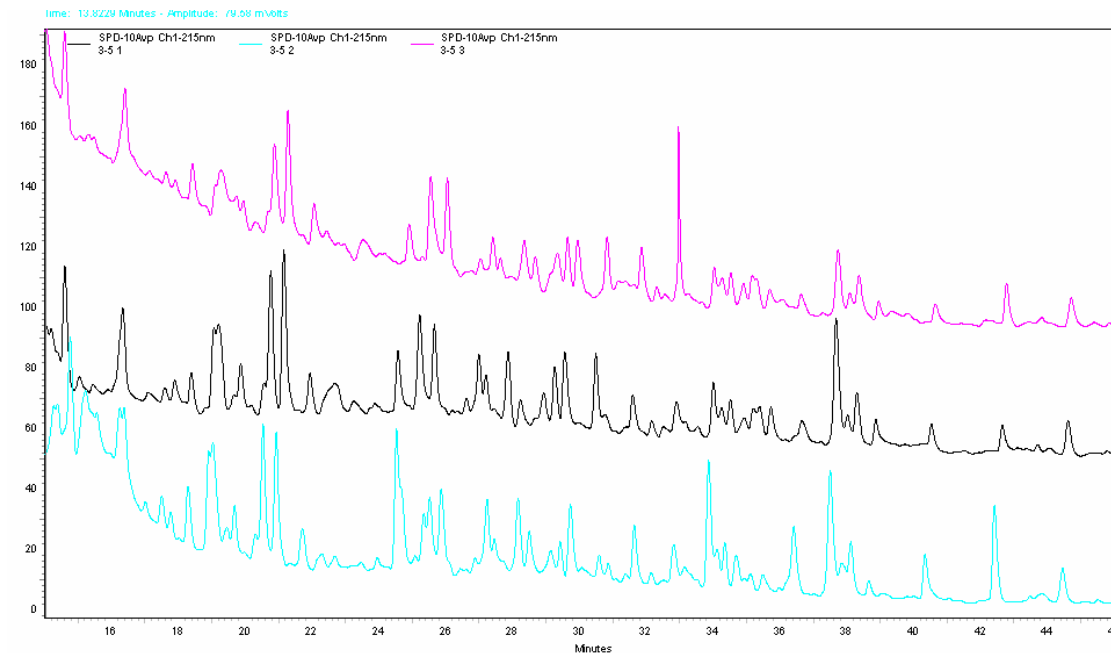
Figure 4-8 Chromatograms of fraction 2 resulted from three repeat solution IEF separations
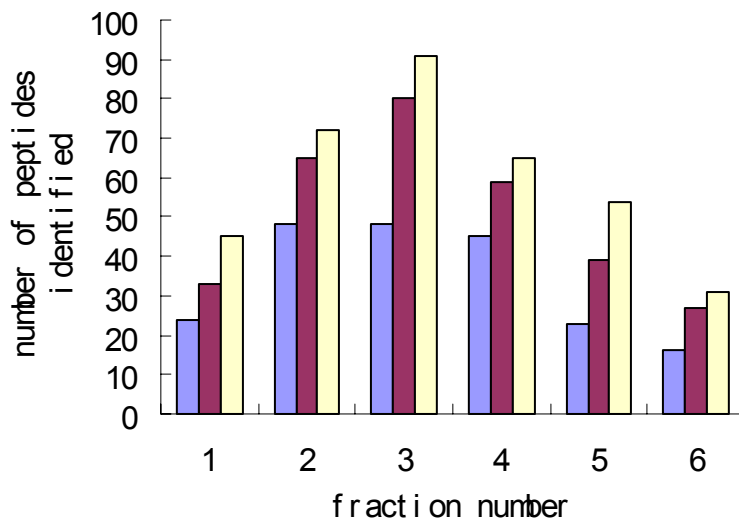
Figure 4-9 Number of peptides identified from 1, 2, and 3 LC-MS/MS analysis. Blue Bars: numbers of peptides identified in each fraction from single LC-MS/MS analyses; purple bars: numbers of peptides identified in each fraction from double LC-MS/MS analysis; white bars: numbers of peptides identified in each fraction from triple LC-MS/MS analysis.

Table 4-2 Percentage of proteins experimentally identified from 1, 3, 6, and 9 combined MudPIT runs[105]

protein copy number per cell[a]	number of proteins[a]	% proteins identified from a different number of combined MudPIT runs			
		1	3	6	9
$>1 \times 10^5$	80	96.3	97.5	97.5	97.5
$(1 \times 10^4)-(1 \times 10^5)$	536	71.6	79.3	83.4	85.6
$(1 \times 10^3)-(1 \times 10^4)$	2184	20.8	24.4	30.6	34.9
$100-1000$	1036	6.5	9.7	13.2	15.5
<100	32	0	3.1	6.3	6.3

[a] Results obtained by Ghaemmaghami et al.[28]

4.2.2 Protein identification and classification

Proteins are identified according to the peptide sequences. Figure 4-10 shows the tandem mass spectra of two peptides with the SEQUEST results list. These two peptides come from the same protein Lamin A/C. Because of their different pIs, one was observed in fraction 2 (pH 3-5) and the other was in fraction 3 (pH 5-6.5). The characterization of several peptides from one protein provides more reliable identification of the protein. Sample loss can't be avoided during separation and detection. If one peptide is lost, the connection with the parent protein is not necessarily lost. But in the top-down strategy, which analyzes intact proteins instead of peptides, loss of a protein is permanent. The proteins identified are listed in Table 4-3 with the Swiss-port access number, name, available sub-cellular location and molecular weight.

The proteins are assigned to a sub-cellular group according to the references, listed in Swiss-port, or using the bioinformatics tool PSORT II (psort.nibb.ac.jp). As shown in Figure 4-11, almost half of the proteins come from nuclei. Mitochondrial proteins and membrane proteins are known to be hard to study because of their poor solubility. Because about one quarter of the identified proteins are classified as mitochondrial or membrane, it seems that the strategy used is effective for insoluble proteins.

These results were also compared with the work of Dr. Fu, another member of the Fenselau group. He isolated nuclei and studied the soluble proteins obtained NaCl buffer, while the present study addresses the centrifugal pellet. Figure 4-12 shows that there is little overlap between proteins identified in the two studies. Actually the sequential extraction used here is used as a prefractionation method in some labs [106].

Our results show that the first extraction was complete and that sequential extraction is effective with the proper buffer system.

Comparison the new two-dimensional solution separation with a one-dimensional LC-MS/MS strategy

The same nuclear peptide mixture was analyzed by LC-MS/MS without fractionation by solution isoelectric focusing. The conditions of LC-MS/MS and SEQUEST identification were identical as described above and used the lower criteria for protein identification. In summary, 54 peptides corresponding to 24 proteins were identified with the single dimensional LC-MS/MS separation, while in the two dimensional solution IEF-LC-MS/MS strategy, 281 peptides corresponding to 167 proteins were identified. There are 26 peptides and 4 proteins identified for the first time with single LC-MS/MS. Most of the proteins identified in the one dimension separation are high abundance proteins, which were identified by multiple peptides. The solution IEF separates the high abundance peptides into several chambers and makes the low abundance peptides, which are covered by high abundant species, more detectable. In the single dimension strategy, about 74 μg of peptides was consumed; while in the two dimensional method only a total of 0.43 μg was used for all eighteen LC-MS/MS analyses. Although the single LC-MS/MS requires more sample, the ion intensity was about 100 times lower than that of the two dimensional strategy. Table 4-4 summarizes the different results of these two methods. This difference in sample consumption resulted from the concentration effect of solution isoelectric focusing. Considering a certain peptide sequence, the molecules are distributed in the whole range of pH 3-10 before isoelectric focusing;

after focusing, all of the molecules are concentrated at the same pH point. The local concentration of a peptide is increased by the focusing process and the peptide is more detectable. So solution isoelectric focusing not only simplifies the complex mixture but also concentrates the diluted sample. The two-dimensional separation method greatly improves the mass spectrometry based protein identification compared with single dimensional separation.

Peptide and protein identification through SEQUEST

Peptides were identified by automatically searching tandem mass spectra against a sequence database. This was realized by the software SEQUEST in this experiment. This software compares an experimental spectrum against the theoretical spectra of all possible peptides that have the precursor ion mass in a sequence database. Each spectrum is assigned to the peptide whose theoretical spectrum has the best match. Two scores are usually used to determine the quality of the assignment: cross correlation (Xcorr) and delta correlation (ΔCn). Xcorr indicates the quality of the match between the spectrum and the peptide; and ΔCn is the difference between the the primary and secondary matches. In this experiment, we used the criteria that ΔCn $\geq$ 0.1, and Xcorr $\geq$ 1.9 for singly charged peptides, Xcorr $\geq$ 2.2 for doubly charged peptides, Xcorr $\geq$ 2.5 for triply charged peptides. By using these criteria, 281 peptides and 167 corresponding proteins were identified. Twenty-two percent of the proteins were identified based on more than two peptides, which means high confidence. But most of the proteins were identified by single peptide. The "one peptide hit" is commonly observed and argued in high-throughput protein identification[107]. In this experiment, each peptide sequence was manually confirmed to be present in its

corresponding protein. But there are still two things to be concerned about: (1) possible false positive identifications of peptides; (2) possible correspondence of one peptide sequence to multiple proteins. The simplest way to get rid of false identifications of peptides is to increase the filtering criteria. If another criteria (ΔCn ≥ 0.1 and Xcorr ≥ 1.9 for +1 ions, Xcorr ≥ 2.5 for +2 ions, Xcorr ≥ 3.75 for +3 ions) were applied to the dataset in this experiment, 212 peptides were considered to be identified corresponding to 110 proteins and 36% of the proteins were identified by multiple peptides. Processing data in this way just simply deletes the peptides which may be wrong, but doesn't increase of confidence of the identifications. Keller and colleagues developed a statistical model to validate the peptide identification made by database searching through SEQUEST[108], and this model was converted to software named PeptideProphet[TM]. This software computes the probabilities that peptide assignments are correct based on the search scores and other information, and assigns a number, which indicates the probability, to each peptide. Processing SEQUEST data in this way can improve the confidence level for identifications. Table 4-5 lists the results processed in different ways. Moreover, the data obtained in this experiment can be processed through another search engine, such as Mascot. Mascot is a probability-based search algorithm[103]. The score obtained from Mascot searching provides the probability of the observation of a match between the experimental data set with each database entry. If the probability of the observed match happening by chance is less than 5%, the match is considered to be significant. The probability is converted to a score by the algorithm for convenience. The best match usually has the highest score and the significant match is typically with a score of the order of 70.

The SEQUEST results can be compared with Mascot results and the overlap identifications are more reliable.

SEQUEST Result
gi|27436946
Peptide:LQEKEDLQELN
DR
Xcorr=4.24
ΔCn=0.17
pI=4.18

SEQUEST Result
gi|27436946
Peptide:NSNLVGAAHEELQQ
SR
Xcorr=3.72
ΔCn=0.30
pI=5.4

93
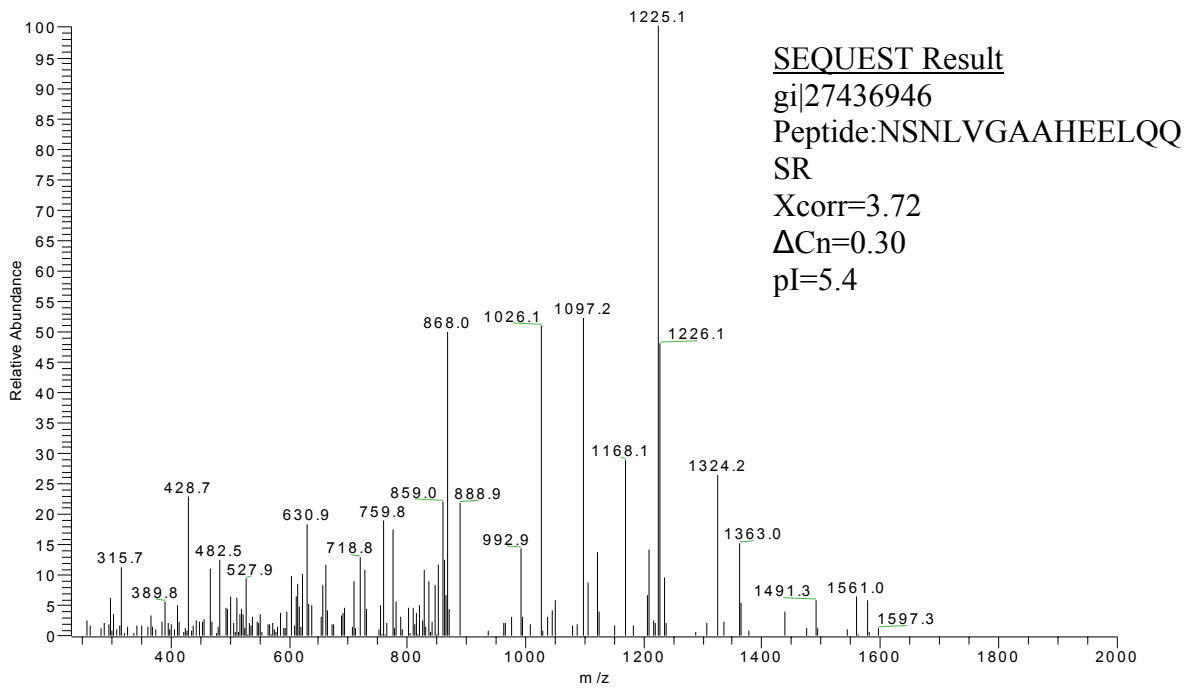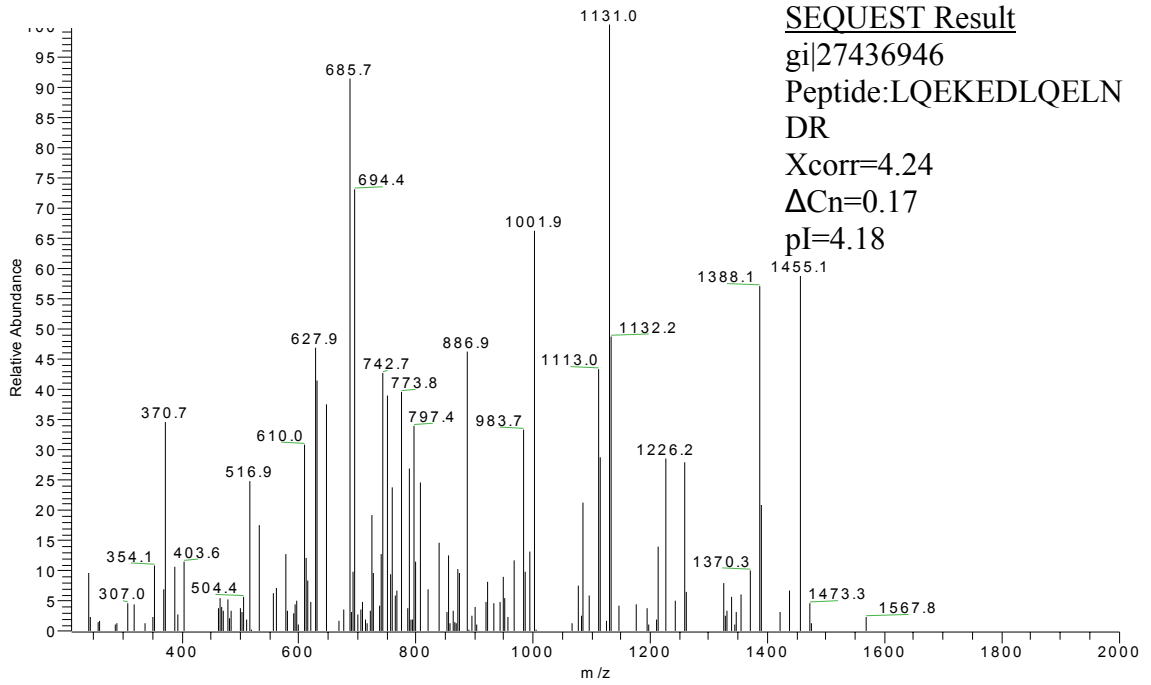
METPSQRRAT RSGAQASSTP LSPTRITRLQ EKEDLQELND RLAVYIDRVR
SLETENAGLR LRITESEEVV SREVSGIKAA YEAELGDARK TLDSVAKERA
RLQLELSKVR EEFKELKARN TKKEGDLIAA QARLKDLEAL LNSKEAALST
ALSEKRTLEG ELHDLRGQVA KLEAALGEAK KQLQDEMLRR
VDAENRLQTM KEELDFQKNI YSEELRETKR RHETRLVEID NGKQREFESR
LADALQELRA QHEDQVEQYK KELEKTYSAK LDNARQSAER
NSNLVGAAHE ELQQSRIRID SLSAQLSQLQ KQLAAKEAKL RDLEDSLARE
RDTSRRLLAE KEREMAEMRA RMQQQLDEYQ ELLDIKLALD
MEIHAYRKLL EGEEERLRLS PSPTSQRSRG RASSHSSQTQ GGGSVTKKRK
LESTESRSSF SQHARTSGRV AVEEVDEEGK FVRLRNKSNE DQSMGNWQIK
RQNGDDPLLT YRFPPKFTLK AGQVVTIWAA GAGATHSPPT
DLVWKAQNTW GCGNSLRTAL INSTGEEVAM RKLVRSVTVV
EDDEDEDGDD LLHHHHGSHC SSSGDPAEYN LRSRTVLCGT CGQPADKASA
SGSGAQVGGP ISSGSSASSV TVTRSYRSVG GSGGGSFGDN LVTRSYLLGN
SSPRTQSPQN CSIM

Figure 4-10 Tandem mass spectra of two peptides coming from one protein: Lamin
A/C. (A) Peptide LQEKEDLQELNDR with pI of 4.18 eluted at 22.87min in
fraction 2 (pH 3-5). (B) Peptide NSNLVGAAHEELQQSR with pI of 5.4 eluted at
24.86min in fraction 3 (pH 5-6.5).

Table 4-3 List of proteins identified from peptides listed in Table 4-1

access No	protein name	location	MW
Q96HG5	actin, beta	C	41005
Q5T9N7	actin, alpha , smooth muscle, aorta	C	16759
Q6MT14	arginine-tRNA ligase	C	63407
P05387	60S acidic ribosomal protein P2	N	11665
P05388	60S acidic ribosomal protein P0	N	34274
P05141	ADP, ATP carrier protein	MM	32764
Q8WWZ7	ATP-binding cassette A5		186561
Q86TL0	APG4-D protein	N	52922
P25705	ATP synthase alpha chain, mitochondrial	M	59751
P06576	ATP synthase beta chain, mitochondrial	M	56560
O75531	Barrier-to-autointegration factor	N	10059
Q71V76	BAT1	C	33142
Q9BX63	BRCA1 interacting protein	N	140878
P35626	beta-adrenergic receptor kinase 2	N	79678
Q9BXX2	breast cancer antigen NY-BR-1.1	M	114250
O75339	Cartilage intermediate layer protein		132538
Q6P1Q4	C9orf76 protein	M	53166
Q15057	centaurin beta 2	C	88029
P00918	carbonic anhydrase II	C	29115
Q13185	chromobox protein homolog 3	N	20823
P45973	chromobox protein homolog 5	N	22225
Q8TB65	cytochrome c oxidase subunit Va	M	16762
Q7Z789	dehydrogenase/reductase(SDR family)member 2, isoform 1	MM	31595
Q9UCE9	D(TTAGGG)N-binding protein B39		2232
Q8N5M0	DDX39 protein	N	36577
Q8WYQ5	DGCR8 protein	N	86045
P49916	DNA ligase III	N	102691
Q5VY62	DEAH (Asp-Glu-Ala-His) box polypeptide 9	C	140957
P33991	DNA replication licensing factor MCM4	N	96576
Q9UHC1	DNA mismatch repair protein Mlh3	N	163712
Q9UNI6	Dual specificity protein phosphatase 12	N	37687
P58107	Epiplakin	C	553094
Q6IPT9	eukaryotic translation elongation factor 1 alpha 1	N	50185
Q96I16	EIF2S2 protein	N	24575
Q6RJW2	Env	C	5514
Q8N671	FMNL 1 protein	N	52433
Q9H469	F-box only protein 37	C	32544
Q86TR5	Full-length cDNA 5-PRIME end of clone CS0DA002YL09 of Neuroblastoma	C	8963
Q6PIJ6	F-box only protein 38	C,N	133954
Q7M4L1	familial Alzheimer's disease protein 1	MM	47043

P41002	G2/mitotic-specific cyclin F	N	87640
Q7Z4Q4	galactokinase 2 variant		49235
Q9H490	GPI transamidase component PIG-U	MM	49921
P52272	hnRNP M	N	77384
P07910	hnRNP C1 / hnRNP C2	N	33688
P22626	hnRNPA2/B1	N	37430
P09651	hnRNP A1	N	38715
P55795	hnRNP H'	N	49264
P61978	hnRNP K	C,N	50976
Q99729	hnRNP A/B	N	36613
P52597	hnRNP F	N	45541
Q00839	hnRNP U	N	90479
Q13151	hnRNP A0	N	30841
Q71UI9	Histone H2A.F/Z variant, isoform 1	N	13509
Q96KK5	H2A histone family, member I	N	13906
P62807	Histone H2B	N	13775
P62805	histone H4	N	11236
P08107	heat shock 70kDa protein 1		70052
P55347	homeobox protein PKNOX1	N	47475
Q96IS6	HSPA8 protein	N	64602
Q15034	HECT domain and RCC1-like domain protein 3	C	117188
Q12906	interleukin enhancer-binding factor 3	N	95384
Q9UKX5	integrin alpha-11	MM	133610
P23229	integrin alpha-6	MM	126619
P06213	Insulin receptor [Precursor]	MM	165307
Q6GMY0	Keratin 8	C,N	53750
P05783	Keratin 18	C,N	47927
P08727	Keratin 19	C,N	44106
Q7Z4S6	kinesin family member 21A	N	187179
Q8N1C2	LOC138046 protein	N	32330
P02545	lamin A/C	N	74139
P20700	lamin B1	N	66277
P24043	laminin alpha-2 chain	MM	342771
Q16891	mitochondrial inner membrane protein	M	83678
Q8NHZ7	methyl-CpG binding domain protein 3-like 2	N	22695
P43243	matrin 3	N	94623
Q9NS69	Mitochondrial import receptor subunit TOM22 homolog	M	15522
Q86YP2	MLL/AF10 fusion protein UPN9610l	C	22425
Q8N108	mesoderm induction early response 1 N1-beta	N	60653
Q96CM7	MGC4562 protein	C	73194
P05114	nonhistone chromosomal protein HMG-14	N	10528
Q9BQD0	NLN protein	M	70304
Q8IVI9	NOSTRIN protein	N	57731
P06748	Nucleophosmin	N	32575

P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit	M	79516
O00148	nuclear RNA helicase	N	49077
P46531	neurogenic locus notch homolog protein 1	MM	272554
Q8N1F7	nuclear pore complex protein Nup93	N	93557
Q9BS12	NUP188 protein	M	94944
Q9UDX0	Oxoglutarate (Alpha-ketoglutarate) dehydrogenase (Lipoamide), isoform 1	M	115934
Q14980	nuclear mitotic apparatus protein 1	N	238274
Q92841	probable RNA-dependent helicase p72	N	72371
O94823	Potential phospholipid-transporting ATPase VB	MM	165391
Q9P1P5	probable G protein-coupled receptor GPR58	MM	34924
Q6P4Q1	PLEKHH2 protein	MM	89122
Q15149	Plectin 1	N	531737
P07205	phosphoglycerate kinase, testis specific	MM	44665
Q9UI79	PRO0195	MM	8193
P23284	peptidyl-prolyl cis-trans isomerase B	C	22742
Q86VH9	polymerase(DNA directed), epsilon	N	259331
Q8IVA1	purkinje cell protein 2 homolog	C	10538
O75340	programmed cell death protein 6		21868
P09874	poly[ADP-ribose] polymerase-1	N	112953
Q15311	RalA binding protein1	MM	75932
O43374	Ras GTPase-activating protein 4	C	90458
Q8WYN9	ribosomal protein S27A	R	14077
P62306	small nuclear ribonucleoprotein F	N	9725
Q9HCY8	S100 calcium-binding protein A14	MM	11662
Q9NR46	SH3 domain GRB2-like protein B2	C	43974
Q07325	small inducible cytokine B9	S	14016
O04376	synaptogyrin-2	MM	24810
Q86TN5	similar to origin recognition complex subunit 3	MM	78275
Q15393	splicing factor 3B subunit 3	N	135592
O75044	SLIT-ROBO Rho GTPase activating protein 2	C	120881
Q96IE3	similar to plectin 1, intermediate filament binding protein	N	113766
Q13242	splicing factor, arginine/serine-rich 9	N	25542
Q9BZZ2	sialoadhesin	MM	182624
Q15459	splicing factor 3 subunit 1	N	88886
P62318	small nuclear ribonucleoprotein Sm D3	N	13916
Q8TF17	SH3 domain and tetratricopeptide repeats containing protein 2	N	144777
Q86XD3	similar to nucleoporin 214kDa	MM	216993
Q5VYJ4	small nuclear ribonucleoprotein polypeptide E-like 1	N	10678
Q9BYV9	transcription regulator protein BACH2	N	92537
Q16762	thiosulfate sulfurtransferase	M	33298
Q13263	transcription intermediary factor 1-beta	N	88550
Q86YP4	transcriptional repressor p66 alpha	N	68063

Q13148	TAR DNA-binding protein-43	N	44740
Q9UDY2	tight junction protein ZO-2	C, N	133972
O60466	TGF beta receptor associated protein-1	M	97176
Q9UL49	transcription factor-like 5 protein	N	48263
P63165	Ubiquitin-like protein SMT3C	N	11557
P22695	ubiquinol-cytochrome-c reductase complex core protein	M	48443
P61086	ubiquitin-conjugating enzyme E2-25	MM	22275
Q9HAR7	uterus-ovary specific putative transmembrane protein UO	MM	40173
O95973	VH4 heavy chain variable region [Precursor] [Fragment]		16315
P21796	Voltage-dependent anion-selective channel protein 1	M	30641
Q9Y277	Voltage-dependent anion-selective channel protein 3	MM	30659
O75717	WD repeat and HMG-box DNA binding protein 1	N	125967
Q9P2P5	KIAA1301 protein	C	176649
Q9P2D5	KIAA1412	MM	154717
O43308	KIAA0425 protein	MM	141147
Q8TF60	KIAA1941 protein	C	139646
Q96PV7	KIAA1931 protein	C	56007
Q9P212	KIAA 1516 protein	C	258947
Q96C57	hypothetical protein FLJ12448	N	28257
Q6ZU46	hypothetical protein FLJ44004		100876
Q96SS4	hypothetical protein FLJ14665	C	46008
Q8N7M6	hypothetical protein FLJ40844	N	42632
Q8N274	hypothetical protein FLJ33834	C	86052
Q8N7T4	hypothetical protein FLJ40379	C	26719
Q8TBC5	hypothetical protein FLJ12895	C	54875
Q6ZSZ5	hypothetical protein FLJ45102	M	130787
Q6ZR28	hypothetical protein FLJ46703	M	18354
Q8NB25	hypothetical protein C6orf60	C	119327
Q6P5T5	hypothetical protein	N	26538
Q86YV1	hypothetical protein FLJ00411	C	46571
Q9H5W0	hypothetical protein FLJ22961	N	53014
Q8N2J3	hypothetical protein FLJ90556	C	30633
Q6P3V1	hypothetical protein FLJ39155		110373
Q8NDM7	hypothetical protein DKFZp434L086	N	116769
Q96MF5	hypothetical protein FLJ32446	M	33239
Q8NB06	Hypothetical protein FLJ34423	N	74342
Q96L26	hypothetical protein	C	6740

N: nuclear; M: Mitochondrial; C: cytoplasmic; MM: membrane;

protein subcellular location

unknown, 4%

membrane 17%

mitochodrial 9%

cytoplasmic 22%

nuclear 48%
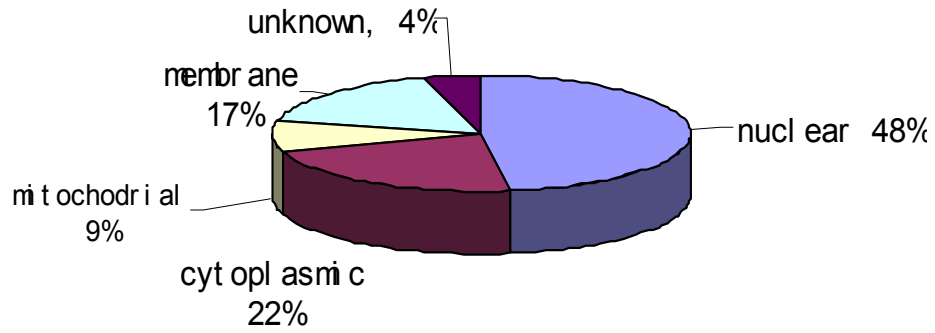
Figure 4-11 Subcellular distributions of proteins identified

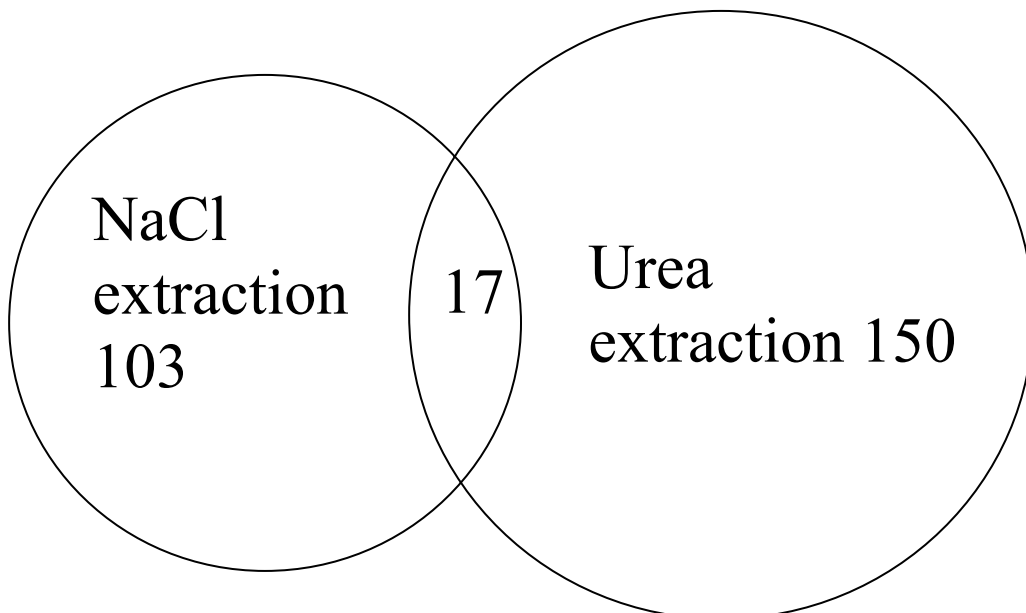NaCl extraction 103

17

Urea extraction 150

Figure 4-12 Comparison of the proteins identified in this study with Dr. Fu's results[98]

Table 4-4 Comparison of two-dimensional IEF-LC-MS/MS with one-dimensional

LC-MS/MS

	Solution IEF-LC-MS/MS	Single LC-MS/MS
Number of peptides identified	281	54/26
Number of proteins identified	167	24/4
Sample consumption	0.47 μg	74 μg
MS spectrum intensity	~10^8 units	~ 10^6 units

Table 4-5 Results of different data processing ways*

Criteria	No. of peptides	No. of proteins	Percentage of proteins identified by multiple peptides
1	281	167	22%
2	212	110	36%
3	192	75	41%

1: Xcorr≥ 1.9 for +1peptides; Xcorr ≥ 2.2 for +2 peptides; Xcorr≥ 2.5; and ΔCn ≥ 0.1

2: Xcorr≥ 1.9 for +1peptides; Xcorr ≥ 2.5 for +2 peptides; Xcorr≥ 3.75; and ΔCn ≥

0.1

3: PeptideProphet[TM] processing: higher than 90% confidence

* Same data sets

4.2.3 Integration of ^{18}O labeling with solution isoelectric focusing

The labeling method introduced in this laboratory [2-3], which decouples digestion and labeling, was employed in this experiment. Lysozyme was digested and labeled with ^{18}O or ^{16}O and the resultant peptides were analyzed with a MALDI-TOF mass spectrometer. Figure 4-13 presents spectra of the singly charged lysozome peptide WWCNDGR with a molecular weight of 993 Da. There is a +4 Da shift between Spectrum A and Spectrum B, which means the peptide was labeled with 2 atoms of ^{18}O. Moreover, in Spectrum B the peak at m/z 994.7, which represents the partially labeled peptide, has abundance less than 5% of highest peak. This means the isotope incorporation is 90%.

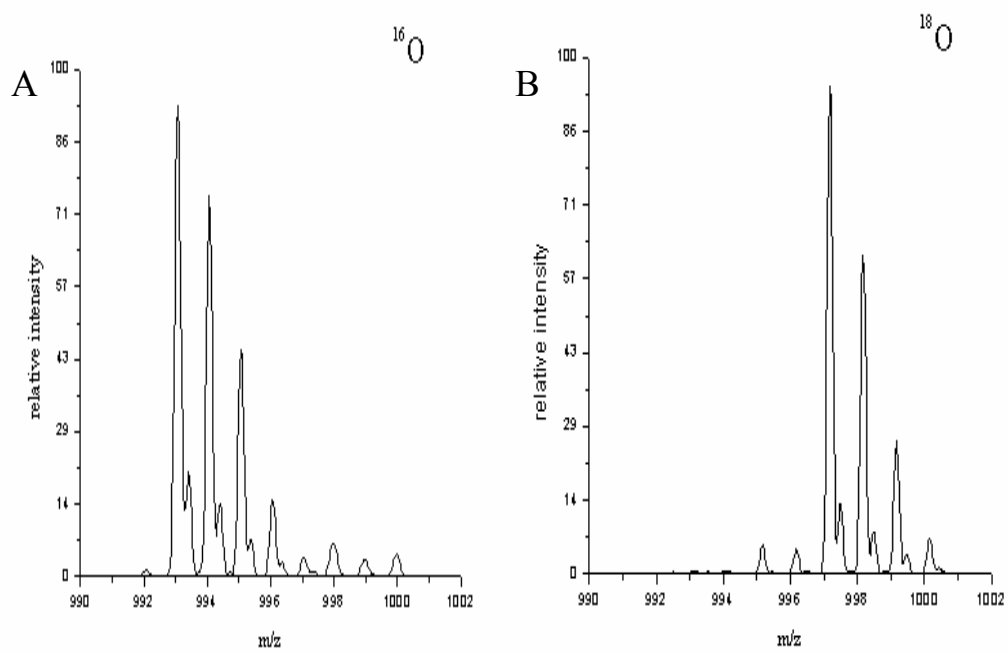Figure 4-13 MALDI mass spectra of peptide WWCNDGR alkylated with iodoacrolymide at the cystine residue. Panel A shows the isotopic distribution of unlabeled peptide; panel B presents peptides incorporated two [18]O atoms at the C-terminal arginine.

Lysozyme was digested and split into two portions. One was incubated in $H_2^{16}O$ and the other in $H_2^{18}O$, both catalyzed with immobilized trypsin for several hours. The two pools were mixed at the ratio of 1:1 and analyzed by mass spectrometry. Figure 4-14 shows the spectra of two peptides pairs: WWCNDGR and GTDVQAWIR. The area of each isotopic peak is listed. The ratios calculated according to equation 1 are 0.97 and 1.01, respectively, in agreement with the mixing ratio. The theoretical isotopic distributions were calculated using the MS-Isotope program in http://prospector.ucsf.edu. The consistency among these peptides suggests that this labeling method is reliable for tryptic peptides. Decoupling the digestion and labeling steps eliminates the need to dry proteins before digestion in $H_2^{18}O$. The nuclear proteins under study here come from the insoluble part of nuclei; some of them tend to precipitate during drying and are difficult to re-dissolve in water. The two-step procedure reduces random sample loss and makes the quantitative analysis more accurate and reliable.
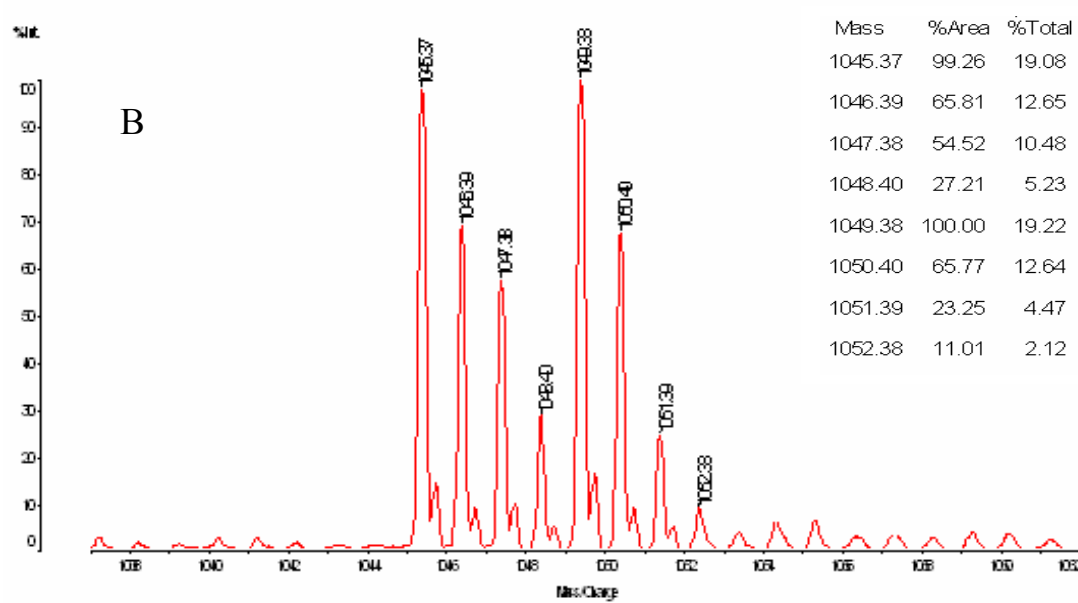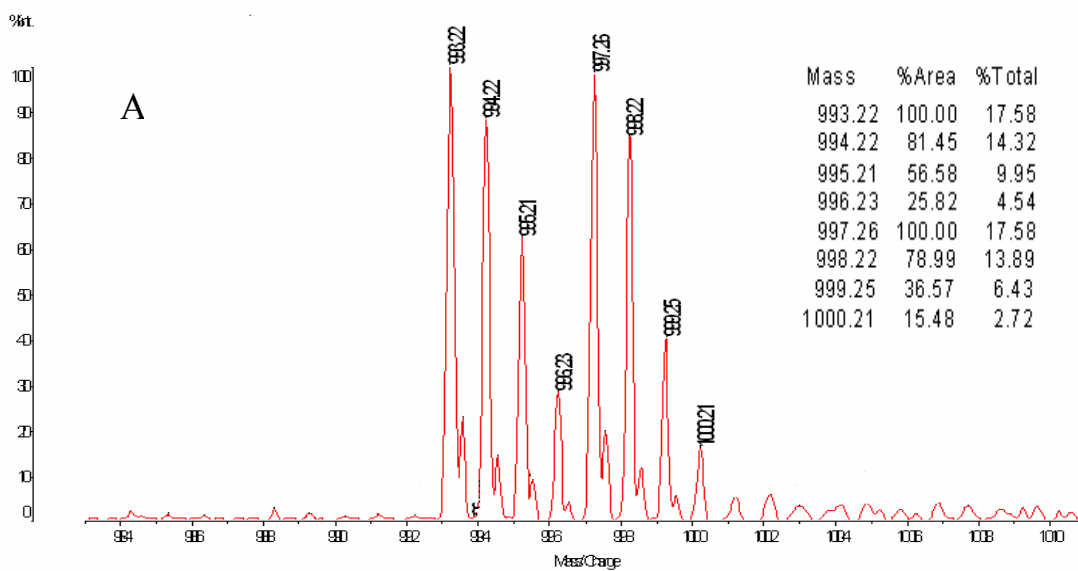
Figure 4-14 MALDI spectra of two $^{16}O/^{18}O$ labeled peptide pairs. A: peptide WWCNDGR with monoisotopic peak at 993.22; B: peptide GTDVQAWIR with monoisotopic peak at 1045.37

The compatibility of ^{18}O labeling with the two-dimensional separation strategy was also investigated using the standard protein mixture described above. The proteins were digested and aliquots were labeled with ^{16}O or ^{18}O. The two pools of peptides were mixed at a ratio of 2:1 and fractionated using solution isoelectric focusing. Then fraction 5-6.5 was desalted and analyzed further with LC-MS/MS. Figure 4-15 presents the partial scan of the doubly charged peptide TGQAPGFTYTDANK from BSA. The labeled and unlabeled peptide pairs co-eluted from the LC column and appeared as isotopic doublets 2 Da apart in the mass spectrum. The introduction of two heavy oxygen atoms does not affect the retention time of the peptide, although a more significant change was observed in ICAT experiments using deuterium[107]. The ratio of ^{16}O/^{18}O was calculated according to equation 1 as 1.96, which agrees with the expected value. The ratio calculated only from the monoisotopic peaks of the unlabeled and labeled peptide is 2.14, in good agreement with the ratio calculated from equation 1.

The peptides are labeled at their C-termini. The labeling helps to identify y-ions in tandem mass spectra, because all y-ions (see Figure 2-2) will carry the double label and be present as doublets in the spectra. Figure 4-16 illustrates this phenomenon. Panel A shows the full mass range of the tandem mass spectrum of the peptide. The spectrum at y11 and y7 are enlarged and presented in panels B and C. Because the y-ions are singly charged, the monoisotopic peaks of labeled and unlabeled peptides are 4 Da apart. The y-ions are much easier to identify because of the doublets and the sequence of the peptides can be assembled by following the y-ions.
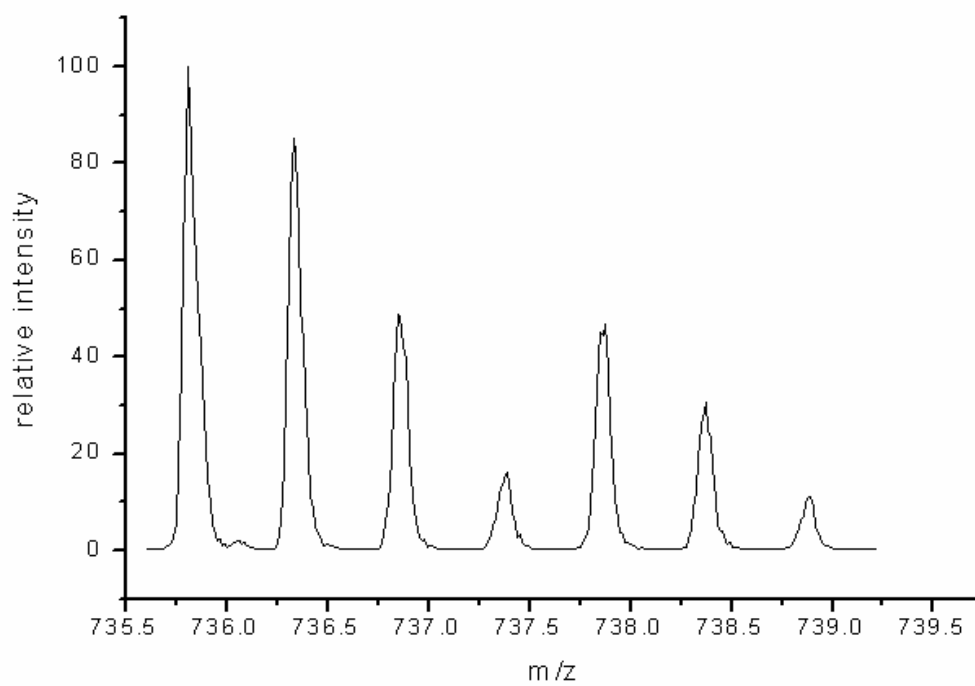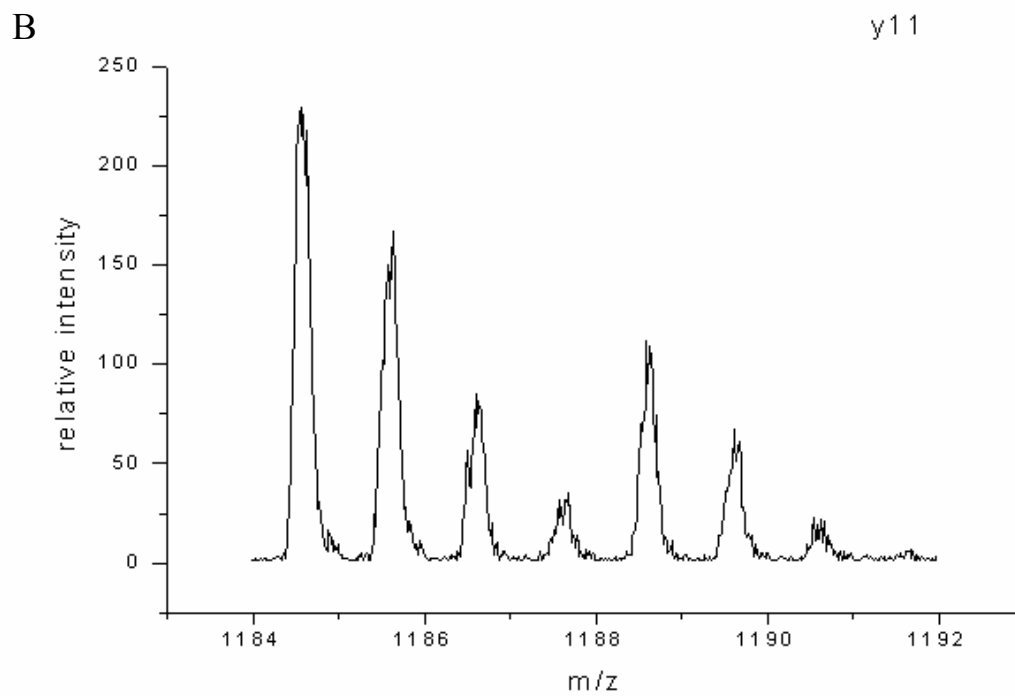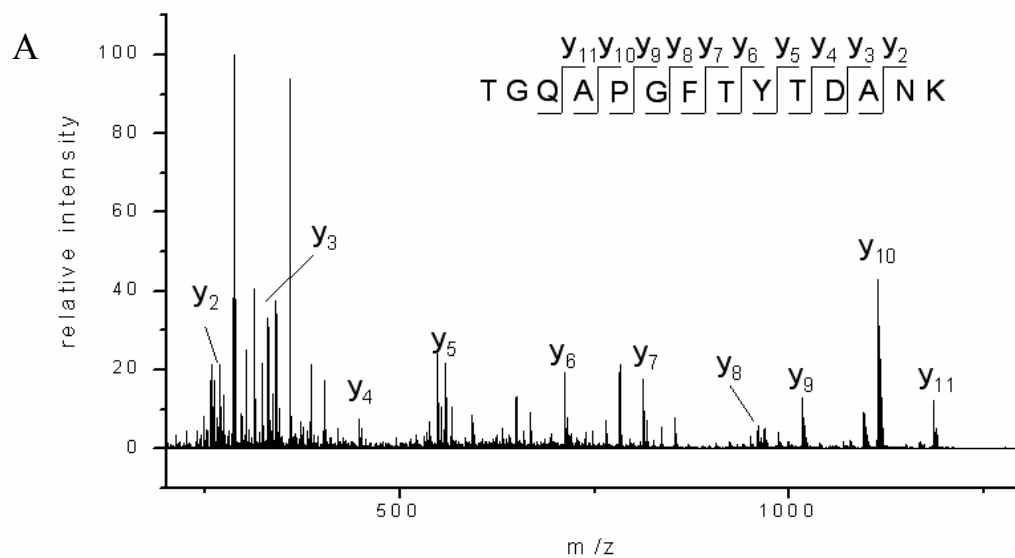
Figure 4-15 ESI mass spectrum of peptide TGQAPGFTYTDANK, [18]O labeled and unlabeled peptides were mixed at a ratio of 1:2.
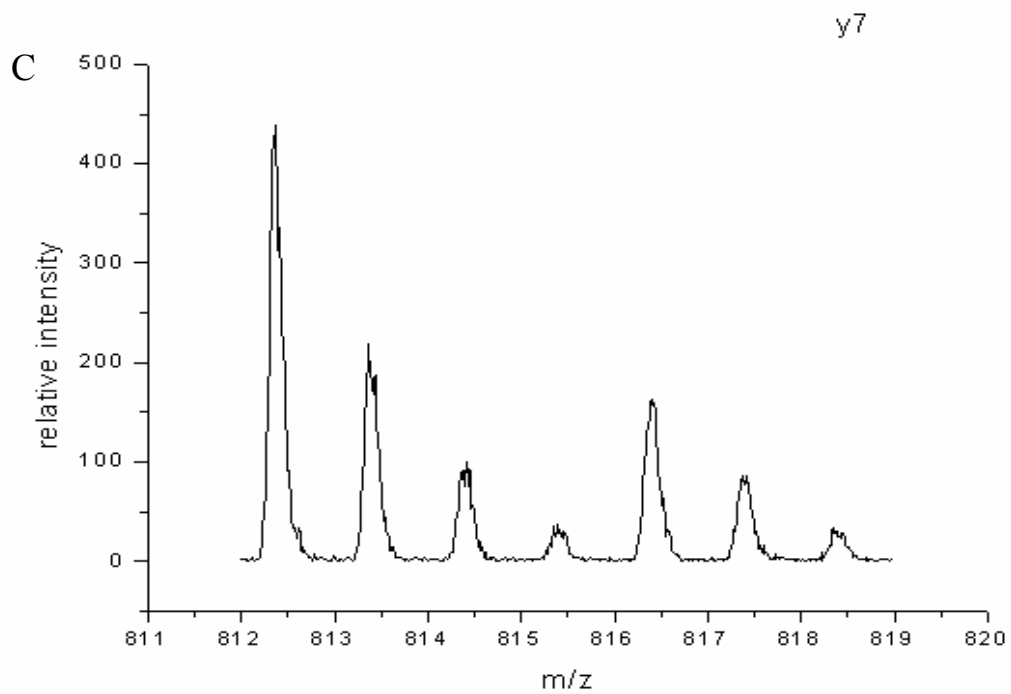
Figure 4-16 Tandem mass spectra of peptide TGQAPGFTYTDANK. A: full mass range; B: enlarged spectrum at m/z 1184; C: enlarged spectrum at m/z 812

4.2.4 Comparative analysis of nuclear proteins of drug susceptible and drug resistant MCF-7 cancer cells

Nuclear proteins extracted from drug susceptible or drug resistant MCF-7 cancer cells were studied using the strategy shown in Figure 4-1. Briefly, the proteins were extracted from nuclei pellets with sample solubilizing buffer (ProteomeSystems, MA) and digested into peptides. The peptides from drug susceptible cells were labeled with $H_2^{16}O$, while the peptides from drug resistant cells were labeled with $H_2^{18}O$. The labeled and unlabeled peptides were mixed, fractionated by solution isoelectric focusing and analyzed by LC-MS/MS. The relative abundances of some proteins were calculated using the peak intensities extracted from mass spectra and are listed in Table 4-4. The protein ratios in the table were calculated based on the spectra from more than two unique peptides. Three drug resistant cancer cell lines described in Chapter 2 were compared with the drug susceptible cell line. We usually consider that the ratio of $^{16}O/^{18}O$ lower than 0.5 or higher than 2.0 is significant. The numbers with red color mean the proteins are of higher abundance in drug resistant cell lines and the blue numbers indicate that the proteins are of lower abundance in drug resistant cell lines.

Table 4-5 Abundance changes of proteins between drug resistant and drug susceptible

MCF-7 cancer cell lines

Protein name	MCF-7/MX	MCF-7/VP	MCF-7/AdrVp
hnRNP A0	ND	ND	9.09±0.03
nonhistone chromosomal protein HMG-14	ND	ND	4.00±0.02
ubiquinol-cytochrome-c reductase complex core protein	ND	5.26±0.22	3.70±0.23
hnRNP F	3.70±0.59	0.82±0.55	ND
hnRNP H2	3.23±0.30	2.27±0.19	2.04±0.30
hypothetical protein 34423	3.03±0.10	ND	2.78±0.06
hsp70	3.03±0.20	2.50±0.15	ND
nucleophosmin	2.70±0.15	1.89±0.31	4.76±0.14
ATP synthase beta	2.70±0.21	2.13±0.18	3.23±0.10
ATP synthase alpha	2.33±0.06	3.33±0.22	ND
Voltage-dependent anion-selective channel protein 1	ND	2.17±0.04	2.22±0.27
hnRNPA2/B1	2.08±0.05	ND	ND
hnRNP K	ND	ND	2.63±0.10
ADT/ATP carrier protein	1.52±0.29	1.92±0.30	2.33±0.10
hnRNP C1 / hnRNP C2	1.43±0.15	2.13±0.09	2.13±0.14
Histone H2A.F/Z variant, isoform 1	1.43±0.19	1.54±0.07	2.27±0.01
Keratin 8, type I cytoskeletal	0.33±0.16	0.41±0.26	0.27±0.10
Keratin 18, type I cytoskeletal	0.26±0.10	0.21±0.06	0.42±0.24
Keratin 19, type I cytoskeletal	0.24±0.09	0.24±0.12	0.33±0.13

ND: not determined

MX: mitoxantrone; VP: etoposide (VP-16); AdrVp: adriamycin

Lower abundance of cytoskeletal keratins

Three cytoskeletal proteins were found to be less abundant in the three drug resistant cells than in drug susceptible cells, including cytokeratin 8, cytokeratin 18 and cytokeratin 19. Cytokeratins function to provide mechanical strength to cells and play a possible role in cell apoptosis[110-113]. Cytokeratins were linked with drug resistance in several cell lines[114-115].It is suggested that keratins may resist to apoptosis by modulating the transportation of apoptotic signaling proteins, however, the role of keratins in resistance to apoptosis remains to be completely investigated. Cytokeratins are common contaminants from human skin, but the *in vivo* isotopic labeling experiments in our lab showed that the cytokeratins were not artifactual[98]. And the observations of lower abundance of cytokeratins in drug resistant cells are similar to Dr. Fu's results[98].

Higher abundance of ATP synthase

ATP synthase was found more abundant in drug resistant cell lines, MCF-7/MX and MCF-7/VP, than in the drug susceptible cell line. ATP snythase is localized in the mitochondrial inner membrane. It catalyzes the production of ATP from ADP with the presence of proton. The ATP synthase complex has two components: the catalytic core including alpha, beta, gamma, delta and epsilon subunits and the membrane proton channel containing a, b and c subunits. The alpha and beta subunits were identified in this experiment. ATP synthase is a high abundance protein and its distribution is not limited in mitochondria. It was also observed by Andersen et. al when they analyze the human nucleolus[116]. The higher abundance of ATP synthase may relate to an energy-dependent enhanced drug efflux process. It has been reported

that the intracellular accumulation of mitoxantrone was significantly decreased in MCF-7/MX cells comparing with drug susceptible MCF-7 cells[117]. A multidrug resistant transporter, breast cancer resistant protein (BCRP), was found to be overexpressed in MCF-7/MX cells[87]. It belongs to the family of ATP-binding cassette transporters which extrude cytotoxic drugs at the expense of ATP hydrolysis[86]. Fanciulli *e*t. al. reported the phenomenon that efflux pump activity is enhanced by the greater ATP availability[118].

Higher abundance of voltage-dependent anion-selective channel protein 1 (VDAC1)

The experiment showed that voltage-dependent anion-selective channel protein 1 had a higher abundance in the etoposide and adriamycin/verapamil drug resistant cells, than in the drug susceptible cells. Voltage-dependent anion-selective channel protein 1 is an outer membrane protein present in the mitochondria and the plasma membrane[119]. It forms a channel through the membrane and allows the diffusion of small hydrophilic molecules. It has been proposed to be involved in the apoptosis[120]. During apoptosis, the apoptogenic protein, cytochrome c, is released from mitochondria into the cytoplasm and actives the death-driving proteolytic proteins known as caspases. The Bcl-2 family of proteins regulates apoptosis by controlling the permeability of mitochondrial membranes[120]. One hypotheses of the regulation mechanism is that the pro-apoptotic proteins in this family interact with VDAC1 and form a large enough pore to release cytochrome c, and that the anti-apoptotic proteins in the family prevent the opening of the pore to inhibit cytochrome c release[121-122]. It is interesting to observe the mitochondrial outer membrane protein in the nuclear

fraction. The mis-location may be resulted from protein translocation in drug resistance or contamination in nuclei isolation. The present findings that VDAC1 has increased abundance in drug resistance may encourage further studies of the regulation process.

Higher abundance of nonhistone chromosomal protein HMG-14

HMG-14 was found to be more abundant in the drug resistant cell line, MCF-7/AdrVp, than in the parental drug susceptible cell line. Chromosomal protein HMG-14 is a non-histone nuclear protein present in higher eukaryotic cells. HMG-14 binds to the inner side of nucleosomal DNA and alters the interaction between the DNA and the histone octamer[123]. It may be involved in the process that maintains transcribable genes in a unique chromotin conformation. The cellular role of this protein is still not clear[124]. The mechanisms postulated to explain the anticancer activity of adriamycin (Adr) include intercalation into DNA bases, inhibiting topoisomerase II-medicated DNA strand breaks, and free radical formation[125]. Increased abundance of HMG-14 may help to maintain the gene chromatin structure and prevent DNA damage introduced by the drug. More work need to be done to clarify the function of HMG-14 and understand its activity in drug resistance.

Higher abundance of heterogeneous nuclear ribonucleoproteins (hnRNPs)

Several heterogeneous nuclear ribonucleoproteins were found to be higher abundance in the drug resistant cell lines than in the drug susceptible cell line. The heterogeneous nuclear ribonucleoprotiens are a large family of nucleic acid binding proteins, which associate with mRNA precursors to form ribonucleoprotein paticles[126]. HnRNPs participate in various processes, such as regulation of

113

transcription[127], splicing[128], and telomere-length maintenance[129]. HnRNPs have been reported overexpressing in several types of cancers[130-132]. The reason for the overexpressions of hnRNPs in cancers is not clear, but it has been associated to the molecular machinery that regulates telomere formation and stabilization[129] and to the control of apoptosis[133-134]. It has been reported that the hnRNPs were dephosphorylated at the early stage of apoptosis. The dephosphorylation may change their activities in statilizing and splicing mRNA and result in the upregulations of mRNA and proteins of the caspase family. The higher abundance of hnRNPs in drug resistant cells may not be sensitive to the dephosphorylation and resistant to apoptosis[133].

Comparison of three drug resistant cell lines

Insoluble proteins from three drug resistant cancer cell lines were studied and several proteins were identified to be significantly changed in abundances compared to the drug susceptible parent line. If we compare the three cell lines, the alterations are not the same. The AdrVp resistant cell line has the most proteins changing significantly (> two folds). This may be related to its high level of drug resistance. More altered proteins provide more clues about mechanism. The proteins, such as hnRNP A0 and HMG-14, were found to change significantly only in the AdrVp resistant cell line and may related to the DNA damage functions of Adr[123, 126]. The changes in abundances of hnRNP F and hnRNP A2/B1 in the MX resistant cell line were not found in the other two lines. HnRNP is reported to be involved in mRNA transportation between the nucleus and cytoplasm[132]. Further studies on these proteins may reveal mechanisms associated with RNA processing in drug resistance.

Chapter 5: Conclusion

The progress of high-throughput proteomic analysis has enabled comprehensive studies of many biological functions. Combinations of high-resolution separations techniques with mass spectrometry are widely used in qualitative and quantitative protein analysis and are continuously being improved. In this study, a free solution isoelectric focusing apparatus was modified and evaluated for protein and peptide separation. Newly commercialized isoelectric membranes, which carry immobilized ampholytes, were used to establish the pH boundaries in the apparatus. The solution isoelectric focusing was coupled with reversed phase liquid chromatography and mass spectrometry for proteomic analysis. A nuclear protein fraction was used to evaluate the performance of the two-dimensional strategy. A total of 281 unique peptides corresponding to 169 proteins were identified by this method. This strategy has been characterized to have high resolution, reproducibility and high sample capacity. The concentration effect in the focusing process dramatically increases the local concentrations of low abundance species and the dynamic range of detection. This two-dimensional method greatly improved protein identification compared with a single-dimensional separation. Because peptides are separated, it can be used to analyze proteins with a wide range of physiochemical properties, such as hydrophobic proteins, and extremely basic proteins.

This two-dimensional separation method is compatible with other proteomic strategies. In the present study, it was combined with a proteolytic isotope labeling

strategy for comparative proteomics. This combination enables analysis of changes of protein abundances in cells or clinical samples under different environments or conditions. The abundances of insoluble nuclear proteins from three drug resistant cancer cell lines were compared with those from a drug susceptible cell line using this method. Nineteen proteins were identified as significantly changed in abundance among the drug resistant cell lines. These proteins require functional confirmation; however, they may be biomarkers for diagnosis of the development of drug resistance, or targets for new anticancer drugs.

References:

1. Mirgorodskays, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen C.P.; Roepstorff, P. Quantitative of peptides and proteins by matrix-assisted laseer desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, 14, 1226-1232.

2 Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal.Chem* **2001**, 73, 2836-2842. Erratum in *Anal. Chem.* **2004,** 76, 2675

3   Yao, X.; Afonso, C.; Fenselau, C.   Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. *J Prot. Research* **2003**, 2, 147-152.

4   Wilkins, M. R.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F. (Eds.) Proteome Research: New Frontiers in Functional Genomics. Springer-Verlag Berlin Heidelberg, New York, **1997**

5   Charlwood J.; Skehel, J.; Camilleri P.   Analysis of N-linked oligosaccharides released from glycoproteins separated by two-dimensional gel electrophoresis. *Analytical Biochemistry* **2000**, 284, 49-59.

6   Oda Y.; Nagase, T.; Chait, B. T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotech.* **2001,** 19, 379-382.

7   Ducret A,; Desponts, C.; Desmarais S.; Gresser M. J.; Ramachandran C.   A general method for the rapid characterization of tyrosine-phosphorylated proteins by mini two-dimensional gel electrophoresis. *Electrophoresis* **2000,** 21, 2196-2208.

8   Gygi, S. P.; Rochon, Y.; Franza B. R.; Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **1999**, 19, 1720-17330.

9   Anderson, L.; Seihamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **1997**, 18, 533-537.

10 Celis, J. E.; Kruhoffer, M.; Gromova, I.; Frederiksen, C.; Qstergaard, M.; Thykjaer, T.; Gromov, P.; Yu, J. S.; Palsdottir, H.; Magnusson, N.; Orntoft, T. F. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **2000**, 480, 2-16.

11 Fenn, B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray

ionization for mass spectrometry of large biomolecules. *Science* **1989**, 246, 64-67.

12 Karas M.; Hillenkamp, F. Laser desorption/ionization of proteins with molecular masses exceeding 100000 daltons. *Anal. Chem.* **1988**, 60, 2299-2301.

13 Takana, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, T. Protein and polymer analyses up to m/z 100000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, 2, 151-153.

14 Edman, P.; Begg, G. A protein sequenator. *Eur. J. Biochem* **1967,** 1, 80-91

15 Blomberg A, B.L., Norbech J, Fey S, Larsen PM, Larsen MR, Roepstorff P, Degand H, Boutry M, Posch A, Gorg A, Interlaboratory reproducibility of yeast protein patterns analyzed by immobilized pH gradient two gel electrophoresis. *Electrophoresis* **1995**, 16, 1935-1945.

16 Righetti, P. G.; Bossi, A. Isoelectric focusing in immobilized pH gradients: an update. *J. Chromatogr. B Biomed. Sci. Appl.* **1997**, 699(1-2), 77-89.

17 Nawrocki, A.; Larsen,M.; Podtelejnikov, A. V.; Jensen, O. N.; Mann, M.; Roepstorff, P.; Gorg, A.; Fey, S. J.; Larsen, M. Correlation of acidic and basic carrier ampholyte and immobilized pH gradient two-dimensional gel electrophoresis patterns based on mass spectrometric protein identification. *Electrophoresis* **1998**, 19, 1024-1035.

18 Norbeck J, Blomberg, A. Two-dimensional electrophoretic separation of yeast proteins using a non-linear wide range (pH 3-10) immobilized pH gradient in the first dimension; reproducibility and evidence for isoelectric focusing of alkaline (pI>7) proteins. *Yeast* **1997**, 13, 1519-1534.

19 Gorg, A.; Obermaier, C.; Boguth, G.; Csordas, A.; Diaz, J. J.; Madjar, J. J. Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis* **1997,** 18, 328-337.

20 O'Farrell, P. H. High-resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, 250, 4007-4021

21 Klose, J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *HUMANGENETIK* **1975,** 26, 231-243

22 Anderson, N. L.; Anderson, N. G. A two-dimensional gel database of human plasma proteins. *Electrophoresis* **1991**, 12, 883-906.

23 Klose, J.; Ursula, K. Two-dimensional electrophoresis of proteins: an updated

protocol and implications for a functional analysis of the genome. *Electrophoresis*, **1995**, 16, 1034-1059.

24 Jungblut, P.; Thiede, B; Zimny-Arndt, U.; Muller, E.-C.; Scheler, C.; Wittmann-Liebold, B.; Otto, A. Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis* **1996,** 17, 839-847

25 Rabilloud, T. Detecting proteins separated by 2-D gel electrophoresis. *Anal. Chem.* **2000**, 72, 48A-55A

26 Damoc, E. Youhnovski, N.; Crettaz, D.; Tissot, J.-D.; Przybylski, M. High resolution proteome analysis of cryoglobulins using Fourier transform-ion cyclotron resonance mass spectrometry. *Proteomics* **2003**, 3, 1425-1433

27 Lehner, I.; Niehof, M.; Borlak, J. An optimized method for the isolation and identification of membrane proteins. *Electrophoresis* **2003,** 24, 1795-1808

28 Gorg, A.; Obermaier, C.; Boguth, G.; Weiss, W. Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: wide pH gradients up to pH 12, longer separation distances and simplified procedures. *Electrophoresis* **1999**, 20, 712-727

29 Gygi, S. P.; Corthals, G.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA*. **2000,** 97, 9390-9395.

30 Dasgupta, P. K. *Anal. Chem.* **1992,** 64, 775A

31 Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. Comprehesive on-line LC/LC/MS of proteins. *Anal. Chem.* **1997**, 69, 1518-1524.

32 Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. III Direct analysis of protein complexes using mass spectrometry. *Nat. Biotech.* **1999,** 17, 676-682.

33 Washburn, M. P.; Wolters, D.; Yates, J. R. III Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotech.* **2001**, 19, 242-247.

34 Wolters, D. A.; Washburn, M. P.; Yates, J. R. III An Automated multidimensional protein identification technology for shotgun proteomics *Anal. Chem.* **2001**, 73, 5683-8690.

35 Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moores, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G., Toward a human blood serum proteome. *Mol. Cell. Proteomics* **2002**, 1, 947-955.

36 Washburn, M. P. Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R. III, Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal. Chem.* **2002**, 74, 1650-1657.

37 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, SALIC as simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, 1, 376-386.

38 Jiang, H.; English, A. M. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *J Prot. Research* **2002,** 1, 345.

39 Wu, C. S. Ed. Handbook of size exclusion chromatography. New York: Marcel Dekker, **1995**

40 Mikkers, F. E. P.; Everaerts, F. M.; Verheggen, T. P. E. M. High performance zone electrophoresis. *J. Chromatogr.* **1979**, 169, 11

41 Bruin, G. J. M.; Chang, J. P.; Kuhlman, R. H.; Zegers, K.; Kraak, J. C.; Poppe, H. Capillary zone electrophoretic separation of proteins in polyethylene glycol-modified capillaries. *J. Chromatogr.* **1989,** 471, 429

42 Bruin, G. J. M.; Huisden, R.; Kraak, J. C.; Poppe, H. Performance of carbohydrate-modified fused silica capillaries for the separation of proteins by zone electrophoresis *J. Chromatogr.* **1989,** 480, 339

43 Caprioli, R. M.; Moore, W. T.; Martin, M.; daGue, B. B.; Wilson, K.; Moring, S. Coupling capillary zone eletrophoresis and continuous-flow fast atom bombardment mass spectrometry for the analysis of peptide mixtures *J. Chromatogr.* **1988,** 480, 247

44 Foret, F.; Sustacek, V.; Bocek, B. J. On-line isotachophoretic sample preconcentration for enchancement of zone detectability in capillary zone electrophoresis *J. Microcol. Sep.* **1990,** 2, 229

45 Shen, Y. F.; Berger, S. T.; Anderson, G. A.; Smith, R. D. High-efficiency capillary isoelectric focusing of peptides. *Anal. Chem.* **2000,** 72, 2154-2159.

46 Jensen, P. K.; Pasa-Tolic, L.; Anderson, G. A.; Horner, J. A.; Lipton, J. A.; Bruce, J. E.; Smith, R. D. Probing proteomes using capillary isoelectric focusing-electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **1999**, 71, 2076-2084.

47 Gysler, J.; Helk, B.; Dambacher, S.;Tjaden, U. R.; Van der Greef, J. Characterization of recombinant cytokine fragments using isotachophoresis-

capillary zone electrophoresis, reversed-phase high performance liquid chromatography, and mass spectrometry. *Pharm. Res.* **1999,** 16, 695-701.

48 Mohan, D.; Lee, C. L. On-line coupling of capillary isoelectric focusing with transient isotachophoresis-zone electrophoresis: a two-dimensional separation system for proteomics. *Electrophoresis* **2002**, 23, 3160-3167.

49 Chen, J.; Balgley, B. M.; DeVoe, D. L.; Lee, C. S. Capillary isoelectric focusing-based multidimensional concentration/separation platform for proteome analysis. *Anal. Chem.* **2003** 75, 3145-3152.

50 Taylor, S. W.; Fahy, E.; Zhang, B.; Glenn, G. M.; Warnock, D. E.; Wiley, S.; Murphy, A. N.; Gaucher, S. P.; Capaldi, R. A.; Gibson, R. A.; Ghosh, S. S. Characterization of the human heart mitochondrial proteome. *Nat. Biotech.* **2003**, 21, 281-286.

51 Rezaul, K.; Wu, L. F.; Mayya, V.; Hwang, S. II; Han, D. A systematic characterization of mitochondrial proteome from a human T Leukemia cells. *Mol. Cell. Proteomics* **2005**, 4, 169-181.

52 Cargile, B. J.; T.alley, D. L.; Stephenson, J. L. Jr. Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. *Electrophoresis* **2004,** 25, 936-945.

53 Cargile, B. J.; Bundy, J. L.; Freeman, T. W.; Stephenson J. L. Jr., Gel based isoelectric focusing of peptides and the utitity of isoelectric point in protein identification. *J Prot. Research* **2004**, 3, 112-119.

54 Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L. Jr. A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics* **2005,** 5, 24-34.

55 Wu, C. C.; MacCoss, M. J.; Mardones, G.; Finnigan, C. Organellar proteomics reveals Golgi arginine dimethylation. *Mol. Biol. Cell.* **2004,** 15, 2907-2919

56 Jiang, X. S.; Zhou, H.; Zhang, L.; Sheng, Q. H. A high-throughput approach for subcellular proteome: identification of rat liver proteins using subcellular fractionation coupled with two-dimensional liquid chromatography tandem mass spectrometry and bioinformatic analysis. *Mol. Cell. Proteomics* **2004,** 3, 441-455

57 Pieper, R.; Gatlin, C. L.; Makusky, A. J.; Russo, P. S.; Schatz, C.R.; Miller, S.S.; Su, Q.; McGrath A.M.; Estock, M.A.; Parmar, P.P.; Zhao, M.; Huang, S.-T.; Zhou, J.;Wang, F.; Esquer-Blasco, R.; Anderson, N.L.; Taylor, J.; Steiner, S., The human serum proteome: display of nearly 3700 chromatographically separatd protein spots on two-dimensional electrophoresis gels and identification of 325

distinct proteins. *Proteomics* **2003**, 3, 1345-1364.

58 Zhou, W. Detection and sequencing of phosphopeptides affinity bound to immobilized metal ion beams by matrix-assisted laser desorption/ionization mass spectrometry. *J Am Soc Mass Spectrom.* **2000**, 11, 273-282.

59 Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hiirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotech.* **2003**, 21, 667-672.

60 Yan, F.; Sreekuman, A.; Laxman, B.; Chinnaiyan, A. M.; Lubman, D. M.; Barder, T. Protein microarrays using liquid phase fractionation of cell lysates *Proteomics* **2003**, 3, 1228-1235

61 Bier, M. Recycling isoelectric focusing and isotachophoresis. *Electrophoresis* **1998,** 19, 1057-1063

62 Kachman, M. T.; Wang, H.; Schwartz, D. R.; Cho, K. R.; Lubman, D. M., A 2-D liquid separations/mass mapping method for interlysate comparison of ovarian cancers. *Anal.Chem* **2002**, 74, 1779-1791.

63 Xiao, Z.;Conrads, T. P.; Lucas, D.A.; Janini, G.M.; Schaefer, C.F.; Buetow, K.H.; Issaq, H.J.; Veenstra, T.D. Direct ampholyte-free liquid-phase isoelectric peptide focusing: application to the human serum proteome. *Electrophoresis* **2004**, 25, 128-133.

64 Herbert, B.; Righetti, P. G. A turning point proteome analysis: sample prefractionation via multicompartment electrolyzers with isoelectric membranes *Electrophoresis* **2000,** 21, 3639-3648

65 Pederson, S. K.; Harry, J. L.; Sebastian, L.; Baker, J.; Traini, M. D.; McCarthy, J. T.; Manoharan, A.; Wilkins, M.R.; Gooley, A.A.; Rightetti, P. G.; Packer, N. H.; Williams K.L.; Herber, B. R. Unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins. *J Prot. Research* **2003**, 2, 303-311.

66 Righetti, P. G.; Castagna, A; Herbert, B.; Reymond, F.; Rossier, J. S. Prefractionation techniques in proteome analysis *Proteomics* **2003,** 3, 1397-1407

67 Zuo X.; Speicher, D.W. A method for global analysis of complex proteomes using sample prefractionation by solution isoelectrofocusing prior to two-dimensional electrophoresis. *Anal. Biochem.* **2000**, 284, 266-278.

68 Zuo X.; Echan, L.; Hembach, P.; Tang, H. Y.; Speicher, K. D.; Santoli, D. Speicher, D. W. Towards global analysis of mammalian proteomes using sample

prefractionation prior to narrow pH range two-dimensional gels and using one-dimensional gels for insoluble and large proteins. *Electrophoresis* **2001**, 22, 1603-1615.

69   Zuo X.; Speicher, D. W. Comprehensive analysis of complex proteomes using microscale solution isoelectrofocusing prior to narrow pH range two-dimensional eletrophoresis. *Proteomics* **2002,** 2, 58-68.

70   Zuo, X.; Hembach, P.; Echan, L.; Speicher, D. W. Enhanced analysis of human breast cancer proteomes using micro-scale solution isoelectrofocusing combined with high resolution 1-D and 2-D gels *J. of Chromato. B* **2002**, 782, 253-265

71   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. USA* **1999**, 96, 6591-9596.

72   Smith R. D.; Pasa-Tolic, L.; Lipton, M. S.; Jensen, P. K.; Anderson, G. A.; Shen, Y.; Conrads, T. P.; Udseth, H. R.; Harkewicz, R.; Belov, M. E. Rapid quantitative measurements of proteomes by fourer transform ion cyclotron resonance mass spectrometry. *Electrophoresis* **2001**, 22, 1652-1658.

73   Smith, R. D.; Anderson, G. A.;Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2002**, 2, 513-523.

74   Lipton, M. S.; Pasa-Tolic, L.; Anderson, G. A.; Anderson, D. J.; Auberry D. L.; Battista, J. R.; Daly, M. J.; Fredrickson, J.; Hixson, K. K.; Kostandarithes, H. Global analysis of the deinococcus radiodurans proteome by using accurate mass tags. *Proc. Natl Acad. Sci. USA* **2002**, 99, 11049-11054.

75   Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton M. S.; Pasa-Tolic, L.; Udseth H. R.; Chrisler, W. B. Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling. *Anal.Chem.* **2001**, 73, 2132-22139.

76   Gu, S.; Pan, S.; Bradbury, E. M.; Chen, X. Use of deuterium-labeled lysine for efficient protein identification and peptide de novo sequencing. *Anal.Chem.* **2002**, 74, 5774-5785.

77   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. Biotech.* **1999**, 17, 994-999.

78   Reynolds K.; Yao, X.; Fenselau, C. Proteolytic 18O labeling for comparative

proteomics:evaluation of endoprotease Glu-C as the catalytic agent. *J Prot. Research* **2002**, 1, 27-33.

79 Brown, K. J.; Fenselau, C. Investigation of doxorubicin resistance in MCF-7 breast cancer cells using shotgun comparative proteomics with proteolytic 18O labeling. *J Prot. Research* **2004**, 3, 455-462.

80 Zang, L.; Toy, D. P.; Hancock, W. S; Sgroi, D. C.; Karger, B. L. Proteomic analysis of Ductal Cartinoma of the breast using laser capture microdissection, LC-MS and 16O/18O isotopic labeling. *J Prot. Research* **2004**, 3, 604-612.

81 Schnolzer, M.; Jedrzejewski, P; Lehmann, W. D. Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **1996,** 17, 945-953

82 Murphy, R. C. Clay, E. L. Synthesis and back exchange of 18O labeled amino acids for use as internal standards with mass spectrometry. *Biomed. Mass Spectrum.* **1979,** 6, 309-314

83 Salmin, S. E.; Sartorelli, A. C. Cancer chemotherapy, in basic and clinical pharmacology. (Katzung, B. G., Ed) Appleton-Lange, **1996,** 881-911

84   Pluen, A. Role of tumor-host interactions in interstitial diffusion of macromolecules: cranial vs. subcutaneous tumors. *Proc. Natl Acad. Sci. USA* **2001**, 98, 4628-4633.

85 Jain, R. K. Delivery of molecular and cellular medicine to solid tumors. *Adv. Drug Deliv. Rev.* **2001**, 46, 149-168.

86 Gottesman, M. M.; Fojo, T.; Bates, S. E. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nature review cancer* **2002,** 2, 48-58.

87 Doyle, L. A, Yang, W.; Abruzzo, L. V.; Krogmann, T.; Gao, Y.; Rishi, A. K.; Ross, D. D. A multidrug resistance transproter from human MCF-7 breast cancer cells. *Proc. Natl Acad. Sci. USA* **1998**, 95, 15665-15670.

88 Litman, T.; Brangi, M.; Hudson, E.; Fetsch, P.; Abati, A.; Ross, D. D.; Miyake, K.; Resau, J. H.; Bates, S. E. The multidrug-resistant phenotype associated with overexpression of the new ABC half-transporter, MXR (ABCG2). *J Cell Sci.* **2000**, 113, 2011-2021.

89 Shen, D.; Pastan, I.; Gottesman, M. M. Cross-resistance to methotrexate and metals in human cisplatin-resistant cell lines results from a pleiotropic defect in accumulation of these compounds associated with reduced plasma membrane binding proteins. *Cancer Res.* **1998,** 58, 268-275

90 O'Brien, M. L.; Tew, K. D. Gluthathione and related enzymes in multidrug resistance. *European Journal of Cancer* **1996,** 32A, 967-978.

91 Cheng, Q.; Lee, H. H.; Li, Y.; Parks, T. P.; Cheng, G. Upregulation of Bcl-x and Bfl-1 as a potential mechanism of chemoresistance, which can be overcome by NF-kappaB inhibition. *Oncogene* **2000**, 19, 4936-4940.

92 Chen, J. S. K.; Konopleva, M.; Andreeff, M.; Multani, A. S.; Pathak, S.; Mehta, K. Drug-resistant breast carcinoma (MCF-7) cells are paradoxically sensitive to apoptosis. *J. Cell. Physiology* **2004**, 200, 223-234.

93 Soule, H. D.; Vazquez, J.; Brennan, M. A cell line from a pleural effusion derived from a breast carcinoma. *J. Natl. Cancer. Inst.* **1973,** 51, 1409-1416

94 Schneider, E.; Horton, J.; Yang, C. H.; Nakagawa, M.; Cowan, K. H. Multidrug resistance-associated protein gene overexpression and reduced drug sensitivity of topoisomerase II in a human breast carcinoma MCF-7 cell line selected for etoposide resistance. *Cancer Res* **1994**, 54, 152-158.

95 Long, B. H.; Musial, S. T.; Brattain, M. G. Comparison of cytotoxicity and DNA breakage activity of congeners of podophyllotoxin including VP-16 and VM-26: a quantitative structure-activity relationship. *Biochemistry* **1984**, 23, 1183

96 Row, T.; Kupfer, G.; Ross, W. Inhibition of epipodophyllotoxin cytotoxicity by interference with topoisomerase-mediated DNA cleavage. *Biochem. Pharmacol.* **1985,** 34, 2483

97 Nakagawa, M.; Schneider, E.; Dixon, K. H.; Horton, J. K.; Kelley, K.; Morrow, C.; Cowan, K. H. Reduced intracellular drug accumulation in the absence of P-glycoprotein (mdr-1) overexpression in the mitoxantrone-resistant human breast cancer cells. *Cancer Research* **1992,** 52, 6175-6181

98 Fu, Z. Comparative proteomics studies of soluble nuclear proteins of drug susceptible and resistant human breast cancer MCF-7 cells. Ph. D dissertation, University of Maryland, **2004**

99 Ross, D. D.; Yang,W.; Abruzzo, L. V.; Dalton, W. S.; Schneider, E. Lage, H. Dietel, M.; Greenberger, L.; Cole, S. P.; Doyle, L. A. A typical multidrug resistance: breast cancer resistance protein messenger RNA expression in mitoxantrone-selected cell lines. *J Natl Cancer Inst* **1999**, 91, 429-433.

100 Bio-Rad protein assay instruction

101 Bradford, M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*

1976, 72, 248-254

102 An, Y.; Fu, Z.; Fenselau, C. Evaluation of solution phase isoelectriofocusing as part of proteomics strategies in press, **2005**

103 Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence database using mass spectrometry data. *Electrophoresis* **1999,** 20, 3551-3567

104 Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* **1994**, 5, 976-989

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

106 Stasyk, T.; Huber, L. A. Zoom in: Fractionation strategies in proteomics. *Proteomics* **2004**, 4, 3704-3716

107 Veenstra, T. D; Conrads, T. P.; Issaq, H. J. What to do with "one-hit-wonders"? *Electrophoresis* **2004,** 25, 1278-1279

108 Keller, A.; Nesvizhskii, A. I.; Koller, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, 74, 5383-5392

109 Sechi, S.; Oda, Y. Quantitative proteomics using mass spectrometry. *Curr. Opin. Chem. Biol.* **2003,** 7, 70-77

110 Hanrahan, J.; Snyder, M.   Cytoskeleton activation of checkpoint kinase. *Molecular Cell* **2003,** 12, 663-673

111 Caulin, C.; Salvesen, G. S.; Oshima, R. G. Caspase cleavage of keratin 18 and reorganization of intermediate filaments during epithelial cell apoptosis. *J. Cell Biol.* **1997,** 138, 1379-1394
112 Caulin, C; Ware, C. F.; Magin, T. M.; Oshima, R. G.   Keratin-dependent epithelial resistance to tumor necrosis factor-induced apoptosis. *J. Cell Biol.* **2000,** 149, 17-22

113 Fuchs, E.; Weber, K. Intermediate filaments: structure, dynamics, function, and disease. *Ann. Rev. Biochem.* **1994**, 63, 345-382

114 Bauman, P. A.; Dalton, W. S.; Anderson, J. M.; Cress, A. E. Expression of cytokeratin confers multiple drug resistance. *Proc. Natl. Acad. Sci. USA* **1994,** 91, 5311-5314

115 Anderson, J. M.; Heindl, L. M.; Bauman, P. A.; Ludi, C. W.; Dalton, W. S.; Cress, A. E. Cytokeratin expression results in a drug-resistance phenotype to six different chemotherapeutic agents. *Clin. Cancer Res.* **1996,** 2, 97-105

116 Anderson, J. S.; Lyon, C. E.; Fox, A. H.; Leung, A. K. L.; Lam, Y. W.; Steen, H.; Mann, M.; Lamond, A. I. Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **2002**, 12, 1-11

117 Rahbar, A. Proteomic analysis of plasma membrane proteins from drug susceptible and drug resistant breast cancer cell lines. Ph.D dissertation University of Maryland, **2004**

118 Fanciulli, M.; Bruno, T.; Giovannelli, A.; Gentife, F. P.;Padova, M. D.; Rubio, O.; Floridi, A. Energy metabolism of human LoVo colon carcinoma cells: correlation to drug resistance and influence of lonidamine. *Clin. Cancer Res.* **2000,** 6, 1590-1597

119 Schwarzer, C.; Barnikol-Watanabe, S.; Thinnes, F. P.; Hilschmann, N. Voltage-dependent anion channel interacts with the dynein light chain Tctex1 and the heat shock protein PBP74 *Int. J. Biochem. Cell Biol.* **2002,** 34, 1059-1070

120 Shimizu, S.; Narita, M.; Tsujinoto, Y. Bcl-2 family proteins regulate the release of apoptotic cytochrome c by the mitochondrial channel VDAC. *Nature* **1999,** 399, 483-387

121 Shimizu, S.; Ide, T.; Yanagida, T.; Tsujinoto, Y. Electrophysiological study of a novel large pore formed by Bax and the voltage-dependent anion channel that is permeable to cytochrome c. *J. Biol. Chem.* **2002,** 275, 12321-12325

122 Shimizu, S.; Konishi, A.; Kodama, T.; Tsujinoto, Y. BH4 domain of antiapoptotic Bcl-2 family members close voltage-dependent anion channel and inhibits apoptotic mitochondrial changes and cell death. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 3100-3105

123 Shick, V. V.; Belyavsaky, A. V.; Mirzavekov, A. D. Primary organization of nucleosomes. Interaction of non-histone high mobility group proteins 14 and 17 with nucleosomes, as revealed by DNA-protein crosslinking and immunoaffinity isolation. *J. Mol. Biol.* **1985**, 185, 329-339

124 Landsman, D.; McBride, O. W.; Soares, N.; Crippa, M. P.; Srikantha, T.; Bustin, M. Chromosomal protein HMG-14 identification, characterization, and chromosome localization of a functional gene from the large human multigene family *J. Biol. Chem.* **1989,** 264, 3421-3427

125 Sinha, B. K. Free radicals in anticancer drug pharmacology *Chem.-Biol.*

Interactions, **1989,** 69, 293-317

126 Choi, Y. D.; Grabowski, P. J.; Sharp, P. A.; Dreyfuss, G. Heterogenous nuclear ribonucleoproteins: role in RNA splicing. *Science*, **1986**, 231, 1534-1539

127 Tomonaga, T.; Levens, D. Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J. Biol. Chem.* **1995**, 270, 4875-4881

128 Van der Houven van Oordt, W. The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J. Cell. Biol.* **2000**, 149, 307-316

129 Ford, L. P.; Wright, W. E.; Shay, J. W. A model for heterogeneous nuclear ribonucleoproteins in telomere and telomerase regulation. *Oncogene* **2002,** 21, 580-583

130 Zhou, J.; Allred, d. c.; Avis, I. Differential expression of the early lung cancer detection marker, heterogeneous nuclear ribonucleoprotein-A2/B1 (hnRNP-A2/B1) in normal breast and neoplastic breast cancer. *Breast Cancer Res. Treat* **2001**, 66, 2217-224

131 Yan-Sanders, Y; Hammons, G. J.; Lyn-Cool, B. D. Increased expression of heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP) in pancreatic tissue from smokers and pancreatic tumor cells. *Cancer Lett.* **2002,** 183, 215-220

132 Pino, I.; Pio, R.; Toledo, G.; Zabalegui, N.; Vicent, S.; Rey, N.; Lozano, M. D.; Torre, W.; Garcia-Foncillas, J.; Montuenga, L. M. Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer. *Lung Cancer* **2003,** 41, 131-143

133 Hermann, R; Hensel, F.; Muller, E.-C.; Keppler, M.; Souto-carneiro, M.; Brandlein, S.; Muller-Hermelink, H. K.; Vollmers, H. P. Deactivation of regulatory proteins hnRNP A1 and A2 during SC-1 induced apoptosis. *Human Antibodies* **2001,** 10, 83-90

134 Brockstedt, E.; Rickers, A.; Kostka, S.; Laubersheimer, A.; Dorken, B.; Wittmann-Liebold, B.; Bommert, K.; Otto, A. Identification of apoptosis-associated proteins in a human Berkitt Lymphoma cell line. *J. Biol. Chem.* **1998,** 273, 28057-28064