INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

University of Alberta

Proteome Profiling of Human Heart Tissues and Squamous Carcinoma Cells

by

Mulu Gebre Gebremedhin

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

requirements for the degree of *Master of Science*

Department of Chemistry

Edmonton, Alberta
Fall 2005

To my parents and family

Abstract

This work focuses on two important areas of proteomics research: comprehensive proteome profiling of heart tissues, and relative quantification of proteins from squamous carcinoma cell lines and heart tissues. Trypsin digestion and microwave-assisted acid hydrolysis (MAAH) were used to generate peptide mixtures that were fractionated by two-dimensional liquid chromatography (LC) and analyzed by electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) tandem mass spectrometry (MS). The proteome information generated by these methods can be used for understanding the protein expression alterations in related cell lines as well as failing heart tissues and other disease states.

Acknowledgments

I would like to thank my supervisor, Dr. Liang Li for his encouragement, guidance and support throughout my study.

Thanks also to Dr. Shaohua Wang and his Secretary Ms. Cherley at the Department of Surgery, University of Alberta for providing the tissue samples and helpful discussions via e-mail and for Cherley for her continued follow up on getting the samples.

I would also like to thank Dr. Manijeh Pasdar and Dr laiji Li, of the Department of Cell Biology at the University of Alberta for their collabrotive work on the cancer cells.

My thanks to all the members in Dr. Liang Li's research group for their collabration and their freindship. Special thanks to Mr. Chengjie Ji who worked with me on the cancer cells and assisted me in labeling tissue samples for quantification.

Thanks to Dr. Monica Palcic and her group for allowing me to use their laboratory and equipment.

I also thank my committee members, Dr. Manijeh Pasdar and Dr. Robert Campbell for their patience and very helpful suggestions to my thesis

My thanks and appreciation to the Department of Chemistry for giving me the opportunity to come here for my study.

I thank all the support from the Department of Chemistry. Especially the kind help and assistance from the general, purchasing, and post offices, the machine and electronic shop and the store room.

I also thank my family and my friends in Edmonton for their encouragement and support which made my stay here much easier and enjoyable.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACN	Acetonitrile
CAD	Collision-activated dissociation
CID	Collision-induced dissociation
CEM	Channel electron multiplier
CPB	Cardiopulmonary bypass (CPB)
DC	Direct current
DTT	Dithiothreitol
DHB	2, 5-dihydroxybenzoic acid
2D	Two dimensional
2-DE	two-Dimensional Gel Electrophoresis
E-cad	E-cadherin
ESI	Electrospray ionization
FT-ICR	Fourier Transform Ion Cyclotron Resonance
FBS	Fetal bovine serum
GRAVY	Grand average of hydrophobicity
HCCA	α-cyano-4-hydroxycinnamic acid
HPLC	High performance liquid chromatography
ICAT	Isotope-Coded Affinity Tags
kDa	Kilo Dalton
LC	Liquid Chromatography
m/z	Mass-to-charge ratio
MALDI	Matrix-Assisted Laser Desorption/Ionization

MAAH	Microwave-assisted acid hydrolysis
MCP	Michrochannel plate
MW	Molecular weight
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MudPIT	Multidimensional protein identification technology
mM	millimolar (1mM = 10^{-3} M)
ns	nanosecond (1ns = 10^{-9} s)
μL	microliter (1 μL = 10^{-6} L)
μg	microgram (1μg = 10^{-6} g)
Pg	Plakoglobin
pI	Isoelectric point
Q	Quadrupole
RF	Radio frequency
RP-HPLC	Reversed phase high performance liquid chromatography
SCC9	Squamous Cell Carcinoma
SCX	Strong cation exchange
SDS	Sodium Dodecyl Sulphate
S/N	Signal to noise ratio
TFA	Trifluoroacetic acid
TOF	Time-of-Flight
UV	ultraviolet

Chapter 1

Introduction:

Proteome Profiling of Human Heart Tissues and Squamous Carcinoma Cells

The terms proteome and proteomics were coined by Marc Williams and colleagues in the early 1990s, and have since been adopted by the research community at large. The proteome refers to the full complement of proteins encoded by the genome. Proteomics can be defined as the qualitative and quantitative comparison of proteome under different conditions to further unravel biological processes. In the rapidly evolving field of proteomics, there is considerable interest in developing methods for large-scale, rapid, and robust analyses of proteins from complex biological samples. Proteomics has traditionally used the separating power of two-dimensional gel electrophoresis for the identification of proteins and quantitative analysis of protein amounts in complex extracts. While this method has been successful in generating useful proteomics data on a number of biological systems, it has some limitations such as difficulties with membrane-associated, very acidic or basic, very low or high Molecular weight, and low abundance proteins which can be overcome by shotgun proteomics. Shotgun proteomics is a gel-free method that uses in-solution digestions of complex sets of proteins, with combinations of multidimensional liquid chromatography coupled with tandem mass spectrometry for peptide sequencing.

Mass spectrometry has become the method of choice for the rapid identification of proteins, determination of changes in protein expression and their modifications in proteome research. In mass spectrometry analysis, biomolecular ions are generated either by electrospray ionization (ESI) or matrix assisted laser desorption ionization (MALDI). These

soft ionization methods make it possible to ionize large, thermally labile biomolecules and transfer them to the gas phase without inducing extensive thermal decomposition. Mass spectrometry can be readily coupled with different separation methods. Two-dimensional gel electrophoresis coupled with mass spectrometry is the traditional way of analyzing complex protein mixtures. More recently, multidimensional liquid chromatography with mass spectrometry has become a powerful technique in identification and characterization of complex protein samples. Despite the different techniques developed, there are still challenges to detect hydrophobic and low abundance proteins. In order to identify and quantify complex protein mixtures, mass spectrometry can be coupled with stable isotope labeling strategies. In the subsequent sections, only the methods/instrumentations relevant to the experiments performed in this project will be discussed.

1.1 Introduction to Mass Spectrometry

Mass spectrometry is an analytical technique that measures the mass-to-charge ratio of individual ions. Molecular weight measurements by mass spectrometry are based upon the production, separation, and detection of molecular ions. A typical mass spectrometer includes an ion source, which ionizes the sample and generates gas phase ions, an analyzer, which separates ions according to individual mass-to-charge ratios, and a detector, which detects and amplifies ions. The choice of ionization method depends on the nature of the sample and the type of information required from the analysis. Soft ionization methods, such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), are widely used for biopolymer analysis.

1.1.1 Ionization Techniques

A variety of ionization techniques are used for mass spectrometry. The most common ionization methods for proteins and peptides are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). MALDI offers higher tolerance toward sample contaminants, such as buffers, salts, and surfactants, higher speed of analysis, and lower sample consumption for each analysis than ESI. However, MALDI is not easily coupled with solution-based separation techniques, as is the case for ESI. ESI, unlike MALDI, can be easily interfaced with modern liquid phase separation techniques.

1.1.1.1 Electrospray Ionization (ESI)

Electrospray Ionization is an atmospheric pressure ionization technique applicable to a wide range of compounds in solution. It was first introduced by Dole in 1968[1]. Electrospray ionization generates ions directly from solution (usually an aqueous or aqueous/organic solvent system) by creating a fine spray of highly charged droplets in the presence of a strong electric field. As the droplet decreases in size, the electric charge density on its surface increases. The mutual repulsion between like charges on this surface becomes so great that it exceeds the forces of surface tension, and ions begin to leave the droplet through what is known as a "Taylor cone". The ions are then electrostatically directed into the mass analyzer. Vaporization of these charged droplets results in the production of singly- or multiply-charged gaseous ions. The number of charges retained by an analyte can depend on such factors as, the composition and pH of the electrosprayed solvent, as well as the chemical nature of the sample. For peptides and proteins, the degree of protonation is dependent upon the number of basic sites. The ionization process and

coupling of ESI to a mass spectrometer has been reported in the literature[2-4]. The ionization process is shown in Figure 1.1. For small molecules (< 2000 Da) ESI typically generates singly or doubly charged ions, while for large molecules (> 2000 Da) the ESI process typically gives rise to a series of multiply-charged species.

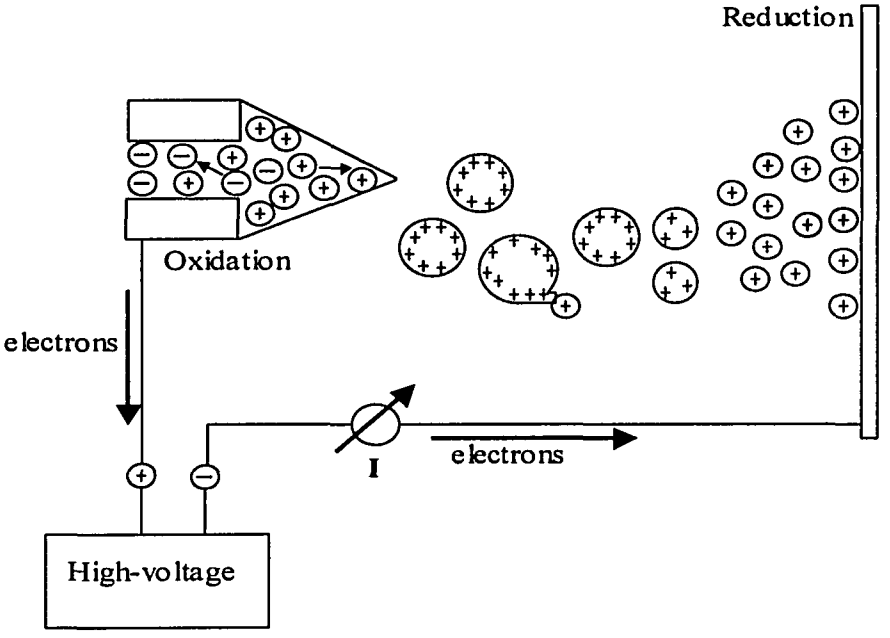

Figure 1.1 Schematic representation of processes in electrospray ionization (ESI).

Because mass spectrometers measure the mass-to-charge (m/z) ratio, the resulting ESI mass spectrum contains multiple peaks corresponding to the different charged states.

ESI allows very sensitive analysis of small, large and labile molecules, such as peptides, proteins, oligosaccharides, and polymers. Another advantage of ESI-MS is that ions are formed directly from solution, a feature that has established the technique as a convenient mass detector for liquid chromatography (LC), in particular, high performance liquid chromatography (HPLC). While past attempts to couple liquid chromatography with mass spectrometry resulted in limited success, ESI has made liquid chromatography-mass

spectrometry routine, adding a new dimension to the capabilities of liquid chromatography characterization. ESI is often coupled to quadrupole or ion trap mass analyzers.

1.1.1.2 Matrix-Assisted Laser Desorption/Ionization (MALDI)

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), introduced in 1987 by two research groups independently[5-7], has become a widespread analytical tool for peptides, proteins and most other biomolecules (oligonucleotides, carbohydrates, natural products, and lipids). The principle of MALDI is shown in Figure 1.2. The efficient and directed energy transfer during a matrix-assisted laser-induced desorption event provides high ion yields of the intact analyte, and allows for the measurement of compounds with high accuracy and sub-picomole sensitivity.

MALDI provides for the nondestructive vaporization and ionization of both large and small biomolecules. In MALDI analysis, the analyte is first co-crystallized with a large molar excess of a matrix compound (matrix to analyte ratio of >500:1), usually a UV-absorbing weak organic acid, after which pulsed UV laser radiation of this analyte-matrix mixture results in the vaporization of the matrix which carries the analyte with it. The matrix therefore plays a key role by strongly absorbing the laser light energy and causing, indirectly, the analyte to vaporize. The matrix also serves as a proton donor and receptor, acting to ionize the analyte in both positive and negative ionization modes, respectively.

Although the ionization mechanism is not well understood, it is widely accepted that, for proteins, ionization occurs via gas phase proton transfer reactions between excited matrix molecules and analyte molecules[8].

The most common matrices used in peptide/protein analysis are α-cyano-4-hydroxycinnamic acid (HCCA), 3, 5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), and 2, 5-dihydroxybenzoic acid (DHB).

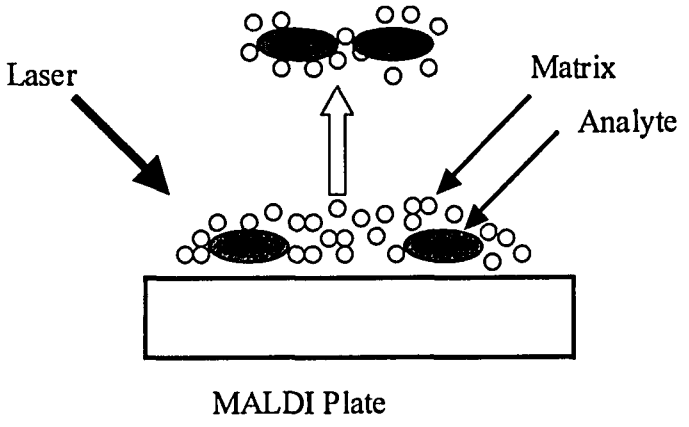

Figure 1.2 Principles of matrix-assisted laser desorption/ionization (MALDI).

Several sample-matrix preparation methods have been developed including dried-droplet[8], vacuum-drying[9], crushed-crystal[10], fast evaporation[11,12], and two-layer method[13]. Of these, the dried-droplet and two layer methods are commonly used in our laboratory.

MALDI has had its biggest impact on the field of protein research. The ability to generate MALDI-MS data on whole proteins and proteolytic fragments is extremely useful for protein identification and characterization. For example, a protein can often be unambiguously identified by the accurate MALDI mass analysis of its constituent peptides (produced by either chemical or enzymatic treatment of the sample). The MALDI sample plate can be used for multi-sample preparation and automated sample analysis. Different parameters can be used to adjust and monitor MALDI autosampling, including: laser position, laser intensity, signal intensity, and mass range.

MALDI can be coupled with many mass analyzers, such as: time-of-flight (TOF), Fourier Transform Ion Cyclotron Resonance (FT-ICR), quadrupole ion trap, and magnetic sector. Among these mass analyzers, TOF is most commonly used with MALDI because its pulsed ion detection mode is well matched with the pulsed ionization in the MALDI process.

1.1.2 Mass Analyzers

After ions are formed in the source region they are accelerated into the mass analyzer by an electric field. The mass analyzer separates these ions according to their mass-to-charge (m/z) value. The selection of a mass analyzer depends upon the resolution, mass range, mass accuracy, scan rate and detection limits required for an application. Each analyzer has different design and operating characteristics. Analyzers are typically described as either continuous or pulsed. Continuous analyzers include quadrupole filters and magnetic sectors. These analyzers are similar to a filter or monochromator used for optical spectroscopy. They transmit a single selected m/z to the detector and the mass spectrum is obtained by scanning the analyzer so that different m/z ratio ions are detected. While a certain m/z is selected, any ions at other m/z ratios are lost, reducing the signal to noise ratio (S/N) for continuous analyzers. Single Ion Monitoring (SIM) enhances the S/N by setting the mass spectrometer at the m/z for an ion of interest. Since the instrument is not scanned, the S/N improves, but any information about other ions is lost. Pulsed mass analyzers are the other major class of mass analyzer. They have some distinct advantages. These instruments collect an entire mass spectrum from a single pulse of ions. This results in a signal-to-noise advantage similar to Fourier transform or multichannel spectroscopic techniques. Pulsed analyzers include: time-of-flight, Fourier ion cyclotron resonance, and quadrupole ion trap mass spectrometers.

Quadrupole Mass Analyzer. A quadrupole mass analyzer is essentially a mass filter that is capable of transmitting only the ion of choice. A quadrupole mass analyzer consists of four parallel rods that have fixed direct current (dc) and alternating radio frequency (RF) potentials applied to them as shown in Figure 1.3. Ions produced in the source of the instrument are then focused and passed along the middle of the quadrupoles. Their motion will depend on the electric fields so that only ions of a particular m/z will be in resonance and thus pass through to the detector. The RF is varied to bring ions of different m/z into focus on the detector and thus build up a mass spectrum.

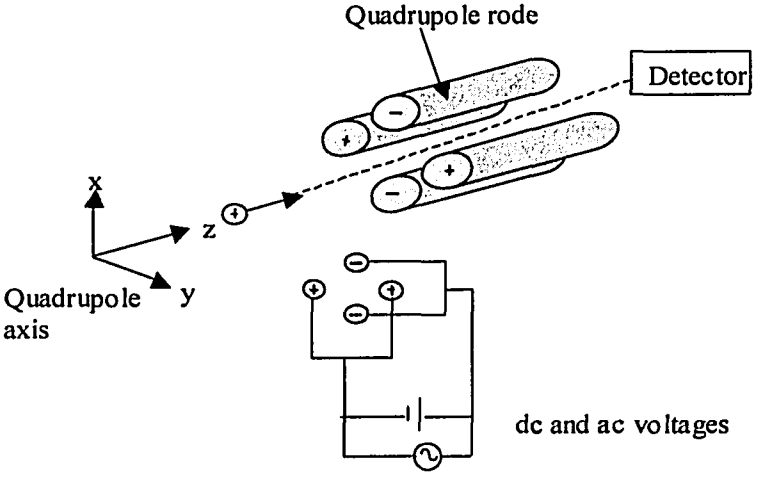

Figure 1.3 Schematic of a quadrupole mass analyzer.

The same absolute potential with different sign is applied to each rod. The two potentials applied are +(U+Vcos(wt)) (labeled '+' on Fig. 1.3) and -(U+Vcos(wt)) where 'U' is the fixed potential and Vcos(wt) is the applied RF of amplitude 'V' and frequency 'w'. The applied potentials on the opposed pairs of rods vary sinusoidally as cos(wt) cycles with time 't'. This results in ions being able to traverse the field-free region along the central axis of the rods but with oscillations amongst the poles themselves. These oscillations result in complex ion trajectories dependent on the m/z of the ions. Specific combinations of the

potentials 'U' and 'V' and frequency 'w' will result in specific ions being in resonance, creating a stable trajectory through the quadrupole to the detector. All other m/z values will be non-resonant and will hit the quadrupoles and not be detected. The mass range and resolution of the instrument is determined by the length and diameter of the rods.

Quadrupole Ion Trap Mass Analyzer. The quadrupole ion trap mass analyzer was introduced at the same time as the quadrupole mass analyzer and by the same person, Nobel Prize winner Wolfgang Paul[14]. The physics behind both of these analyzers is very similar. Quadrupole ion trap consists of two end cap electrodes and a ring electrode as shown in Figure 1.4. In an ion trap, a radio frequency voltage is applied to the ring electrode while the two end-cap electrodes are held at ground potential. One method of using an ion trap for mass spectrometry is to generate ions externally with electrospray and then inject them into the trapping volume. The ions are then ejected and detected as the radio frequency field is scanned[15]. It is also possible to isolate one ion species by ejecting all others from the trap. The isolated ions can subsequently be fragmented by collisional activation and the fragments detected to generate a fragmentation spectrum. The primary advantage of quadrupole ion traps is that multiple collision-induced dissociation experiments can be performed without having multiple analyzers. Other important advantages include its compact size and the ability to trap and accumulate ions to increase the signal-to-noise ratio of a measurement.

Ion trap mass spectrometry has recently undergone very rapid development and is emerging as a high performance technique that shows signs of becoming one of the leading tools in the discipline. These instruments allow tandem mass spectrometry experiments which are possible only by the use of combinations of multiple quadrupoles, such as the highly successful triple quadrupole instrument. Extension to high mass/charge

measurements, the development of high resolution capabilities and the very recent demonstration of non-destructive, broad-band Fourier transform capabilities, all suggest an increased role in the future. Limitations occur in dynamic range, accurate mass measurement, and quantitative precision.

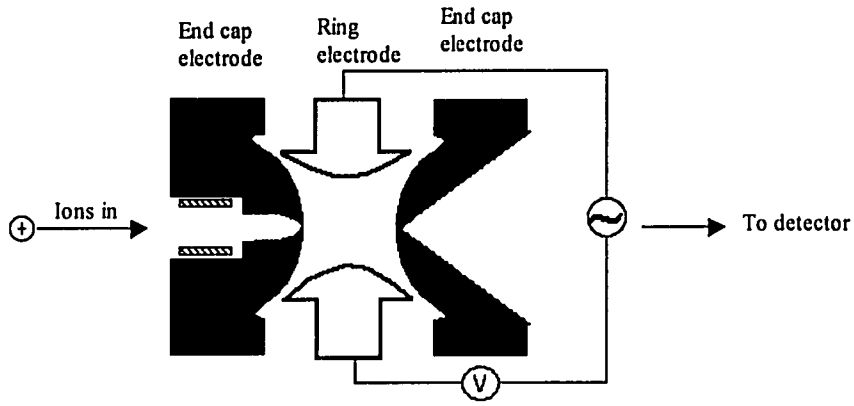

Figure 1.4 Schematic of ion trap mass spectrometer.

Time-of-Flight Mass Analyzer (TOF). A time-of-flight (TOF) analyzer is one of the simplest mass analyzing devices and is commonly used with MALDI. It separates ions in time as they travel down a flight tube[16,17] (Figure 1.5). In the source of a TOF analyzer, a packet of ions is formed by a very fast (nano seconds) ionization pulse. These ions are accelerated into the flight tube by an electric field (typically 2-25 kV) applied between the backing plate and the acceleration grid. Since all the ions are accelerated across the same distance by the same force, they have the same kinetic energy. Because velocity (v) is dependent upon the kinetic energy ($E_{kinetic}$) and mass (m), lighter ions will travel faster.

$$E_{kinetic} = 1/2mv^2 \qquad \textbf{equation 1.1}$$

$E_{kinetic}$ is determined by the acceleration voltage of the instrument (V) and the charge of the ion (e × z). Equation 1.1 rearranges to give the velocity of an ion (v) as a function of acceleration voltage and m/z value.

$$v = [(2V.e)/ (m/z)]^{1/2} \qquad \textbf{equation 1.2}$$

After the ions accelerate, they enter the flight tube. The ions drift through this field-free region at the velocity reached during acceleration. At the end of the flight tube they strike a detector. The time delay (t) from the formation of the ions to the time they reach the detector depends upon the length of the drift region (L), the mass to charge ratio of the ion, and the acceleration voltage in the source.

$$t = L. [(m/z)/ (2V.e)]^{1/2} \qquad \textbf{equation 1.3}$$

Equation 1.3 shows that low m/z ions will reach the detector first. The mass spectrum is obtained by measuring the detector signal as a function of time for each pulse of ions produced in the source region. Because all the ions are detected, TOF instruments have very high transmission efficiency which increases the S/N.

The mass resolution ($R = m/\Delta m$) of a linear TOF mass spectrometer is usually very poor (generally less than 500). The following factors contribute to the poor resolution of the process: (a) different time of formation or acceleration of ions; (b) different initial locations of ions in space; and (c) different initial velocities of ions before acceleration. Time focusing can be achieved, in the case of laser desorption (LD) or matrix-assisted laser desorption/ionization (MALDI), by using pulsed drawout fields with sharp rise times or short laser pulses. A dual-stage extraction[18] method is normally used for correction of the initial spatial distribution of ions in an ion source. And finally, initial velocity (or energy)

distribution can be corrected partially by a time-lag focusing technique which is explained briefly below.

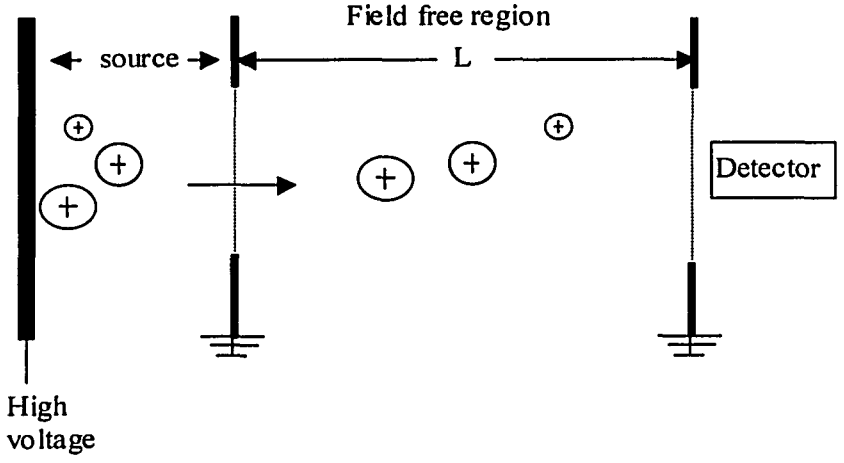

Figure 1.5 A schematic diagram of a linear time of flight (TOF) mass analyzer.

Peak broadening can be reduced by using a time-lag focusing TOF mass analyzer[19-23]. In conventional MALDI-TOF instruments, the ions produced by a pulsed laser beam are immediately extracted from the source. An electric field created by continuously applying dc voltages to the source plates, including the repeller and the extractor, expels the ions from the source to the flight tube for mass analysis. In MALDI-TOF systems equipped with time-lag focusing, the general procedure is to introduce a time delay between ionization and extraction of the ions out of the ion source into the drift tube of the analyzer. During the time delay, the ions drift with the initial velocity imparted by the desorption event. Ions with a greater initial velocity drift faster than ions of lesser initial velocity; therefore, ions of greater initial velocity move farther away from the point of desorption than ions with less initial velocity. After some time, the time delay, a voltage pulse is applied to create an electric field and accelerate the ions. The time delay and extraction potential are tuned to minimize the arrival

time distribution of the ions at the detector. One major drawback of using time-delayed extraction for initial velocity distribution correction is that it is mass dependent, which is problematic for a TOF mass spectrometer intended to record the entire mass range simultaneously.

Alternatively, an approach to mass independent compensation for the initial velocity distribution of ions has been possible with the invention of the ion mirror or reflectron. The reflectron, located at the end of the flight tube, is used to compensate for the difference in flight times of the same m/z ions of slightly different kinetic energies by means of an ion reflector (Figure 1.6). The idea is that the ions of higher kinetic energy penetrate deeper into the field and take a longer time to return than slow ions. Fast and slow ions are focused in time at the detector. Therefore, with proper geometry and voltage on the reflectron, the initial energy spreads are largely compensated, and the mass resolution is greatly improved. For a reflectron TOF, resolution is generally over 2000 and in some cases can be as high as 20,000. The spectral recording speed of TOF can be as high as 10,000 spectra/second, while other mass analyzers usually can only record up to 10 spectra/second. TOF has unlimited mass range; an ion of up to approximately 1 million m/z can be detected but detection is usually limited by the effectiveness of the ion detector.

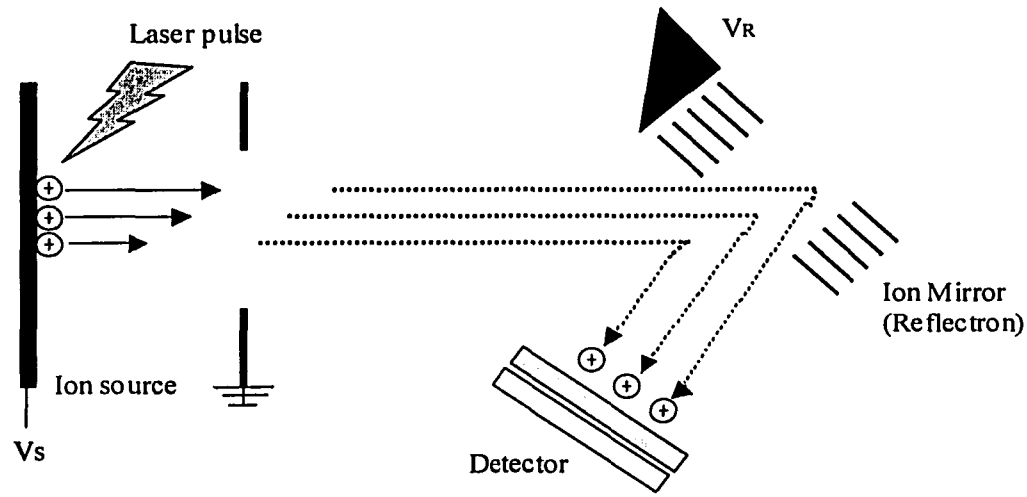

Figure 1.6 A schematic diagram of a reflectron time of flight (TOF) mass analyzer. Vs and V_R are voltages applied on the repeller and ion mirrors, respectively.

Fourier Transform Ion Cyclotron Resonance (FTICR) Mass Analyzer. The Fourier-transform ion cyclotron resonance mass spectrometer (FTICR-MS) [24, 25] offers two distinct advantages: high resolution and the ability to perform tandem mass spectrometry experiments. First introduced in 1974 by Comisarow and Marshall, FTICR-MS is based on the principle of a charged particle orbiting in the presence of a magnetic field. While the ions are orbiting, a radio frequency (RF) signal is used to excite them and as a result of this RF excitation, the ions produce a detectable image current on the cell in which they are trapped. The time-dependent image current can then be Fourier-transformed to obtain the component frequencies of the different ions which correspond to their m/z.

Tandem Mass Spectrometry. To obtain more information on the molecular ions generated for example, in electrospray ionization and MALDI ionization sources, it has been necessary to apply techniques, such as tandem mass spectrometry (MS/MS), to induce fragmentation. Tandem mass spectrometry (abbreviated MS^n, where n refers to the number of generations of fragment ions being analyzed) allows one to induce fragmentation and mass

analyze the fragment ions. This is accomplished by collisionally generating fragments from a selected ion and then mass analyzing the fragment ions. Fragmentation can be achieved by inducing ion/molecule collisions by a process known as collision-activated dissociation (CAD). Collision-activated dissociation is accomplished by selecting an ion of interest with a mass analyzer and introducing that ion into a collision cell. The selected ion then collides with a collision gas (typically argon or helium) resulting in fragmentation. The fragments are then analyzed to obtain a fragment ion spectrum.

One example is the quadrupole-time-of-flight tandem mass spectrometer (QqTOF-MS), which is one of the most popular instruments currently used in proteomics research[26]. Its popularity is due to the high sensitivity, mass accuracy and mass resolution of the TOF instrument for both precursor and product ions. QqTOF consist of a mass-resolving quadrupole (Q), an RF-only quadrupole (q), and a reflecting time-of-flight (TOF) mass spectrometer with orthogonal injection of ions. There is an additional RF quadrupole Q0 in which collisional damping and focusing of ions takes place. Therefore, QqTOF consists of three quadrupole and one TOF mass analyzers. Triple Quadrupole (QQQ)[27] mass spectrometry offers the highest absolute sensitivity for targeted compounds that require the measurement of only a few types of ions, however, the Q-TOF may provide better S/N, due to the increased specificity afforded by the higher resolution of Q-TOF systems for certain analytical situations. The first and third quadrupoles in triple quadrupole are mass filters, and the middle one is a collision cell. This allows the study of fragments that are useful in structural studies.

1.1.3 Ion Detection

Once the ion passes through the mass analyzer it is then detected by the ion detector, the final element of the mass spectrometer. The detector allows a mass spectrometer to generate a signal current from incident ions by generating secondary electrons, which are further amplified. Alternatively, some detectors operate by inducing a current generated by a moving charge. Some of the commonly used detectors are: electron multiplier, channel electron multiplier (CEM) or channeltron, microchannel plate multiplier, and scintillation.

An electron multiplier is one of the most common means of detecting ions. It is made up of a series of dynodes maintained at ever increasing potentials. Ions strike the dynode surface, resulting in the emission of electrons. These secondary electrons are then attracted to the next dynode where more secondary electrons are generated, ultimately resulting in a cascade of electrons. Typical amplification or current gain of an electron multiplier is one million.

Microchannel plate (MCP) [28], which consists of two microchannel plates made of a lead-glass inner surface placed in a chevron configuration, is part of continuous dynode multiplier detectors. Each ion that strikes the MCP creates a pulse of electrons at the anode. The pulse is amplified and is used to trigger a timing pulse which is sent to the ion counting device. Typically transient recorders and time-to-digital converters are used to digitize the ion current to form a mass spectrum. To be detected, the large ions generated must be converted into either electrons or low-mass ions at a conversion electrode. These electrons or low-mass ions are then used to start the electron multiplication cascade for amplication. The yield of secondary electrons from the conversion electrode is a function of the velocity of the ions to be detected[29, 30]. To increase the detection sensitivity or increase the yield of

secondary electron generation, the acceleration potential may be increased, or post-acceleration in the flight tube may be applied. This detector is commonly used for TOF and Q-TOF mass analyzers.

The photo multiplier conversion dynode or scintillation[31] counter detector is similar to an electron multiplier where the ions initially strike a dynode, resulting in the emission of electrons. However, with the photo multiplier conversion dynode detector electrons then strike a phosphor screen. The phosphor screen, much like the screen on a television set, releases photons once an electron strikes. These photons are then detected by a photo multiplier, which operates with a cascading action much like an electron multiplier. The primary advantage of the conversion dynode setup is that the photo multiplier tube is sealed in a vacuum (photons pass through sealed glass), unexposed to the internal environment of the mass spectrometer. Thus the possibility of contamination is removed. A five year or greater lifetime is typical and, with sensitivity similar to electron multipliers, photo multiplier conversion dynode detectors are becoming more widely used in mass spectrometers.

1.2 HPLC Separation of Peptides

Proteins of interest to biological researchers are usually part of a very complex mixture of other proteins and molecules that co-exist inside the cellular medium. In biological samples, different proteins tend to be present in widely differing amounts. If such a mixture is ionized, for example using Electrospray or MALDI, the more abundant species have a tendency to drown out signals from less abundant ones. Additionally, even if this effect were negligible, the mass spectrum from a complex mixture would be very difficult to interpret due to the overwhelming number of mixture components that would be present.

This problem is intensified by the fact that enzymatic digestion of a protein gives rise to a large number of peptide products.

Two methods are widely used to fractionate proteins, or their peptide products from an enzymatic digestion. The first method, generally used to fractionate complex whole protein mixtures is Two-Dimensional Gel Electrophoresis. The second method, High Performance Liquid Chromatography is used mainly to fractionate peptides after enzymatic digestion. In some situations, it may be necessary to combine both of these techniques.

2D-gel electrophoresis has become a standard technique in Proteomics for proteome separation. However, there are some drawbacks associated with this method: First of all, 2D-gel electrophoresis is time consuming and it is difficult to obtain reproducible results. Secondly, 2D-gel electrophoresis may not cope with a large number of proteins that might be of particular interest, e.g. acidic, basic, low abundance, or hydrophobic membrane proteins. Furthermore, laborious preparation steps are necessary to determine protein identities with MALDI-TOF.

Multidimensional HPLC[32,33] for separation, coupled with mass spectrometry of proteome samples offers an alternative way to overcome the limitations associated with 2D-gel electrophoresis. A peptide mixture that results from digestion of a protein mixture is fractionated by one or two steps of liquid chromatography. The eluate from the chromatography stage can be either directly introduced to the mass spectrometer through electrospray ionization, or laid down as a series of small spots for later mass analysis using MALDI. The first dimension is usually strong cation exchange chromatography, coupled either off-line or on-line with the second dimension, reversed-phase chromatography.

1.2.1 Strong Cation-Exchange HPLC

Strong cation-exchange chromatography is probably the most useful mode of high-performance ion-exchange chromatography for peptide separations. The major separation mechanism of this mode of high performance liquid chromatography (HPLC) is electrostatic in nature. Ion-exchange packings may also exhibit significant hydrophobic characteristics, giving rise to mixed-mode contributions to solute separations[34]. The utility of strong cation exchange chromatography packings, generally containing sulfonate functionalities, lies in their ability to retain their negatively charged character in the acidic to neutral pH range. At low pH, the side chain carboxyl groups of acidic amino acid residues are protonated, emphasizing any positively charged character of the peptides. Thus, by manipulating the pH of the mobile phase, the net charge of a peptide may be varied. In addition to overall net charge, other factors which may affect the retention behavior of peptides during ion-exchange chromatography include peptide conformation, polypeptide chain length, charge distribution, and charge density. To understand peptide retention behavior during ion-exchange chromatography completely, it is not sufficient merely to demonstrate that these various factors have an effect on peptide retention, it is also necessary to quantitate the relative contribution each factor makes to retention behavior. In ion-exchange methods, salt concentration and pH are typically the most important factors affecting selectivity and resolution.

Off-line fractionation of peptide mixtures using strong cation exchange chromatography is commonly used in proteomic research. Strong cation exchange chromatography is not only used for separation or fractionation of peptide mixtures but is

also used to remove some detergents, such as SDS, that are used to solubilize proteins in the proteome analysis.

1.2.2 Reversed-Phase HPLC

Reversed-phase high performance liquid chromatography (RP-HPLC) has become a widely used, well-established tool for the analysis and purification of biomolecules[35, 36]. The reason for the central role that RP-HPLC now plays in analyzing and purifying proteins and peptides is resolution: RP-HPLC is able to separate polypeptides of nearly identical sequences, not only for small peptides such as those obtained through trypsin digestion, but even for much larger proteins. Polypeptides which differ by a single amino acid residue can often be separated by RP-HPLC. The separation in RP-HPLC relies on the interaction between the analyte (e.g., hydrophobic amino acid residues contained in the peptide) and the "hydrophobic" functions on the bonded phase of the packing surface. Usually, peptides, in order of increasing hydrophobicity, are eluted from the column by a gradient of increasing organic solvent (e.g., acetonitrile) concentration. The bonded phase in an RP column is non-polar or hydrophobic (usually C8 or C18 on silica columns). Less polar (more hydrophobic) analytes are more attracted to the hydrophobic bonded phase and spend longer associated with the bonded phase and are eluted later; the reverse is true for polar analytes.

Reversed-phase liquid chromatography separations can easily be interfaced on-line to electrospray or nanoelectrospray, or off-line to MALDI sources because of compatibility of solvents and lack of salts and detergents[37].

1.3 Protein Identification using Mass Spectrometry

Mass spectrometry (MS) has become a routine technique for protein identification due to its speed and high sensitivity. There are two approaches widely in use for high throughput protein identification: peptide mass fingerprinting and peptide fragmentation fingerprinting. Peptide mass mapping[38,39] involves the digestion of the protein of interest by a proteolytic enzyme, such as trypsin, followed by identification of the resulting peptides using mass spectrometry. This approach, coupled to an understanding of the cleavage process, affords a peptide map that is unique for each protein and allows its identification by searching existing databases. The peptide map can be produced by either MALDI or ESI MS. MALDI-MS is commonly used because of its tolerance to samples with buffers and salts and is more sensitive. Another advantage of MALDI is that it almost exclusively produces singly charged ions for low mass peptides, which makes it more suitable for direct mixture analysis. When the m/z of peptides is the sole parameter in a peptide mass mapping, a sufficient number of proteolytic peptides with highly accurate masses are required to make an unambiguous identification of a protein, especially when dealing with large proteome databases. To some extent, the use of either external or internal mass standards for calibrating peptide masses can increase the peptide mass mapping mass accuracy. A modern MALDI-TOF instrument can routinely provide mass accuracy of better than 50 ppm with external mass calibration and better than 20 ppm with internal calibration. Since peptide mass mapping is often not sufficient for unambiguous identification of proteins, other methods are required for confirmation.

The other method for protein identification relies on the fragmentation of peptides in mixtures to obtain sequence information using tandem mass spectrometry[40-43]. The

fragmentations are produced either in a collision cell in a tandem mass spectrometer such as MALDI/ESI Qq-TOF, or within an ESI-ion trap. In an MS/MS based protein identification experiment multiple peptides are usually found and all of their fragment spectra are used to correlate to a protein. The types of fragment ions observed in an MS/MS spectrum depend on many factors, including: primary sequence, the amount of internal energy, how the energy was introduced, and charge state. In tandem mass spectrometry, two fragmentation-energy regimes of either high or low energy collisionally induced dissociation (CID) are used to generate fragmentation. Low energy CID is commonly used by most instruments (triple quadrupole, ion trap and quadrupole-TOF). The nomenclature shown in Figure 1.7 was first developed by Roepstorff and Fohlman (1984)[44] and later modified by Biemann (1990)[45]. Fragments will only be detected if they carry at least one charge. If this charge is retained on the N- terminal fragment, the ion is classified as either a, b or c and if the charge is retained on the C- terminal, the ion type is either x, y or z. A subscript indicates the number of residues in the fragment. Among the series of ion types mentioned, b and y-type ions are often observed in low energy CID.

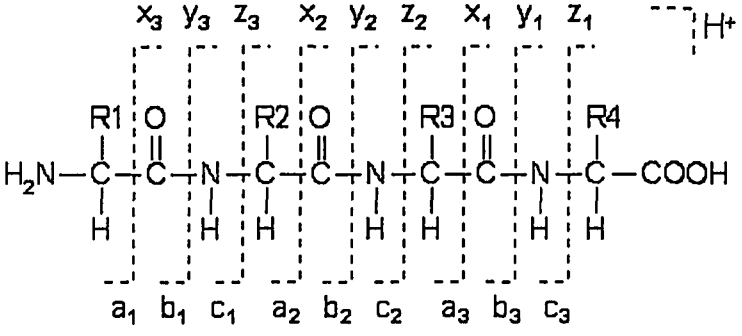

Figure 1.7 Nomenclature of peptide fragmentation pattern under low energy CID.

If both peptide mass mapping and peptide fragment ion fingerprinting fail to identify the protein, de novo sequencing of peptides can be carried out by using tandem MS. This method involves manual interpretation of the fragment ion spectra to determine the amino acid sequence.

1.4 Scope of the Thesis Work

In this thesis, proteomic analysis was applied to heart tissues and cancer cell lines. Full characterization of the human heart proteome is essential in understanding the protein expression alterations in failing heart tissues. Shotgun proteomics, which is based on the peptide analysis, has emerged as a powerful technique for heart tissue proteome identification. One critical step in applying the shotgun proteomics approach to generate a comprehensive proteome profile of tissues is to digest all the proteins, including soluble and insoluble proteins. In chapter 2, two dimensional liquid chromatography coupled with tandem mass spectrometry (2D LC-MS/MS) and microwave-assisted acid hydrolysis (MAAH) coupled with 2D LC-MALDI MS/MS techniques for more comprehensive coverage of proteome profiles from diseased human heart tissues are presented. Two steps were used to analyze extracted proteins. SDS soluble proteins were digested by trypsin, and the resulting peptide mixture was fractionated by strong cation exchange (SCX) chromatography into 10 fractions. Each fraction was further analyzed by an LCQ DECA tandem mass spectrometer. SDS-insoluble proteins were suspended in 25% TFA and digested with microwave irradiation for 10 minutes. The digests were fractionated by SCX chromatography into 9 fractions. Each fraction was analyzed by LC-MALDI MS/MS. From

these combined methods, a significant number of proteins were obtained and presented in this thesis.

Relative quantification of proteins is another focus of today's proteomic research. In chapter 3, the application of an isotope labeling strategy coupled with liquid chromatography laser desorption/ionization mass spectrometry (LC-MALDI) to quantify and identify differentially expressed proteins between E-cadherin-deficient human carcinoma cell line (SCC9) and its transfectants expressing E-cadherin (SCC9-E-cad) and plakoglobin (Pg-SCC9) is presented. This labeling strategy, known as stable dimethyl isotope labeling, labels the amino groups of N-termini and lysine to determine the proteins that are differentially expressed between two samples. Stable isotope dimethyl labeling strategy, similar to that used for squamous carcinoma cell line extracts, but with a slightly different protocol, was applied to quantify and identify heart tissue extracts.

1.5 Literature Cited

1. Dole, M. *J. Chem. Phys*, **1968**, 49, 2240-2249.
2. Yamashita, M.; Fenn, J.B. *J. Phys. Chem.* **1984**, 88, 4451-4459.
3. Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Anal. Chem.* **1985**, 57, 675 – 679.
4. Kebarle, P.; Tang, L. Anal. Chem. **1993**, 65, 972A-986A.
5. Karas, M.; Bachmann, D.; Bahr, Y.: Hillenkamp, F. *Int. J. Mass Spectrom. Ion Process*, **1987**, 78, 53.
6. Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, 60, 2299-2301.

7. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, 2, 151.

8. Bokelmann, V.; Spengler, B.; Kaufmann, R. *Eur. Mass Spectrom.* **1995**, 1, 81-93.

9. Weinberger, S. R.; Boernsen, K. O.; Finchy, J. W.; Robertson, V.; Musselman, B. D. In proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics; San Francisco, CA, May 31-June 4, 1993; PP 775a-b.

10. Xiang, F.; Beavis, R. C. *Rapid Commun. Mass Spectrom.* **1994**, 8, 199-204.

11. Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994**, 66, 3281-3287.

12. Edmondson, R. D.; Campo, K. K; Russell, D. H. In Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics; Atlanta, Georgia, May 21- May 26, 1995; p 1246

13. Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1996**, 68, 2721-2725.

14. Paul, W. *Angew. Chem. Int. Ed. Engl.* **1990**, 29, 739.

15. March, R. E. *J. Mass Spectrom* **1997**, 32, 351-369.

16. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, 63, 1193A-1203A.

17. Cotter, R. J. Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research; American Chemical Society: Washington, D. C., **1997.**

18. Wiley, W. C.; McLaren, I. H. *Rev. Sci. Instrum.* **1995**, 26, 1150-1157.

19. Colby, S. M.; King, T. B.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1994**, 8, 865-868.

20. Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, 9, 1044-1050.

21. Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, 67, 1998-2003.

22. Whittal, R. M.; Russon, L. M.; Weinberger, S. R.; Li, L. *Anal. Chem.* **1997**, 69, 2147-2153.

23. Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, 67, 1950-1954.

24. Buchanan, M.V.; Hettich, R. L. *Anal. Chem.* **1993**, 65, 245A-259A.

25. Wilkins, C. L.; Gross, M. L. *Anal. Chem.* **1981**, 53, 1661A-1676A.

26. Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **2000**, 14, 1047-1057.

27. Yost, R. A.; Enke, C.G. *J. Am. Chem. Soc.* **1978**, 100, 2274-2275.

28. Wurz, P.; Gubler, L. *Rev. Sci. Instrum.* **1996**, 67, 1790-1793.

29. Beuhler, R. J. *J. Appl. Phys.* **1983**, 54, 4118-4126.

30. Spengler, B.; Kirsch, D.; Kaufmann, R.; Karas, M.; Hillenkamp, F.; Giessmann, U. *Rapid Commun. Mass Spectrom.* **1990**, 4, 301-305

31. Daly, N. R., *Rev. Sci. Instrum.* 1963, 34, 1116-1120.

32. Wolters, D. A.; Washburn, M. P.; Yates, J. R. III. *Anal. Chem.* **2001**, 73, 5683-5690.

33. Washburn, M. P.; Woters, D.; Yates, J. R. III. Nat. Biotechnol. **2001**, 9, 242-247.

34. Lorne Burke, T. W.; Mant, C. T., Black, J.A.; Hodges, R. S. *J. Chromatogr.* **1989**, 476, 377-389.

35. Yates, J. III. *J. Mass Spectrom.* **1998**, 33, 1-19.

36. Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, 70, 437-473.

37. Le Bihan, T.; Duewel, H. S; Figeys, D., *J. Am. Soc. Mass Spectrom.* **2003**, 14, 719-727.

38. Gibson, B. W.; Biemann, K. Proc. Natl. Acad. Sci. **1984**, 81, 1956-1960.

39. Greer, F. M.; Morris, H. R.; Forstrom, J.; Lyons, D. Biomed. Environ. *Mass Spectrom.* **1988**, 16, 191-195.

40. Mclafferty, F. W., Ed.; Tandem Mass Spectrometry, Wiley-Interscience: New York, **1983.**

41. Biemann, K.; Scoble, H. A. *Science*, **1987**, 237, 992-998.

42. Biemann, K. Biomed. *Environ, Mass Spectrom.* **1988**, 16, 99-111.

43. Yates, J. R., II; Specicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, 214, 397-408.

44. Roepstorff, P.; Fohlman, *J. Biomed. Mass Spectrom.* **1984**, 11, 601.

45. Biemann, K. *Methods Enzymol.* **1990**, 193, 886-887

Chapter 2

Comprehensive Proteome Analysis of Diseased Human Heart Tissue

2.1 Introduction

Full characterization of the human heart proteome is essential in understanding the protein expression alterations in failing heart tissues. The causes of most heart diseases that result in heart failure are not well known, but are likely due to protein and gene expression alterations [1-4]. In order to compare the normal heart tissue with that of diseased heart tissue, the complete proteome profile of the healthy heart tissue is required. There are a number of alternative methods to separate and identify complex protein mixtures. One challenging step in most of these techniques is to extract and digest all the extracted proteins in order to identify all the protein components in the heart tissue.

Effective protein isolation methods, better protein separation and high-throughput identification techniques need to be established in order to identify all possible proteins in heart tissue. Two-dimensional gel electrophoresis (2DE) in combination with mass spectrometry has been used to separate and identify proteins from human heart tissues [5]. About three thousand spots were detected using 2DE from human heart tissue protein extracts, but only two hundred proteins were reported from these spots [5-8]. As a result of a number of limitations in 2DE, some alternative technologies have been introduced. Recently, shotgun proteomics, which is based on peptide analysis [9-10], has emerged as a powerful technique for heart tissue proteome identification [11]. One critical step in applying the shotgun proteomics approach to generate a comprehensive proteome profile of tissues is to digest all proteins including soluble and insoluble proteins. In most cases, protein pellets

extracted from cells or tissues are not easily solubilized in most common solvents or buffers. Some detergents, such as SDS, can be used in solublizing such extracts, but a very high concentration of SDS can not be used as it can affect trypsin digestion, complicates the protein purification using the 2D liquid chromatography and also interferes with the mass spectrometry.

A new protein digestion method, microwave-assisted acid hydrolysis[12], developed by our group, was used to analyze membrane proteins. This method has become a promising technique to solve some of the problems encountered in shotgun proteomics techniques. MAAH is detergent-free and digests proteins in the absence of enzymes. SDS-insoluble proteins can be easily digested using this technique.

A method is presented to analyze the human heart tissue proteome using a combination of traditional shotgun proteomics techniques and MAAH for more comprehensive coverage of proteome profiles from diseased human heart tissues.

2.2 Experimental Section

2.2.1 Materials and Reagents

Trizol was purchased from Invitrogen Canada (Burlington, ON, Canada). 2, 5-dihydroxybenzoic acid (DHB), bovine trypsin, dithiothreitol (DTT), iodoacetamide, trifluoroacetic acid (TFA), chloroform and sodium dodecylsulfate (SDS) were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). HPLC grade acetonitrile (ACN), acetic acid, isopropanol and acetone were from Fisher Scientific Canada (Edmonton, Canada). Water was obtained from a Milli-Q plus purification system (Millipore, Bedford, MA).

2.2.2 Extraction of proteins

Human heart tissue was obtained from a person with heart disease and provided by Dr. Shaohua Wang (Cardiac Surgery Division, University of Alberta Hospital). The tissue was frozen in liquid nitrogen and stored at -80 ^{o}C. The thawed tissue was cut into pieces and homogenized in Trizol (60-100 mg of tissue/1 mL of Trizol). Samples were mixed at room temperature overnight. The debris was removed by centrifugation at 12,000 x g for 10 minutes at 4 ^{o}C. Chloroform was added to separate the solution into an aqueous phase and an organic phase. The aqueous phase was carefully removed after centrifugation at 12, 000 x g for 15 minutes at 4 ^{o}C. Isopropanol was added at the last step to precipitate the protein from the organic phase. Detailed instruction was provided with the Trizol reagent and can be obtained in the manufacturer's manual. The protein pellet was then solubilized in 1% SDS. The SDS-insoluble protein pellet was kept at -80 ^{o}C. Protein concentration was estimated using the Bradford assay and approximately 6.5 μg total protein/mg of tissue was extracted.

2.2.3 In-Solution Tryptic Digestion of SDS Soluble Proteins

The protein solution in 1% SDS was reduced with dithiothreitol (DTT) and then alklyated with iodoacetamide followed by acetone precipitation at -20 ^{o}C. The protein pellet was resuspended in 0.1% SDS. Solubilization of the proteins was achieved with 0.2 M NaOH [13]. The mixture was diluted to a final SDS concentration of 0.05% and the pH was adjusted with 1mM $NaHCO_3$ to ~ 8.5. 20 mM $CaCl_2$ was added to the mixture to a final concentration of 2 mM $CaCl_2$. Finally, trypsin was added to the protein solution at a protein: enzyme ratio of 40:1 by weight and the solution was incubated at 37 ^{o}C overnight.

2.2.4 Acid Hydrolysis of the SDS-insoluble Protein Pellet

The SDS-insoluble protein pellet was washed three times with water and the supernatant was removed by a pipette. The sample was suspended in a 25% TFA solution and then divided into three polypropylene centrifuge tubes. Each tube was capped and sealed with Teflon tape and placed in a domestic 900W (2450 MHz) microwave oven. In order to absorb extra microwave energy, water in a loosely covered container was placed beside the sample vials. The sample was then exposed to microwave irradiation for 10 minutes. The entire sample was collected into one vial and dried in a vacuum centrifuge to remove the acid. The dried sample was dissolved in 0.1% TFA.

2.2.5 2D LC-ESI MS/MS

The tryptic peptides were acidified with 0.1% TFA, supplemented with 20% ACN, and then loaded onto a strong cation exchange column (Biochrom, Terre Haute, IN, 2.1 x 150 mm). Fractions were collected using a fraction collector (Agilent) every one minute during a 40 min gradient from 2% to 80% solvent B (solvent A: 0.1% TFA and 20% ACN in water; B: 1M NaCl in 0.1% TFA and 20% ACN) at flow rate of 0.25 mL/min. Ten fractions were collected in 40 minutes. All collected fractions were reduced in volume to ~ 0.01 mL by vacuum centrifugation. Each fraction was injected onto a capillary C18 column (Vydac, 1 x 150 mm) with a gradient from 5 to 80% solvent B (solvent A: 0.1% acetic acid in water; 0.1% acetic acid in ACN). The LC-ESI MS/MS system used was a ThermoFinnigan LCQ DECA ion trap mass spectrometer (San Jose, CA). The mass spectrometer was operated in six segment data dependent mode (one mass spectrum followed by five CID spectra).

2.2.6 2D LC-MALDI MS/MS

ACN was added to the SDS-insoluble 0.1%TFA peptide solution to a final 20% ACN concentration and loaded onto a strong cation exchange column (2.1 x 150 mm). Nine fractions were collected during solvent gradient development, as described in the 2LC-ESI MS/MS section above. All collected fractions were reduced in volume to ~0.02 ml to remove ACN. Each fraction was diluted to ~ 0.1mL with 0.1% TFA. 0.1mL of each fraction was injected onto a capillary C18 column (Vydac, 1.0 x 150 mm). Gradient elution was performed with solvent A (0.1% TFA and 4% ACN in water) and B (0.1% TFA in ACN) at a flow rate of 40 μL/min. Each fraction was collected every one minute and directly deposited onto a 100 well gold MALDI plate by the heated droplet interface. 1M DHB in 50% acetonitrile was used as a matrix. Peptides were analyzed by a QSTAR MALDI MS/MS mass spectrometer (MDS SCIEX, Concord, Canada).

2.2.7 Data Processing

All MS/MS spectra were searched using the MASCOT algorithm (http://www.matrixscience.com). For the tryptic digest part, over 20,000 MS/MS spectra were searched against SwissProt with variable modification of methionine and cysteine. For the SDS-insoluble acid hydrolysis, the collected MS/MS spectra were searched against the database using the following criteria: no specification of enzyme type, variable modification of methionine and asparagine or glutamine (deamidation). The MS/MS spectra of the matched peptides were examined manually.

2.2.8 Safety Considerations

Procedures and cautions in handling liquid samples for microwave experiments can be found in the literature [14]. Although the sample volume in each vial used in our experiment was about 33 μL, care was taken during the microwave irradiation. To avoid any problems associated with the high concentration of the acid used, the capped sample vial was allowed to cool inside the microwave and opened under a fume-hood.

2.3 Results and Discussion

A flowchart of the experiment is presented in Figure 2.1. The entire extracted protein pellets were solublized in 1% SDS. However, it was observed that there was a substantial amount of protein pellet that was very difficult to dissolve in 1% SDS. Therefore, proteins were analyzed by two steps: SDS-soluble protein analysis and SDS-insoluble protein analysis.

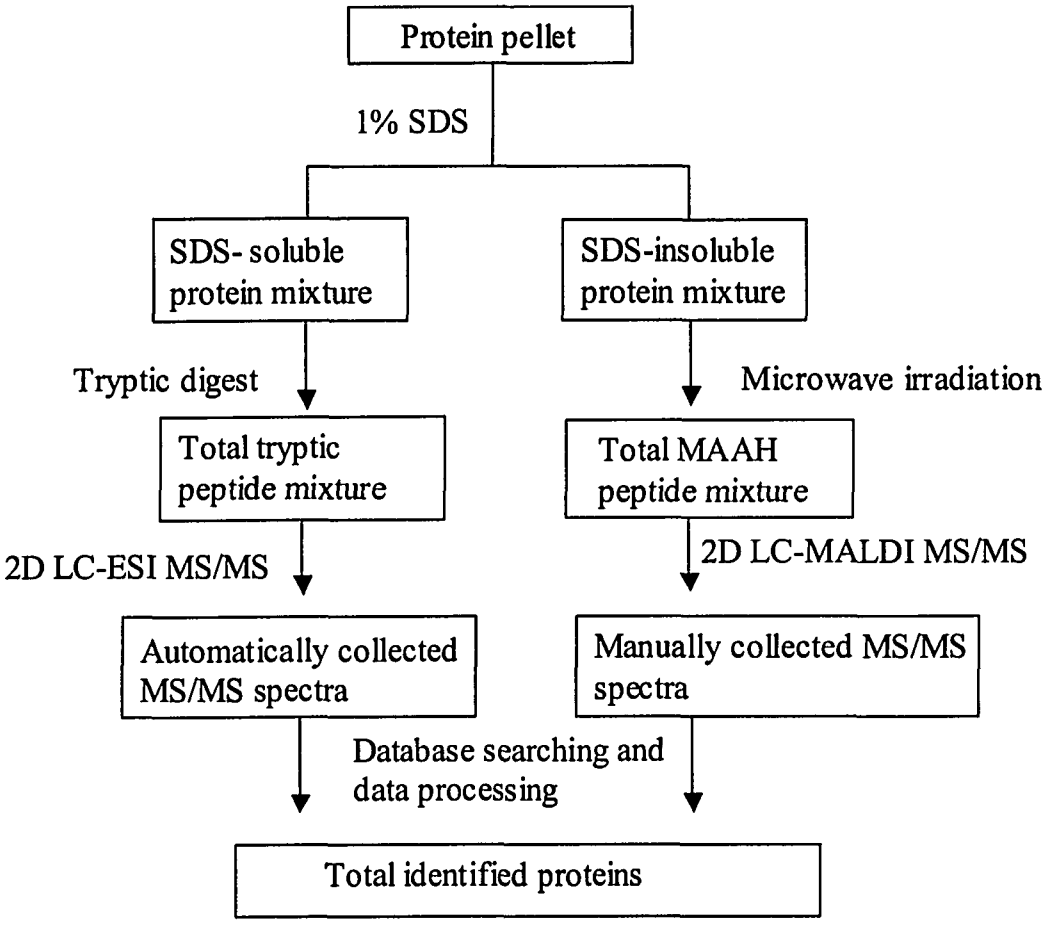

Figure 2.1 A flow Chart of the experiment.

2.3.1 SDS-soluble Protein Analysis

The 1% SDS-soluble proteins were reduced and alkylated, followed by acetone precipitation. The protein pellet was then resuspended in 0.1% SDS. The protein pellet was not completely solubilized in 0.1% SDS. Therefore, 0.2 M NaOH was used for solubilization. The sample was trypsinized into peptides in the presence of SDS (0.05%). The resulting peptides were separated by strong cation exchange (SCX) chromatography. A total of 10 fractions were collected every one minute using a fraction collector. The UV-

absorbance chromatogram of the cation exchange separation of the tryptic digest from the SDS-solublized protein mixture is shown in Figure 2.2A

Each fraction was injected manually onto a reversed phase column and analyzed by LC-ESI MS/MS. The LC-ESI base peaks of the first three fractions are shown in Figure 2.3. Overall, more than 20, 000 MS/MS spectra were automatically collected and searched against the database.

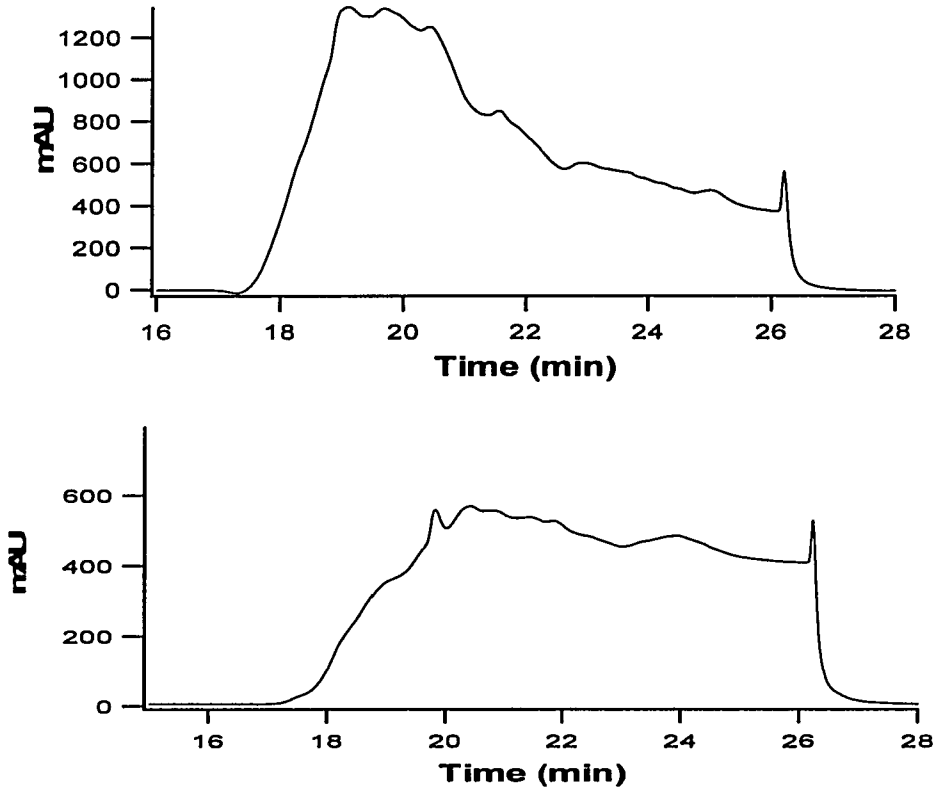

Figure 2.2 UV chromatogram of cation exchange separation of the tryptic digests (A) and MAAH (B) of protein mixtures.

The most abundant proteins, such as myosin (light and heavy chains), serum albumin, tropomysin (alpha and beta) were observed in almost all fractions. However, the peptides that match to a single protein were obtained in different fractions, indicating the efficiency of

the fractionation. The most important proteins, low abundance proteins, were the focus of our study. Overall, 236 proteins were identified using 2D LC-ESI MS/MS. Of these proteins, 21% are membrane and membrane-associated proteins.

2.3.2 Microwave-Assisted Acid Hydrolysis

The SDS-insoluble proteins were suspended in 25% TFA and digested with microwave irradiation for 10 minutes. The acid was removed by vacuum centrifugation and the residue was dissolved in 0.1% TFA. The digests were fractionated by SCX into 9 fractions. As shown in Figure 2.2B, the intensity of the UV-absorbance chromatogram of the SCX was similar to that obtained from the soluble protein pellets, indicating that a large number of peptides were generated by acid hydrolysis of the insoluble pellets. Each fraction was further separated by reversed-phase chromatography and deposited onto a 100 well gold MALDI plate by the heated droplet interface method [15]. 1μL of 1M DHB in 50% acetonitrile was deposited onto each sample.

A MALDI Qq-TOF instrument was used to obtain both MS and MS/MS spectra as this instrument has high mass accuracy and resolution. Each MS/MS spectrum was collected manually using MALDI Qq-TOF and searched against the database.

The peptides generated by MAAH were from the C- and N-termini as well as peptides from internal fragmentation. A single or multiple amino acid cleavage from a single peptide is the other feature of this method. Although other amino acid cleavages are observed, cleavage at glycine is the most common. Glycine cleavage is easily identified from the MALDI MS spectra from the 57 Da mass differences.

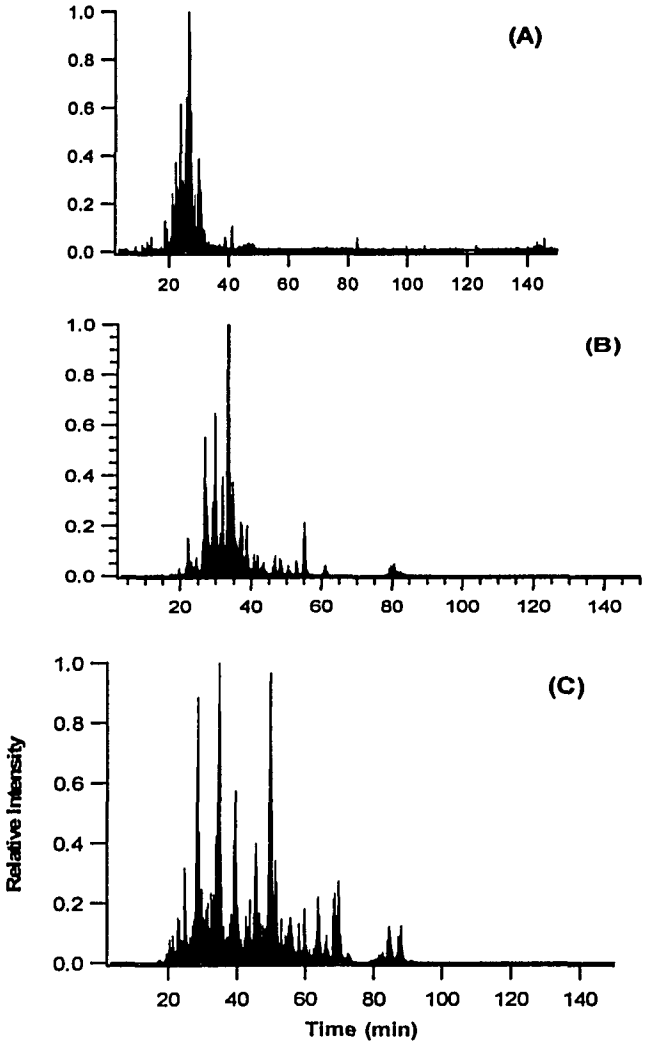

Figure 2.3 LC-ESI base peaks for fraction 1(A), 2(B) and 3(C), respectively.

Figure 2.4B is a MALDI MS spectrum that shows the glycine cleavage for peptides with mass of 1405.79 and 1462.79 Da eluted at 50 minutes from fraction 3 (Figure 2.4A). Two peptide sequences, AHYKLVPQQLAH and GAHYKLVPQQLAH, were identified based on the two MS/MS spectra shown in Figure 2.5A and B respectively and represent a protein with ID of P10606. Peptide pairs, such as APSRKFFGG and APSRKFFG, represent a protein with ID of P60174, were eluted in two consecutive spots at 22 and 23 min respectively in fraction 3 (not shown). KHSLPDLPYDYG and KHSLPDLPYDY,

representing a protein with ID of P04179, were observed at 28 and 30 min, respectively in fraction 3. PIISDRHGGYKPTD and PIISDRHGGYKPT, representing a protein with ID of P06732, were observed in the same spot at 27 min in fraction 3. CTGAHERTF and ACTGAHERTF, representing a protein with ID of Q13232, were also observed at 25 and 26 min, respectively, in fraction 3. These observations generally indicate that two peptides with a single amino acid difference have similar chromatographic retention times and can be easily identified from the MALDI MS spectra.

For demonstration purposes, three membrane proteins which were obtained only from the undissolved protein pellet have been selected. Protein Q8TB96 matched three peptides and an MS/MS spectrum of one peptide is shown in Figure 2.6A. Protein Q96F46 matched two peptides and an MS/MS spectrum of one peptide is shown in Figure 2.6B. Protein Q9Y4D7 matched two peptides and an MS/MS spectrum of one peptide is shown in Figure 2.6C.

Overall, 148 proteins were identified using this method. Of these 148 proteins, 121 were not identified by 2D LC-ESI MS/MS. These proteins could then be lost due to the lack of suitable solubilizing solvents or detergents. Therefore, microwave-assisted acid hydrolysis was demonstrated to be a promising technique to analyze very hydrophobic or proteins that are insoluble in detergents, such as SDS. The comparison of the two techniques, as applied to heart tissue proteins, is presented below. About 25% of the identified proteins are membrane or membrane-associated proteins.

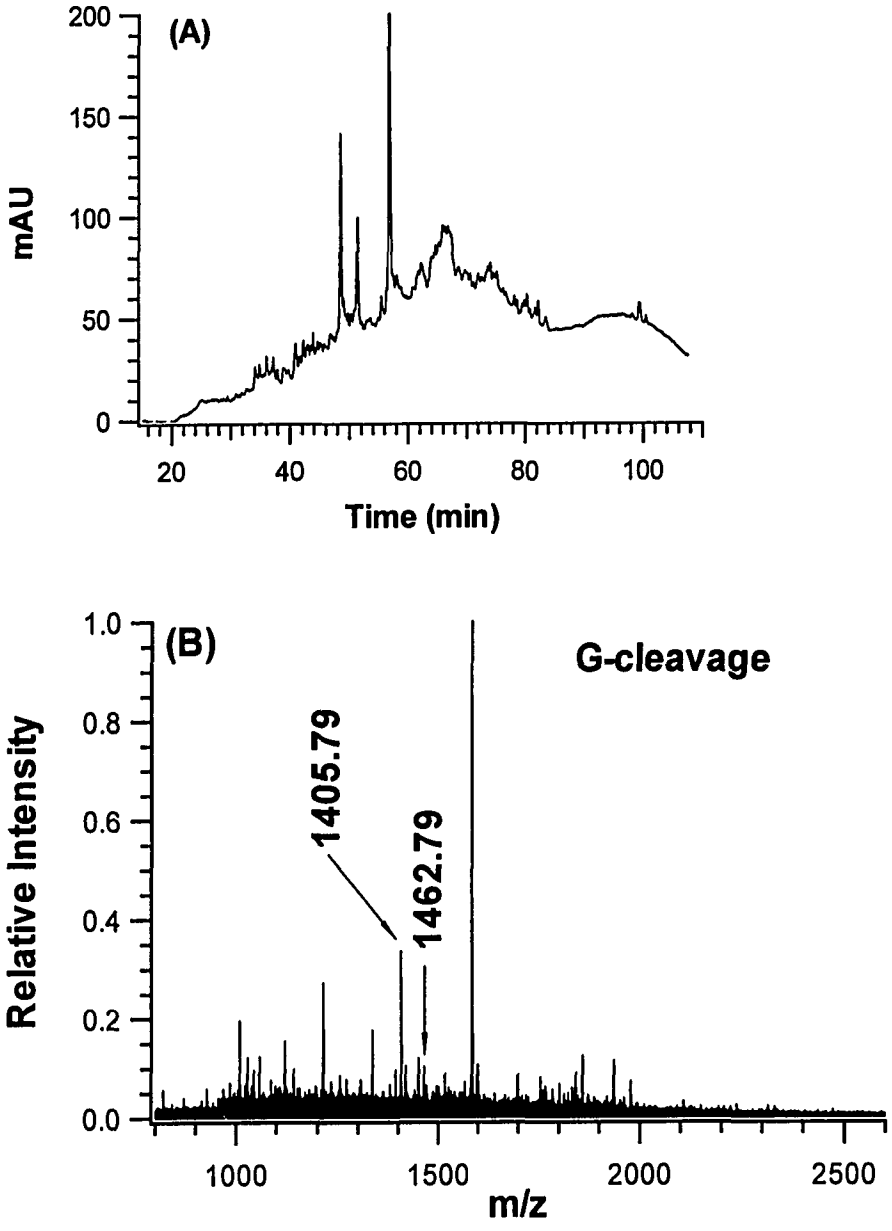

Figure 2.4 **(A)** Reversed phase HPLC chromatogram of hydrolysates for fraction 3 , (B) and LC-MALDI MS spectrum showing glycine cleavage obtained from HPLC fraction 3 (figure 2.4A) for peptides with mass 1405.79 and 1462.79 Da,.

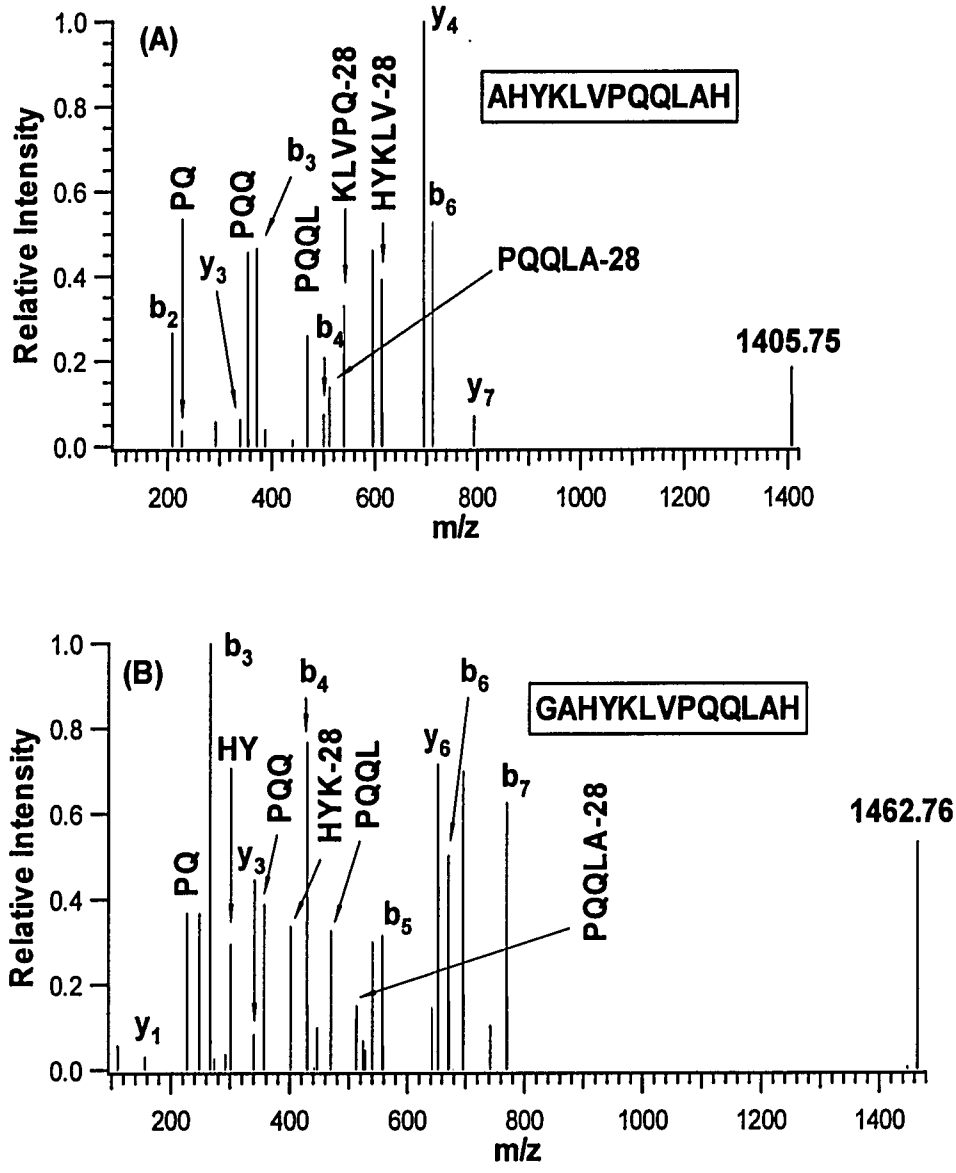

Figure 2.5 MALDI MS/MS spectra of peptides (A) 1405.79 Da and (B) 1462.79 Da

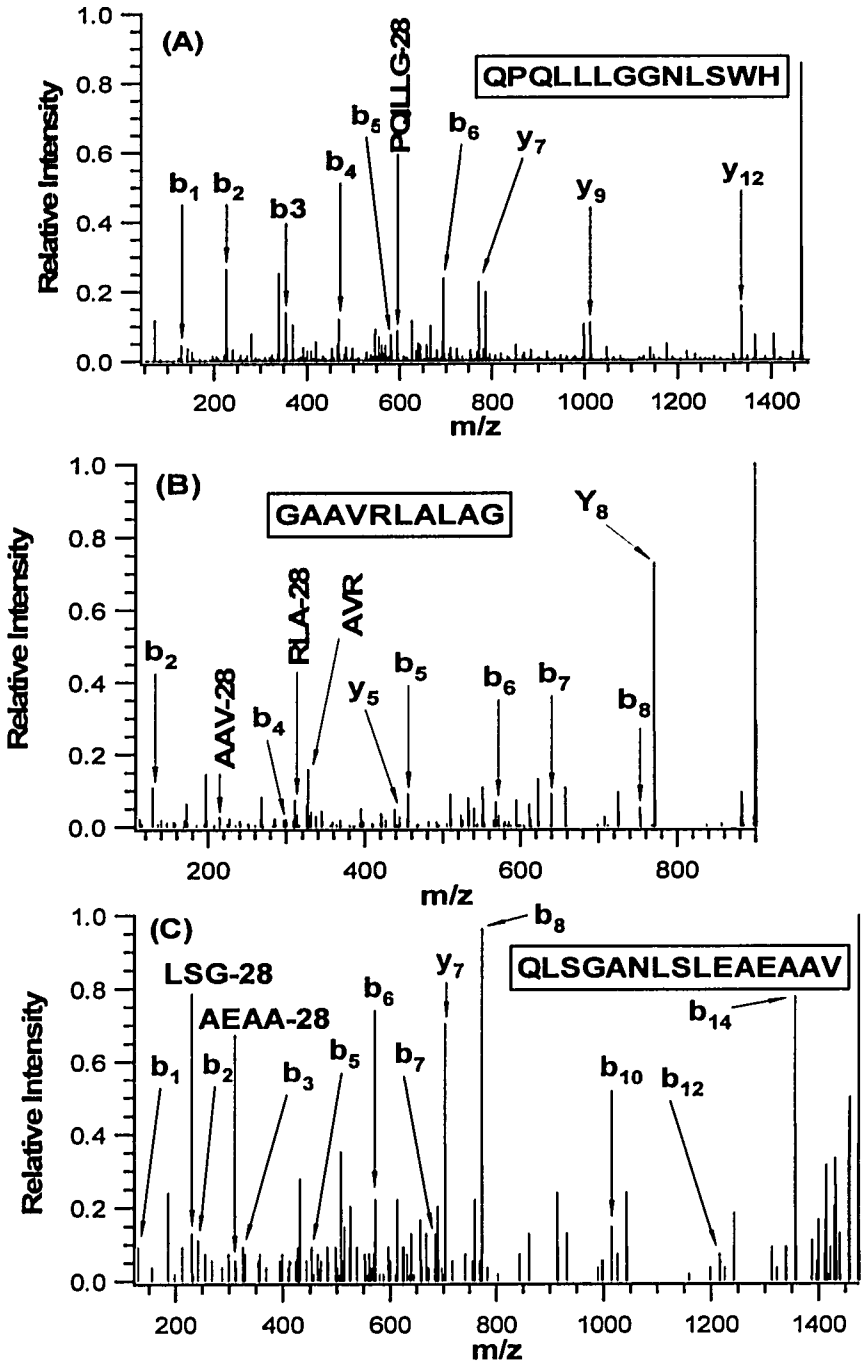

Figure 2.6 MALDI MS/MS spectra of selected membrane proteins: Three peptides that represent three proteins with ID of Q8TB96 (A), Q96F46 (B) and Q9Y4D7 (C), respectively.

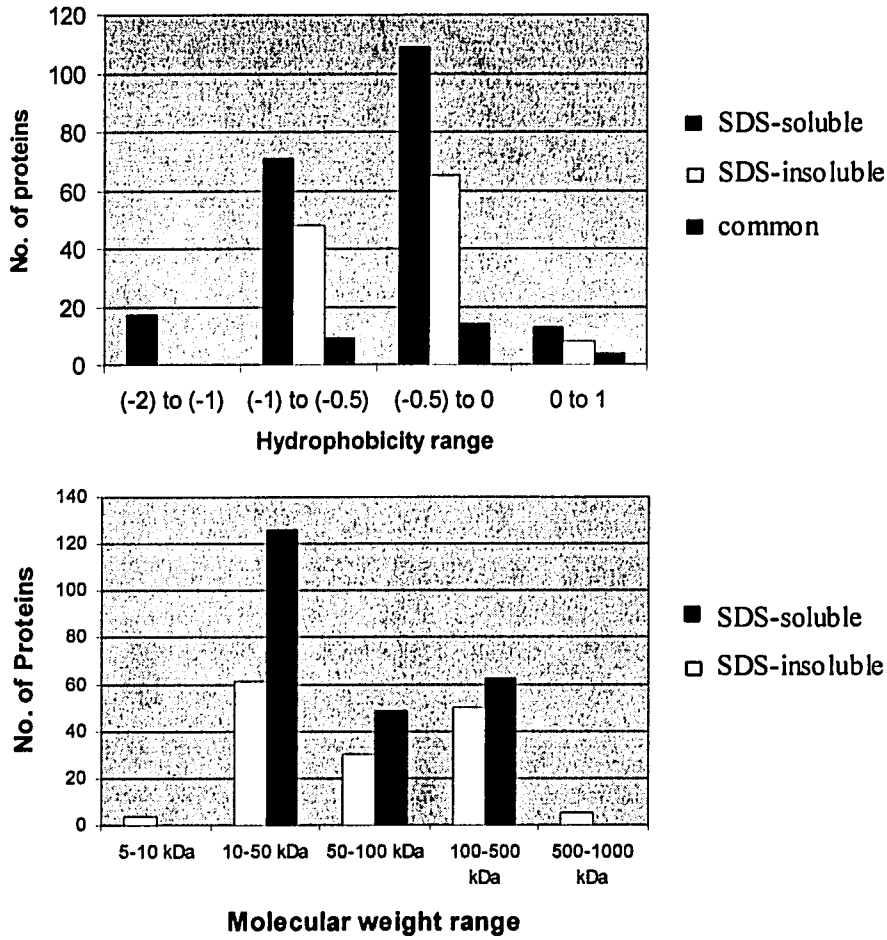

Figure 2.7 histograms showing the hydrophobicitiy ranges for all proteins identified from SDS-soluble, SDS-insoluble and some common proteins (A) and molecular weight ranges for all proteins identified from SDS-soluble and SDS-insoluble mixtures (B).

2.3.3 Tryptic Digestion versus Microwave-Assisted Acid Hydrolysis

MAAH is much faster than in-solution tryptic digestion. Only 10 minutes were required to generate peptide mixtures. By comparison, in-solution tryptic digestion is usually performed overnight. The proteins identified using MAAH followed by 2D-LC MALDI Qq-TOF were 25% membrane or membrane-associated proteins.

The hydrophobicities of all identified proteins were estimated with the ProtParam program available at the EXPASY web site (http: us.expasy.org/tools/protparam.html). The

program calculates the grand average of hydrophobicity (GRAVY). Positive values are considered as hydrophobic and negative values are considered as hydrophilic. For comparison, the plot of number of proteins detected in SDS-soluble, SDS-insoluble and common proteins found in both mixtures against hydrophobicitiy is shown in Figure 2.7A.

Significantly lower and higher molecular weight proteins were obtained from the undissolved protein pellet than the SDS-soluble protein pellet. The smallest molecular weight protein obtained from the undissolved protein pellet was 7 kDa. However, the smallest molecular weight protein obtained from the SDS-soluble protein pellet was 10 kDa. Proteins as high as 1011 kDa (not shown in the histogram) were obtained from the undissolved protein pellet, but the highest molecular weight obtained from the SDS-soluble protein pellet was 479 kDa. A plot of number of proteins against molecular weight is given in Figure 2.7B for all proteins identified using LC-ESI MS/MS and LC-MALDI MS/MS. The undissolved protein pellet was washed three times with water to remove salts, detergents and soluble proteins. However, as shown in Figure 2.8A, this result shows that 27 proteins were obtained in both fractions.

Although extremely hydrophobic proteins were expected in the undissolved protein pellet, the identified proteins were not all very hydrophobic. One possible reason is that more washing could be used to make sure that all soluble proteins are removed. The other reason could be related to the aggregation of the proteins after the acetone precipitation which can complicate the digestion of the proteins with trypsin and this suggests that MAAH can solve such problems. However, lower and higher molecular weight proteins were obtained from the undissolved protein pellet than the SDS-soluble protein pellet.

The subcellular locations of all the proteins are presented in Figure 2.8B. The list of all detected proteins using 2D-LC-ESI MS/MS and 2D-LC-MALDI MS/MS are shown in Table 2.1. The list of the common proteins which were obtained in both the tryptic and MAAH digests are shown in Table. 2.2.

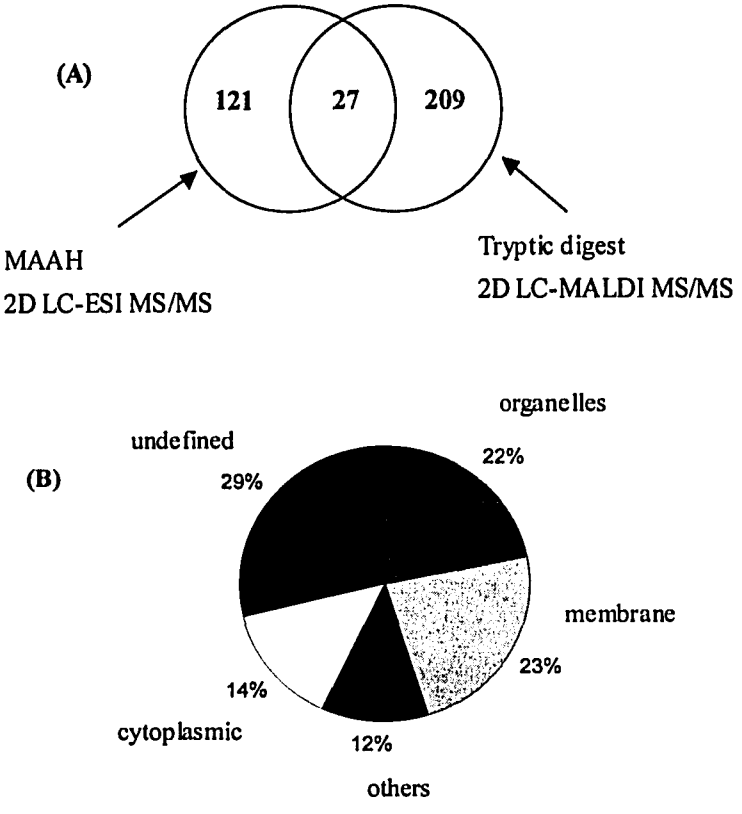

Figure 2.8 Schematic graph showing: (A) the total number and overlap of proteins identified from the in-solution tryptic digestion coupled with 2D LC-ESI MS/MS and microwave-assisted acid hydrolysis coupled with 2D LC-MALDI MS/MS and (B) subcellular location of all the proteins identified: 22% were from organelles (mitochondrail, nuclear, ribosomal, endoplasmic recticulum, Golgi, lysosomal), 14% were cytoplasmic, 23% were membrane proteins (including integral, membrane-associated, type I, II and IV membrane proteins), 12% were classified as others (secreted, extracellular, interacellular, fibers, filaments and with multiple location). The location of 29% of proteins was not defined.

Table 2.1. All Identified Proteins Using 2DLC-ESI-MS/MS and 2DLC-MALDI MS/MS

Access ID	Unique Peptide Sequence	Score	MW (kDa)	Subcellular Location	Soluble in 1% SDS?
P02768	AACLLPK	35	69.32	Secreted	Y
	AEFAEVSK	45			
	LCTVATLR	37			
	DDNPNLPR	34			
	FQNALLVR	57			
	QTALVELVK	41			
	TYETTLEK	30			
	LDELRDEGK	42			
	CCTESLVNR	51			
	LVNEVTEFAK	47			
	FKDLGEENFK	49			
	AVMDDFAAFVEK	85			
	AAFTECCQAADK	63			
	ETYGEMADCCAK	53			
	YICENQDSISSK	73			
	TCVADESAENCDK	110			
	ADDKETCFAEEGK	36			
	CCAAADPHECYAK	44			
	ADDKETCFAEEGKK	47			
	KVPQVSTPTLVEVSR	79			
	QNCELFEQLGEYK	86			
	QEPERNECFLQHK	45			
	VFDEFKPLVEEPQNLIK	47			
	SHCIAEVENDEMPADLPSLAA DFVESK	57			
	EFNAETFTFHADICTLSEKER	45			
	HPYFYAPELLFFAK	33			
	KQTALVELVK	56			
	LKECCEKPLLEK	85			
	MPCAEDYLSVVLNQLCVLHEK	49			
	RPCFSALEVDETYVPK	30			
P17661	FANYIEK	46	53.37	Cytoplasmic	Y
	GTNDSLMR	33			
	LLEGEESR	37			
	AQYETIAAK	41			
	ADVDAATLAR	63			
	TSGGAGGLGSLR	82			
	VSDLTQAANK	71			
	KLLEGEESR	40			
	QVEVLTNQR	38			
	DNLLDDLQR	35			
	EYQDLLNVK	55			
	VAELYEEELR	52			
	IESLNEEIAFLK	31			
	FASEASGYQDNIAR	74			
	TNEKVELQELNDR	64			
	INLPIQTYSALNFR	47			

	FLEQQNAALAAEVNR	98			
	TFGGAPGFPLGSPLSSPVFPR	37			
P13533	ADIAESQVNK	30	223.55	Thick filaments	Y
	AGLLGLLEEMRDER	38			
	ANSEVAQWR	37			
	AQLEFNQIK	37			
	AVVEQTER	26			
	DIDDLELTLAK	60			
	DIDDLELTLAKVEK	43			
	DTQIQLDDAVR	23			
	GTLEDQIIQANPALEAFGNAK	70			
	HADSVAELGEQIDNLQR	36			
	KHADSVAELGEQIDNLQR	43			
	KLEGDLK	33			
	LELDDVTSNMEQIIK	29			
	LQDAEEAVEAVNAK	69			
	LQDLVDKLQLK	51			
	LQNEIEDLMVDVER	34			
	MVSLLQEK	32			
	NALAHALQSAR	52			
	NLQEEISDLTEQLGEGGK	81			
	NLQEEISDLTEQLGEGGKNVH ELEK	55			
	NLTEEMAGLDEIIAK	62			
	SLNDFTTQR	34			
	TECFVPDDKEEFVK	36			
	TKYETDAIQR	32			
P12883	NALAHALQSAR	32	222.97	Thick filaments	Y
	AGLLGLLEEMRDER	38			
	KHADSVAELGEQIDNLQR	43			
	HADSVAELGEQIDNLQR	67			
	QKYEESQSELESSQK	25			
	ADIAESQVNK	30			
	ANSEVAQWR	37			
	AQLEFNQIK	37			
	AVVEQTER	37			
	DIDDLELTLAK	60			
	DTQIQLDDAVR	23			
	GTLEDQIIQANPALEAFGNAK	70			
	KLEGDLK	32			
	LELDDVTSNMEQIIK	29			
	LQNEIEDLMVDVER	34			
	MVSLLQEK	32			
	NLTEEMAGLDEIIAK	62			
P08670	TKYETDAIQR	32	53.488		Y
	EYQDLLNVK	56			
	FADLSEAANR	64			
	FANYIDKVR	29			
	FLEQQNK	45			
	GTNESLER	45			
	ILLAELEQLKGQGK	49			
	ISLPLPNFSSLNLR	46			
	KLLEGEESR	40			

	KVESLQEEIAFLK	101			
	LLQDSVDFSLADAINTEFK	77			
	LQDEIQNMKEEMAR	34			
	MFGGPGTASRPSSSR	38			
	NLQEAEEWYK	41			
	QDVDNASLAR	57			
	QQYESVAAK	49			
	QVDQLTNDKAR	57			
	QVQSLTCEVDALKGTNESLER	67			
	RQVDQLTNDK	33			
	SKFADLSEAANR	41			
	FADLSEAANR	64			
	TNEKVELQELNDR	64			
	TYSLGSALRPSTSR	24			
	VEVERDNLAEDIMR	31			
P06732	ELFDPIISDR	38	43.07	Cytoplasmic	
	EQQQLIDDHFLFDKPVSPLLLA S GMAR	52			Y
	GGDDLDPNYVLSSR	47			
	GTGGVDTAAVGSVFDVSNAD R	69			
	GYTLPPHCSR	33			
	GQSIDDMIPAQK	45			
	KLEK	22			
	LGSSEVEQVQLVVDGVK	131			
	LNYKPEEEYPDLSK	58			
	LSVEALNSLTGEFK	78			
	PFGNTHNK	31			
	PFGNTHNKFK	28			
	SFLVWVNEEDHLR	31			
	TDLNHENLK	28			
	VLTLELYKK	24			
P62736	AGFAGDDAPR	56	41.98	Cytoplasmic	Y
	AVFPSIVGRPR	63			
	DIKEK	28			
	DLYANNVLSGGTTMYPGIADR	44			
	DSYVGDEAQSK	50			
	DSYVGDEAQSKR	35			
	EITALAPSTMK	40			
	FRCPETLFQPSFIGMESAGIH ET	34			
	HQGVMVGMGQK	36			
	KDLYANNVLSGGTTMYPGIAD R	69			
	LDLAGRDLTDYLMK	34			
	RGILTLK	35			
	SYELPDGQVITIGNER	43			
	TTGIVLDSGDGVTHNVPIYEG YA	26			
	VAPEEHPTLLTEAPLNPK	37			

P68133	AGFAGDDAPR	56	42.02	Cytoplasmic	Y
	GYSFVTTAER	33			
	DSYVGDEAQSK	50			
	DSYVGDEAQSKR	35			
	EITALAPSTMK	40			
	DLTDYLMK	33			
	HQGVMVGMGQK	39			
	LDLAGRDLTDYLMK	24			
	QEYDEAGPSIVHR	29			
	SYELPDGQVITIGNER	43			
	VAPEEHPTLLTEAPLNPK	37			
	LCYVALDFENEMATAASSSSL EK	33			
P60709	AGFAGDDAPR	80	41.74	Cytoplasmic.	Y
	AVFPSIVGRPR	47			
	DSYVGDEAQSK	67			
	DSYVGDEAQSKR	58			
	EITALAPSTMK	56			
	HQGVMVGMGQK	47			
	SYELPDGQVITIGNER	85			
	DLYANTVLSGGTTMYPGIADR	33			
P12829	ALGQNPTNAEVLR	75	21.42		Y
	HVLATLGEK	51			
	IDFTADQIEEFK	77			
	IDFTADQIEEFKEAFSLFDR	55			
	MLDFETFLPILQHISR	74			
	NKEQGTYEDFVEGLR	72			
	SVKIDFTADQIEEFK	56			
	VFDKESNGTVMGAELR	45			
	VLGKPKPEEMNVK	26			
P49454	DLQEK	22	367.37	Nuclear	Y
P07951	ATDAEADVASLNR	73	32.83		Y
	EAQEKLEQAEK	29			
	HIAEDSDR	35			
	IQLVEEELDRAQER	35			
	KATDAEADVASLNR	83			
	KYEEVAR	35			
	LATALQKLEEAEK	51			
	LEEAEKAADESER	49			
	RIQLVEEELDR	45			
P12882	ADIAESQVNK	33	222.98	Thick filaments	Y
	ANSEVAQWR	37			
	DIDDLELTLAK	60			
	DIDDLELTLAKVEK	43			
	HADSVAELGEQIDNLQR	36			
	KHADSVAELGEQIDNLQR	45			
	KLEGDLK	33			
	TKYETDAIQR	32			
P11055	ADIAESQVNK	30	223.89	Thick filaments	Y
	ANSEVAQWR	37			

	DIDDLELTLAK	60			
	DIDDLELTLAKVEK	70			
	EQDTSAHLER	36			
	HADSVAELGEQIDNLQR	36			
	KLEGDLK	33			
	TKYETDAIQR	32			
Q9Y623	ADIAESQVNK	30	222.87	Thick filaments	Y
	ANSEVAQWR	37			
	DIDDLELTLAK	60			
	EQDTSAHLER	36			
	KLEGDLK	33			
	TKYETDAIQR	32			
O15230	IRIDSLSAQLSQLQK	26	399.45	basement membranes	Y
	KLESTESR	31			
	LEEALQRK	39			
	LQEKEDLQELNDR	41			
	SCRCDIGGALGQSCEPR	31			
P06753	LKEAETR	40	32.80		Y
	SLEAQAEK	43			
	KYEEVAR	33			
	LEEAEKAADESER	66			
	SVAKLEKTIDDLEDELYAQK	39			
Q01449	GKVAATK	29	19.45		Y
	GVVNKDEFK	38			
	DGIICK	24			
	ETYSQLGK	41			
	GVVNKDEFK	38			
	EAFSCIDQNR	46			
	LNGTDPEEAILSAFR	29			
	VSVPEEELDAMLQEGK	71			
	GSSNVFSMFEQAQIQEFK	101			
	ETYSQLGKVSVPEEELDAMLQ EGK	54			
P04792	AQLGGPEAAK	56	22.78	Cytoplasmic	Y
	YTLPPGVDPTQVSSSLSPEGT LTV EAPMPK	49			
	KYTLPPGVDPTQVSSSLSPEG TLTVEAPMPK	41			
	VSLDVNHFAPDELTVK	27			
P21810	DLPETLNELHLDHNK	24	41.65	Secreted	Y
	DLPETLNELHLDHNKIQAIELE DLLR	44			
	EISPDTTLLDLQNNDISELR	42			
	ELHLDNNK	31			
	LGLGHNQIR	57			
	NHLVEIPPNLPSSLVELR	51			
P01028	HLVPGAPFLLQALVR	56	192.77		Y
	LELSVDGAK	31			
P10721	EGEEFTVTCTIK	26	109.86	Type I	Y

	ENSQTK	24		membrane	
P19021	IPVDEEAFVIDFKPR	42	108.33	Type I membrane	Y
P00739	AVGDKLPECEAVCGKPK	45	39.01	Secreted	Y
P00354	AGAHLKGGAK	56	35.88	Cytoplasmic.	Y
	VVDLMAHMASK	45			
Q9H3K6	FEGKPLLQR	34	10.12		Y
P07900	RAPFDLFENR	50	84.54	Cytoplasmic.	Y
Q9UII2	EAGGAFGKR	33	12.25	Mitochondrial.	Y
	EQLAALKK	32			
Q9Y490	AHATGAGPAGR	38	269.72		Y
P02787	HQTVPQNTGGK	37	77.05	Secreted.	Y
	APNHAVVTR	33			
O43920	ECKIEYDDFVECLLR	42	12.39	Mitochondrial inner membrane	Y
O75781	RQAEIENK	40	42.08	Membrane-associated	Y
P30086	YVWLVYEQDRPLK	54	20.93	Cytoplasmic	Y
Q8NCM2	QVAEILK	39	111.91	Integral membrane	Y
P51888	VLEKLPGLVFLYMEK	37	43.81	Secreted	Y
P06576	ADKLAEEHSS	39	56.56	Mitochondrial.	Y
	AIAELGIYPAVDPLDSTSR	58			
	EGNDLYHEMIESGVINLK	35			
	FLSQPFQVAEVFTGHMGK	58			
	IMDPNIVGSEHYDVAR	35			
	LVLEVAQHLGESTVR	39			
	TIAMDGTEGLVR	45			
	VLDSGAPIKIPVGPETLGR	62			
P07355	RAEDGSVIDYELIDQDAR	62	38.47	plasma membrane.	Y
	KELASALK	25			
	KLMVALAK	24			
	AYTNFDAERDALNIETAIK	45			
	TPAQYDASELK	37			
	TDLEKDIISDTSGDFR	42			
	TNQELQEINR	41			
P18669	HYGGLTGLNK	44	28.67		Y
	VLIAAHGNSLR	36			
	KAMEAVAAQGK	66			
	ALPFWNEEIVPQIK	32			

	AMEAVAAQGK	31			
P02743	IVLGQEQDSYGGKFDR	62	25.39	Secreted.	Y
	DNELLVYK	44			
	AYSDLSR	35			
Q12797	RSNEVLR	31	85.50	Type II membrane	Y
P62807	EIQTAVR	34	13.77	Nuclear	Y
	HAVSEGTK	49			
	LAHYNK	28			
	LAHYNKR	32			
	QVHPDTGISSK	31			
Q9BPU6	QKAMGK	36	61.42	Cytoplasmic	Y
P02749	KFICPLTGLWPINTLK	44	38.30	Secreted	Y
P19429	AKESLDLR	39	23.88		Y
	AYATEPHAK	38			
	ISADAMMQALLGAR	83			
	KNIDALSGMEGR	69			
P09382	DSNNLCLHFNPR	56	14.585		Y
O96000	AFDLIVDRPVTLVR	43	20.64	Mitochondrial inner membrane	Y
P52294	KEAAWAITNATSGGSAEQIK	39	60.25	Cytoplasmic and nuclear.	Y
P02144	HGATVLTALGGILK	65	17.05		Y
	HGATVLTALGGILKK	60			
	HLKSEDEMK	32			
	KDMASNYK	30			
	VEADIPGHGQEVLIR	51			
	GHHEAEIKPLAQSHATK	42			
P17540	EVENVAITALEGLK	33	47.52		Y
	GTGGVDTAAVADVYDISNIDR	62			
	HTTDLDASK	34			
	ITQGQFDEHYVLSSR	62			
	LIDDHFLFDKPVSPLLTCAGMAR	57			
	LSEMTEQDQQR	47			
	RGTGGVDTAAVADVYDISNIDR	59			
P13073	AHESVVK	26	19.58		Y
	DHPLPEVAHVK	41			
	HLSASQK	40			
	SEDFSLPAYMDRR	28			
P01160	NLLDHLEEK	41	16.71		Y
O43242	HDADGQATLLNLLLR	51	60.98		Y
P02545	RVDAENR	35	74.14		Y
	LKDLEALLNSK	29			
	KLESTESR	26			

P62258	LICCDILDVLDKHLIPAANTGESK	50	29.17	Y
	LAEQAERYDEMVESMK	32		
	AAFDDAIAELDTLSEESYK	33		
P04406	VIHDNFGIVEGLMTTVHAITATQK	69	35.92	Y
	AGAHLQGGAKR	57		
	GALQNIIPASTGAAK	54		
P24752	TPIGSFLGSLSLLPATK	78	45.20	Y
	IHMGSCAENTAK	42		
Q13439	QENLLKR	37	261.14	Y
	QEVVDVMK	33		
Q14055	MAAATASPR	41	65.13	Y
P62937	SIYGEKFEDENFILK	38	17.88	Y
	VSFELFADKVPK	36		
P08238	RAPFDLFENK	37	83.13	Y
P02511	APSWFDTGLSEMR	49	20.16	Y
	HEERQDEHGFISR	21		
	HFSPEELK	53		
	KQVSGPER	30		
	QDEHGFISR	26		
	VLGDVIEVHGK	43		
	VLGDVIEVHGKHEER	47		
	RPFFPFHSPSR	26		
P30048	HLSVNDLPVGR	48	27.69	Y
P35442	IRLCNSPVPQMGGK	45	129.95	Y
O94966	SSCAKVQTR	36	151.34	Y
P48735	GKLDGNQDLIR	32	50.91	Y
	GRPTSTNPIASIFAWTR	57		
	HAHGDQYK	25		
	LIDDMVAQVLK	40		
	LNEHFLNTTDFLDTIK	29		
	NILGGTVFREPIICK	45		
	TIEAEAAHGTVTR	33		
P99999	TGPNLHGLFGR	44	11.62	Y
	GIIWGEDTLMEYLENPKK	43		
P47985	RLEVLDSTK	44	29.65	Y
	EIEQEAAVELSQLRDPQHDLDR	36		
O14727	IMVAAKNK	36	141.84	Y
P04075	ADDGRPFPQVIK	34	39.29	Y
	ALANSLACQGK	26		
	FSHEEIAMATVTALR	43		
	GILAADESTGSIAK	23		
	IGEHTPSALAIMENANVLAR	50		
	YTPSGQAGAAASESLFVSNHAY	77		
	ELSDIAHR	43		

	AAQEEYVKR	37			
	ELSDIAHR	34			
Q16718	KLEDQLQGGQLEEVILQAEHE LNLAR	39	13.33		Y
P69905	TYFPHFDLSHGSAQVK	36	15.13		Y
	VGAHAGEYGAEALER	71			
	MFLSFPTTK	29			
	VLSPADKTNVK	40			
P51970	HCAEPFTEYWTCIDYTGQQLF R	33	19.97		Y
	FDECVLDKLGWVRPDLGELS K	29			
P20774	RLDFTGNLIEDIEDGTFSK	75	33.92		Y
	ESAYLYAR	54			
	DFADIPNLR	36			
P30533	HAESVGDGER	42	41.47		Y
Q9H2C1	FGTKCAGCAQGISPSDLVR	45	44.41	Nuclear	Y
O60237	LQEAQLELADIK	51	110.42	Cytoplasmic	Y
P11310	ANWYFLLAR	67	46.59	Mitochondrial	Y
P05155	GVTSVSQIFHSPDLAIR	38	55.15	Secreted.	Y
Q9NYQ8	SSNTALLNR	39	479.39		Y
	KGELQVAK	34			
Q14112	IESALLDGSER	54	151.39	Secreted	Y
	EGTSLGEVGGPDLK	67			
Q8WUJ3	LVQYLNAVPDGRILSVAVNDE GSR	47	152.99		Y
	AVVDVPMPKK	52			
Q9P0L0	HEQILVLDPPTDLK	36	27.32	Type IV	Y
	FKGPFTDVVTTNLK	27		membrane	
Q86XX4	EDGRLVIEFK	57	442.93	Type I membrane	Y
Q9Y6M9	QHPQPYIFPDSPGGTSYER	40	21.7	Mitochondrial inner membrane	Y
P07099	GGHFAAFEEPELLAQDIR	37	52.95	Membrane-bound	Y
P51884	SLEDLQLTHNK	69	38.43	Secreted	Y
	NNQIDHIDEK	45			
	VANEVTLN	32			
Q99466	AGPCPPRGCSNGGTCQLMPE K	63	209.62	Type I membrane	Y
	ELRDQAGLAPADVAHQR	57		protein	
Q06830	TIAQDYGVLKADEGISFR	69	22.11	Cytoplasmic.	Y
P00367	GASIVEDKLVEDLR	43	61.39	Mitochondrial	Y
		55			
	DSNYHLLMSVQESLER				
	HGGTIPIVPTAEFQDR	31			

P49748	AGLGSGLSLSGLVHPELSR	46	70.39	Mitochondrial inner membrane	Y
P01023	TEHPFTVEEFVLPK	54	163.28		Y
P11783	NKEQGTYEDFVEGLR	35	10.62		Y
	VFDKESNGTVMGAELR	39			
P01011	NLAVSQVVHK	63	47.65	Extracellular	Y
P60660	VLDFEHFLPMLQTVAK	44	16.80		Y
Q9P2G4	DICELVINK	45	100.35		Y
	LQPATLHCR	35			
P04083	GVDEATIIDILTKR	72	38.58		Y
	KGTDVNVFNTILTTR	52			
Q9Y277	LTLDTIFVPNTGKK	43	30.66	Outer mitochondrial membrane	Y
P35498	RDSLFVPR	60	228.97	Integral membrane	Y
Q92629	GLKLEGDSEFLQPLYAK	63	32.18	Type II membrane	Y
P12109	VFSVAITPDHLEPR	34	108.55		Y
	VAVVQYSGTGQQRPER	41			
P68366	IHFPLATYAPVISAEK	37	49.92		Y
Q13705	SDLTAVLADFGLAVR	71	57.64	Type I membrane	Y
P52732	DEVYQILEKGAAK	49	119.27		Y
Q16082	AALSHDGILNLEAPR	54	20.23		Y
P50213	SNVTAVHK	40	39.59	Mitochondrial.	Y
P01842	AGVETTTPSK	48	11.24		Y
	SYSCQVTHEGSTVEK	46			
P01859	VVSVLTVVHQDWLNGK	65	35.88		Y
P49815	IKVLDVLSFVLLINR	58	200.75	Cytoplasmic	Y
Q16891	KVQAAQSEAK	42	83.68	Mitochondrial inner membrane	Y
	LRACQLSGVTAAAQSCLCGK	38			
P13804	LLYDLADQLHAAVGASR	75	35.08	Mitochondrial	Y
O95169	GGDPSKEPER	43	21.76	Mitochondrial inner membrane	Y
Q9C0A0	SPLGGFQGCMR	39	145.32	Type I membrane	Y
Q96RV3	TSSTNSAKTR	41	258.65	Integral membrane	Y
	DILGGPISLGNIR	56			
Q9P0M6	HILLAVANDEELNQLLK	38	39.93	Nuclear	Y
P32119	KEGGLGPLNIPLLADVTR	85	21.89	Cytoplasmic	Y
	LSEDYGVLKTDEGIAYR	40			
Q9Y512	TKDDIIICEIGDVFK	52	51.96	Integral membrane	Y
Q9P2S2	QLTIFNSQAAIK	34	184.98	Type I	Y

	QLTIFNSQAAIKIGGR	47		membrane	
Q6NZI2	AHATTSNTVSK	32	43.47	plasma membrane	Y
P63267	EIVR	25	41.85	Cytoplasmic	Y
	AGFAGDDAPR	72			
	EITALAPSTMK	60			
	HQGVMVGMGQK	47			
	AVFPSIVGRPR	47			
	DSYVGDEAQSK	80			
	DSYVGDEAQSKR	49			
	SYELPDGQVITIGNER	69			
	VAPEEHPTLLTEAPLNPK	40			
	DLYANNVLSGGTTMYPGIADR	21			
	KDLYANNVLSGGTTMYPGIADR	65			
	LCYVALDFENEMATAASSSSLEK	125			
P63261	EIVR	25	41.77	Cytoplasmic	Y
	AGFAGDDAPR	72			
	EITALAPSTMK	60			
	HQGVMVGMGQK	47			
	AVFPSIVGRPR	47			
	DSYVGDEAQSK	80			
	DSYVGDEAQSKR	49			
	SYELPDGQVITIGNER	69			
P13535	KLEGDLK	33	222.63	Thick filaments	Y
	ANSEVAQWR	37			
	ADIAESQVNK	30			
	EQDTSAHLER	36			
	TKYETDAIQR	32			
	DIDDLELTLAK	60			
Q8IUG5	AGVISRLEK	47	285.01	Cytoplasmic	Y
Q9UKX3	KLEGDLK	33	223.54	Thick filaments	Y
	MVSLLQEK	32			
	ADIAESQVNK	30			
	EQDTSAHLER	36			
	TKYETDAIQR	32			
O43526	AGGAGAGKPPKR	43	95.85	Integral membrane	Y
	EHVDRHGCIVK	47			
P29016	QVKPEAWLSSGPSPGPGR	55	36.939	Type I membrane	Y
Q07507	YFESVLDREWQFYCCR	48	24.005	Secreted	Y
Q14118	EQIAGLSRR	64	97.581	extracellular and type-I membrane	Y
Q9NX02	DLAAVLVVSR	49	120.52	Cytoplasmic	Y
O95672	VCLGQANR	65	87.82	Type II membrane	Y
P36542	SEVATLTAAGK	59	32.99	Mitochondrial.	Y
P02763	YVGGQEHFAHLLILR	72	23.51	Secreted.	Y
Q96KN8	KMVNK	60	30.28		Y
P30049	AQAELVGTADEATR	61	17.49	Mitochondrial	Y

Q13151	KLFVGGLK	43	30.84	Nuclear	Y
P14854	NCWQNYLDFHR	57	10.06		Y
	FPNQNQTR	38			
P12814	EALER	42	103.06	Cytoplasmic.	Y
	DGLKLMLLLEVISGER	38			
O94851	EMASAQEPDKLSMVMYLSK	36	126.69	Cytoplasmic	Y
Q12873	MPDKDDIR	54	220.69	Nuclear	Y
P12277	LAVEALSSLDGDLAGR	52	42.64	Cytoplasmic	Y
P46100	KDYTALTK	35	282.57	Nuclear	Y
	KVQDGLSDIAEK	30			
P00558	VLPGVDALSNI	40	44.48	Cytoplasmic	Y
P06733	AAVPSGASTGIYEALELR	61	47.04	Cytoplasmic	Y
	FTASAGIQVVGDDLTVTNPK	36			
Q86UR5	MTDLGRLGAFITK	41	189.08		Y
	GTASDAER	41			
P07339	EGCEAIVDTGTSLMVGPVDEV R	71	44.55	Lysosomal.	Y
P21796	WNTDNTLGTEITVEDQLAR	72	30.64	Outer	Y
	TDEFQLHTNVNDGTEFGGSIY QK	47		membrane of mitochondria and plasma membrane	
P24539	HVVQSISTQQEK	56	28.91	Mitochondrial.	Y
	LAQLEEAK	49			
	SQQALVQK	45			
P56159	FLNFFK	60	51.46	Attached to the membrane	Y
P14625	LISLTDENALSGNEELTVK	60	92.47	Endoplasmic reticulum	Y
Q9H0E7	KMELIQPK	45	81.15		Y
P42684	INTTADGK	51	128.34	Cytoplasmic.	Y
P00403	VVLPIEAPIR	37	25.56	Integral membrane	Y
O15068	EDLRLALK	51	123.98		Y
P12235	TAVAPIER	41	32.93	Integral	Y
	LAADVGK	34		membrane	
P30044	ALNVEPDGTGLTCSLAPNIISQ L	72	22.03	Mitochondrial, peroxisomal and cytoplasmic	Y
Q9NYA3	QEVEGLEGK	62	79.92		Y
P05413	QVASMTKPTTIIEK	55	14.73	Cytoplasmic	Y
P42338	EAGLDLR	42	122.76		Y
P42261	GLSVLQK	56	101.54	Integral membrane	Y
P42765	DGTVTAGNASGVADGAGAVII ASEDAVK	80	41.92	Mitochondrial	Y
P49411	HYAHTDCPGHADYVK	49	49.542	Mitochondrial.	Y
	DKPHVNVGTIGHVDHGK	34			
Q8NFH8	QALKSTINEALPK	41	71.534	Cytoplasmic.	Y

P45379	ALSNMMHFGGYIQK	34	35.771		Y
	DLNELQALIEAHFENR	67			
	EEEENRR	27			
	KAEDEAR	42			
	KVLAIDHLNEDQLR	41			
	MEKDLNELQALIEAHFENR	30			
	VLAIDHLNEDQLR	39			
Q9UKX2	EQDTSAHLER	34	222.91	Thick filaments	Y
	HADSVAELGEQIDNLQR	36			
	KHADSVAELGEQIDNLQR	43			
	DIDDLELTLAKVEK	42			
P08123	GPSGPQGIR	26	129.46		Y
	GVVGPQGAR	31			
	GPAGPSGPAGK	27			
	HGNRGETGPSGPVGPAGAV GPR	38			
P08133	GLGTDEDTIIDIITHR	56	75.74	Stress fibers	Y
	GFGSDKEAILDIITSR	81			
	DLEADIIGDTSGHFQK	33			
	GTVRPANDFNPDADAK	29			
	SLHQAIEGDTSGDFLK	30			
P11021	DNHLLGTFDLTGIPPAPR	55	72.33	Endoplasmic reticulum	Y
	ITITNDQNR	24			
	VYEGERPLTK	25			
	AKFEELNMDLFR	60			
	IINEPTAAAIAYGLDKR	45			
	ALSSQHQAR	39			
	KSDIDEIVLVGGSTR				
	VEIIANDQGNR	39			
P07919	EQCEQLEK	33	10.74	Mitochondrial inner membrane	Y
	ERLELCDER	40			
	SHTEEDCTEELFDFLHAR	70			
P20674	ILEVVKDK	32	16.774	Mitochondrial inner membrane	Y
	WVTYFNKPDIDAWELR	47			
P01009	FNKPFVFLMIEQNTK	54	46.737	Secreted	Y
	GTEAAGAMFLEAIPMSIPPEV K	48			
P68871	VHLTPEEK	38	15.87		Y
	VNVDEVGGEALGR	65			
	SAVTALWGK	54			
	LLVVYPWTQR	20			
	FFESFGDLSTPDAVMGNPK	76			
P07437	SGPFGQIFRPDNFVFGQSGA GNNWAK	28	49.67		Y
	GHYTEGAELVDSVLDVVR	37			
	GHYTEGAELVDSVLDVVRK	39			
	LHFFMPGFAPLTSR	32			
P31946	LAEQAERYDDMAAAMK	52	27.95	Cytoplasmic	Y
	EMQPTHPIR	33			
P00505	ISVAGVTSSNVGYLAHAIHQVT K	34	47.48	Mitochondrial	Y

	KAEAQIAAK	32			
	NLDKEYLPIGGLAEFCK	41			
P14618	IENHEGVR	27	57.81		Y
	KASDVHEVR	48			
	LAPITSDPTEATAVGAVEASFK	69			
	LNFSHGTHEYHAETIK	33			
	RFDEILEASDGIMVAR	37			
P01857	TTPPVLDSDGSFFLYSK	46	36.11		Y
	GPSVFPLAPSSK	37			
	FNWYVDGVEVHNAK	58			
	THTCPPCPAPELLGGPSVFLF PPKPK				
P25705	AVDSLVPIGR	46	59.75	Mitochondrial	Y
	EAYPGDVFYLHSR	61		inner	
	EGDIVKR	31		membrane	
	EIVTNFLAGFEA	46			
	EPMQTGIK	24			
	GMSLNLEPDNVGVVVFGNDK	59			
	NVQAEEMVEFSSGLK	69			
	TGAIVDVPVGEELLGR	47			
	TGTAEMSSILEER	64			
	VLSIGDGIAR	35			
	VVDALGNAIDGK	56			
	HALIIYDDLSK	32			
	ILGADTSVDLEETGR	108			
P40925	LGVTANDVK	30	36.29	Cytoplasmic	Y
	EVGVYEALKDDSWLK	67			
	ESAFEFLSSA	27			
	GEFVTTVQQR	44			
P14555	SQLCECDKAAATCFAR	34	16.08	Membrane-	Y
	EAALSYGFYGCHCGVGGR	35		associated	
	CCVTHDCCYK	27			
Q99798	AKDINQEVYNFLATAGAK	33	85.43	Mitochondrial	Y
	DLGGIVLANACGPCIGQWDRK	43			
	FNPETDYLTGTDGKK	27			
	GHLDNISNNLLIGAINIENGK	29			
	KQGLLPLTFADPADYNK	37			
	LNRPLTLSEK	36			
P07585	KASYSGVSLFSNPVQYWEIQP STFR	37	39.75	Secreted	Y
	VPGGLAEHK	29			
	AHENEITK	31			
	VVQCSDLGLDKVPK	29			
P40926	KGEDFVK	26	35.53	Mitochondrial	Y
	SQETECTYFSTPLLLGK	57			
	LTLYDIAHTPGVAADLSHIETK	83			
	VAVLGASGGIGQPLSLLLK	72			
	VDFPQDQLTALTGR	67			
P08574	HGGEDYVFSLLTGYCEPPTG	40	35.39	Mitochondrial	Y

	VSLR				
P09493	QLEDELVSLQKK	48	32.69		Y
	LATALQKLEEAEK	51			
	KATDAEADVASLNR	83			
	LEEAEKAADESER	49			
	SKQLEDELVSLQK	65			
	IQLVEEELDRAQER	35			
	SIDDLEDELYAQKLK	81			
	VLSDKLK	34			
	SVTKLEK	35			
	KYEEVAR	39			
	HIAEDADR	38			
	KLVIIESDLER	58			
	RIQLVEEELDR	45			
	LVIIESDLERAEER	28			
Q9H9E3	GVTSAVNIMHSSLQQGK	61	89.09	Golgi	Y
O14647	ENKENKEK	50	200.56	Nuclear	Y
Q8IWV2	IEVQFPETVPTAKGATVK	37	113.45	Attached to the membrane	Y
O60879	KKVKELR	55	126.23		Y
O95793	LAQIQQAK	49	63.51	Cytoplasmic	Y
O15061	LADSSRTLRHIAPGPK	47	172.77	Cytoplasmic	Y
Q05639	STTTGHLIYK	39	50.78	Nuclear	Y
Q8IUD2	DLKEK	45	128.24	Cytoplasmic and membrane-associated	Y
Q96C19	KQIKDMEK	43	26.69		Y
Q92945	SGEMIKKIQNDAGVR	49	72.71	Nuclear	Y
P14136	GTNESLER	45	49.91		Y
	FLEQQNK	45			
Q08380	TLQALEFHTVPFQLLAR	53	65.33	Secreted	Y
Q9BVI0	DKEKNKEK	73	116.74	Nuclear	Y
P51991	KIFVGGIK	62	39.80	Nuclear	Y
P02042	VHLTPEEK	38	16.03		Y
P34931	VEIIANDQGNR	49	70.73		Y
P10809	KGVITVKDGK	53	61.05	Mitochondrial	Y
Q86UE4	SDSDKSSSQVPPILQETDKSK	42	63.837	Type II membrane	Y
P61626	TPGAVNACHLSCSALLQDNIA DAVACAK	37	16.537		Y
Q9NZU5	VKGGDGIRIYK	35	40.833		Y
P07195	LKDDEVAQLKK	43	36.769	Cytoplasmic.	Y
P21757	EEQVHLEQEIK	63	50.187	Type II membrane	Y
P78559	DEVLQQKDK	44	307.60		Y
P41218	LKLVCGSHSFIKVIK	38	45.836	Nuclear and cytoplasmic	Y
P12036	EYQDLLNVK	56	112.64		Y

P49790	KIKTAVR	51	153.89	Nuclear	Y
P51857	EKIAEGKVR	32	37.38	Cytoplasmic	Y
Q13415	LLLVEPSRNDLLLR	47	97.36	Nuclear	Y
P49116	IQIVTDSASVERLLGK	50	65.41	Nuclear	Y
P32248	NDLFKLFK	41	42.87	Integral membrane	Y
P07858	GQDHCGIESEVVAGIPR	71	38.75	Lysosomal	Y
P15259	AMEAVAAQGK	43	28.79		Y
	KAMEAVAAQGK	66			
	ALPFWNEEIVPQIK	32			
Q8TB72	DAETDGPEKGDQKGK	42	114.216	Cytoplasmic	Y
Q9BVV6	DGAAMYSLINALSTNR	50	169.26		Y
Q13029	KKVSHSSK	43	188.92	Nuclear.	Y
P51531	KGKGGAK	30	180.76	Nuclear	Y
P61981	DSTLIMQLLR	57	28.32	Cytoplasmic	Y
	TAFDDAIAELDTLNEDSYK	65			
P63104	DSTLIMQLLR	45	27.90	Cytoplasmic.	Y
O14626	DIKEK	39	37.53	Integral membrane	Y
P31947	DSTLIMQLLR	57	27.871	Cytoplasmic	Y
Q9BXT6	KQVEKIR	43	137.09		Y
P25786	QGSATVGLK	49	29.56	Cytoplasmic and nuclear	Y
P30405	GSGDPSSSSSSGNPLVYLDV	43	22.04	Mitochondrial	Y
	DANGKPLGR				
	IVITDCGQLS	41			
O95831	RVEHHDHAVVSGR	39	66.90	Mitochondrial	Y
Q9NTI2	RLNER	37	130.641	Integral membrane	Y
P52815	EIKNYIQGINLVQAK	63	21.348	Mitochondrial.	Y
P46783	SAVPPGADKK	54	18.898	Cytoplasmic	Y
Q14192	GFLTERDDILCPDCGKDI	33	32.19		Y
Q99523	EQFLRLR	40	92.10	Type I membrane	Y
P23327	HRGHGSEEDEDVSDGHHHH	41	80.24	Sarcoplasmic reticulum	Y
	GPSHR				
Q8NI27	ETKERTPK	30	169.58	Nuclear	Y
P46939	EKLAGLNQR	34	396.47	Neuromuscular junction	Y
Q9H1J7	EGELSTCGCSRTARPKDLPR	42	40.32	secreted	Y
P10606	HYKLVPQQLAH	49	13.69	Mitochondrial inner membrane	N
	GAHYKLVPQQLAH	80			
	AHYKLVPQQLAH	45			
	TGLEREIMLAAKKGLD	32			
P05976	FVEGLRVFDKE	41	21.00		N
	TVMGAELRHVLATLGEK	31			
	PKKDVKKPVAAAAAAPAPAPA	26			
	PAP				

	APKKDVKKPVAAAAAAPAPAP APAP	35			
	PKKDVKKPVAAAAAAPAPAPA PAPA	32			
O00268	PAAGGPAGVSGQPGPGAAAA APAP	45	110.05	Nuclear	N
	APAAGGPAGVSGQPGPGAAA AAPAP	45			
	AAGGPAGVSGQPGPGAAAAA PAP	29			
P04198	AGPAVASGAGIAAPAGAPGVA P	35	49.53	Nuclear	N
	AGPAVASGAGIAAPAGAPGVA PPRPG	38			
Q14686	GAPQLQANQN	35	218.98	Nuclear	N
	QPQPQLPQQQ	45			
	IPAAPLTTN	39			
Q13232	GDFCIEVGKNLIHG	32	19.00		N
	ACTGAHERTF	49			
	CTGAHERTF	45			
P87889	QPLSGNEQRGQPQAPQ	39	73.94	Cytoplasmic membrane	N
	QVPVRLQPQK	49			
Q02221	GNHTLFHNSHVNPLPTGYEH P	43	10.84	Mitochondrial inner membrane	N
	SHVNPLPTGYEHP	34			
Q9BV73	QGVQLGEVSG	37	280.96 7	centrosome	N
	QQEQQQAQGQ	32			
P09211	SARPKLKAFLA	42	23.21		N
	GSLKASCLYG	28			
P63316	GRIDYDEFLEFMK	29	18.39		N
	NDGRIDYDEFLEFMKGVE	30			
	GDKNNDGRIDYDEFLEFMKG VE	31			
P60174	APSRKFFVG	33	26.52		N
	APSRKFFVGG	22			
	SLKPEFVDIINAKQ	30			
P24311	SHQKRTPDFH	51	9.15		N
	SHQKRTPDFHD	33			
	SPVGRVTPKEWRNQ	44			
	SHQKRTPDFHDKYG	39			
	SHQKRTPDFHDKY	29			
	SHQKRTPDFHDKYGNAVLA	27			
P15954	TPFLVVRHQLLKT	45	7.24		N
	SAFATPFLVVRHQLLKT	33			
	SHYEEGPG	35			
	SHYEEGPGK	36			
Q9Y6Y8	QPDPESVVLGTDG	45	111.07 6	Endoplasmic reticulum	N
	QPDPESVVLGTD	30			

P25205	QEMPEKAPAGQL	47	90.98	Nuclear	N
P42356	AANPGPAEGAVGPKVALK	35	231.29		N
	AANPGPAEGAVG	37			
P78325	AANPGPAEGAVGPKVALK	43	88.67	Type I	N
	AANPGPAEGAVG	32		membrane	
Q99797	NVKPQGSRLDLFGE	65	80.61	Mitochondrial	N
	NVKPQGSRLDLF	67			
Q92610	KGAAPGSQTGKKQQS	45	137.55	Nuclear	N
	AAPGSQTGKKQQS	35			
Q9UGR2	WQQMEAHAGKASSSM	41	111.58	Nuclear	N
	KLAASVLDALDPPGPT	36			
P08559	PPFEVRGANQWIKFKSVS	65	43.30	Mitochondrial	N
	PFEVRGANQWIKFKSVS	35			
Q8TCE6	GQLIVQSAEDPEKSESHVIQ	40	40.51		N
	KLHKEM	33			
P23193	VQTRSADEPMTT	55	33.97	Nuclear	N
	FVVCNECGNRWKF	47			
P29966	GAPPEQEAAPAEEPA	50	31.41		N
	DKEEPAAAGSGAASPS	40			
Q12962	AAPVSAGGAAPPEGAI	59	21.71	Nuclear	N
Q15596	QGNMGGNSMFSQQSPPHF	46	159.16	Nuclear	N
	AQRQREILNQHLRQR	37			
Q14999	QLNDSAAEPGA	73	191.19		N
Q9Y5A6	KPYVCTKCGKAFSH	39	53.66	Nuclear	N
	QVSTPPNEQKPVW	64			
P62805	SGRGKGGKGLGK	49	11.23		N
	SGRGKGGKGLGKG	46			
Q14571	PNNGQEVL	57	308.08	Integral	N
	IMCTGPEAGNTE	33		membrane	
Q06033	RSPFRLL	75	99.12		N
	RSPFRLLG	64			
Q9NQT8	ISAKDVPT	45	202.67		N
P08246	QLNGSATINANVQV	52	28.52		N
	ALLLGGTALASEIVGGRR	60			
Q9BQQ3	QPAGGAEGFHLH	43	46.35	peripherical	N
	QPAGGAEGFHLHG	68		membrane	
Q9NZM4	AGVSPQGAGLVIQKNL	42	152.89		N
	EPGALPQQPKAPQNLTFM	35			
	QPQAQQPPQAP	29			

	LPNGSPGGAT	37			
Q96L91	NPEAKAAA	29	343.43		N
	APGALTTPGGSAPAQVVH	40			
	AQGPAAVQQ	33			
Q9Y463	NEVYYAKK	31	69.20	Nuclear	N
	SYNLYDLLRNTH	79			
	PELSIIHCDLKPE	64			
P42658	VLSKIPHGDPQSLDPP	43	97.59	Type II	N
	ASHLLGGQGPEEDGGAGAKP	39		membrane	
Q9P2D1	KEADKSLLIGVFKH	43	252.48	Nuclear	N
	QPEAGAVSRGKNFDEESNA	39			
Q03692	IGKPGAAGAPGQPGIPG	75	66.16		N
P02458	KGPPGPQGPAGEQG	45	134.49		N
	PGPRGRDGEPGTLGNPGPPG PPG	88			
	ERGRTGPAGAAGARGNDGQ PGPAG	70			
	AGPTGKQGDRGEAGAQPGP MG	40			
	AGPQGKVGPSGAPGED	26			
Q9BX69	KTQGGASNPALQ	38	116.49		N
O94812	NLLAKDPN	37	131.90		N
	LEALWELLLQAILQALG	31			
P22004	QQQQQLPR	38	57.19		N
	GVHVHPRAAGLVG	45			
	RRLKTQEK	31			
O94833	NQFGDSQQL	39	590.89	Cytoplasmic	N
	AAQLQEALLH	64			
P02452	PGEAGRPGEAGLPGAK	77	138.88		N
O95477	KEDSVSQSSSDAG	44	254.29		N
Q9Y696	KPADLQNLAPGTHPPF	55	28.77	trans-Golgi	N
Q8WWM	KPQPLQQPSQPQQPP	64	113.37	Membrane-	N
7	KISLAPTDVKELSTKEPG	62		associated	
P07307	GAKLEKQQQ	45	35.19	Type II membrane	N
P53367	KMRNDVSVKLKFLEE	36	41.74		N
	KMRNDVSVKLKFLEEN	34			
Q9H2W1	QDSLKKH	37	26.94	Integral	N
	PEPTNQGQDSLK	45		membrane	
Q12968	LFQQDAT	48	115.52	Cytoplasmic	N
	EPEDREP	30			
O0048	KLKKERPDF	38	9.36	Mitochondrial	N

	SKLKKERPDF	56		inner	
	VDYSKLKKERPDF	52		membrane	
	MLRQIIGQAKKH	42			
	SVNVDYSKLKKERPDF	45			
O95298	EIFEKFHPIR	37	14.18	Mitochondrial	N
	GEIFEKFHPIR	56		inner	
	TYGEIFEKFHPIR	57		membrane	
O15031	QVNKLIHA	41	205.10	Type I membrane	N
Q9BXC0	VSCESFIMESANGWH	50	39.295	Integral	N
	VSCESFIMESAN	37		membrane	
Q9Y4D7	QLSGANLSLEAEAAV	75	211.95	type I	N
	EQLDCGAAHLQHPLSILQP	64	9	membrane	
Q10571	QSLQQQQQQQ	59	135.86		N
	PQQPPQQQPPPPPGLLVRQ	71			
	RQSTPHSGPGVN	29			
P01133	QDLKNGGQPVAGE	36	133.86		N
	ECQLGVHSC	88			
Q8WYP3	LSILDRLLH	54	100.16	Cytoplasmic	N
Q13332	ILQPIEGIMG	48	217.09	Type I membrane	N
P23468	QTGVPGQP	51	214.76	Type I	N
	LREDQIPR	36		membrane	
	QTPGMASHPPI	68			
Q9H4A3	QPQATQPTTLAS	28	250.60	Cytoplasmic.	N
	QQPPAAAAPGEQAVAGPA	49			
	QQPPAAAAPGEQAVAGPAP	30			
	QQPPAAAAPGEQAVAG	31			
	QQPPAAAAPGEQAV	29			
P23975	PQVQPENNGADTGP	40	69.33	Integral	N
	PQVQPENNGADTGPEQPLRA	56		membrane	
P38646	TGEQKEDQKEEKQ	68	73.68	Mitochondrial	N
Q9UPX8	QGSSMEIDPQAP	84	134.80	Cytoplasmic	N
Q7Z7G0	QNPPTNLTVVTVEG	55	118.57		N
	KTVVGSKKV	41			
Q8TB96	QPQILLGGNLSWH	43	68.06	Type I	N
	QPQILLGGNLSWHP	55		membrane	
	QPQILLGGNLSWHPA	31			
Q96AY4	YQALQRVL	34	177.62		N

	GAPGTMTSKR	31			
P08138	VGLVAYIAF	90	45.15	Type I membrane	N
Q9BYJ9	QAPSPQAAPQPQ	30	60.87		N
	GGAGSDSNSPGNVQPN	34			
	SPGNVQPNSAPSVESH	81			
O14949	SPFEQRAYPHVF	38	9.77	Mitochondrial inner membrane	N
	TEEFERSKRKNPAAYENDK	59			
	TEEFERSKRKNPAAYEN	30			
P14138	NTPEQTVPYGLSNYR	72	25.44	Secreted	N
P00387	TGITPMLQVIRAIMKD	45	34.08	cytoplasmic	N
	ITPMLQVIRAIMKD	35			
Q10469	SLRKAQGIDNVL	46	51.55	Type II membrane	N
Q9UM54	KKREDDE	55	148.62		N
	KEERNYH	45			
P04179	KHSLPDLPYDY	48	24.71	Mitochondrial	N
	KHSLPDLPYDYG	57			
	GEPKGELLEAIKRDFG	68			
P08519	KPQVEPK	38	501.32		N
Q9UIF8	KKLHVKGKKTNE	54	220.71		N
Q9H0R1	KPKISAHRK	44	54.767		N
Q15744	VAHCGQTAM	36	30.702	Nuclear	N
O75390	YGHAVLRKTD	42	51.712	Mitochondrial	N
P98088	NGIVVSRI	46	130.07	Secreted	N
O14686	QQVSLLAQ	51	563.83	Nuclear	N
Q13415	LKPRTRCAAP	33	97.36	Nuclear	N
Q00888	KRRDGTGGV	42	47.07	Secreted	N
Q92545	QPQEPQPERLSPAP	42	197.59		N
Q9UEW8	PVTAAAAAAPAAATAAPAPAA P	53	59.64	Cytoplasmic and nuclear	N
	PVTAAAAAAPAAATAAPAPAA				
Q9ULV3	VQPQVQPQAH	43	100.03	Nuclear	N
	QLLQLQQLLQQS	33			
	VQPQVQPQ	29			
P41220	KQQAFIKPSP	92	24.38		N
	KQQAFIKPSPE	83			
P36957	PAEAPAAAPKAEPTAAAVPP	71	48.64	Mitochondrial	N
	PAEAPAAAPKAEPTAAAVPPP	58			

	A				
	PAEAPAAAPKAEPTAAAVPPP AA	37			
Q8NH93	SDPRLQNPMYFF	41	36.63	Integral membrane	N
P08473	ESAIDSRGGEP	44	85.38	Type II membrane	N
P35555	DINECAQNPL	34	312.31		N
	DINECAQNPLD	54			
Q9H251	NVTGAVDADEGPNAIVYYF	45	369.53	Type I membrane	N
P23526	KPQVDRY	39	47.58	Cytoplasmic	N
Q9UGJ1	KTPPTAVTEHD	59	76.09	Centrosome	N
P38570	ARVQNITQVGSVTK	43	130.09	Type I membrane	N
Q6VAB6	KSHEFQLGHR	38	93.58	Cytoplasmic and membrane-associated	N
P78415	DNPRRSPPGAG	45	52.12	Nuclear	N
P54725	QPQFQNMRQ	48	39.61	Nuclear	N
Q7Z2W7	QGGGKETLKAIN	34	127.65	Integral membrane	N
Q8NB12	KVPNENIRL	49	56.62	Nuclear and cytoplasmic	N
Q9UBE8	AAAAAAAAAAQMLNPGQ	48	57.05	Nuclear	N
	AAAAAAAAAAQMLNPGQQ	53			
Q99497	KDKMMNG	36	19.89	Nuclear, and	N
	KDKMMNGG	40		cytoplasmic	
Q14767	KPVCEPPCQN	39	195.06	Secreted	N
P51816	HSEQSTF	54	144.74		N
	GIPKNSVPQNPNNKNEP	40			
P56696	IQAAQWRL	38	77.09	Integral membrane	N
Q00056	NTKMRSSN	35	34.48	Nuclear	N
	QPPAQAKGPAHGLHAS	40			
Q9HBL0	GASPLSSQPL	51	185.68	Focal adhesions	N
	DELPNQDGHSAGSM	65			
Q9Y6N7	KGEPATLNCKAEGR	44	180.93	Type I membrane	N
	VVDMRTNPGDP	53			
Q9Y6V0	QMQLTPGSSPTQAPIGED	64	566.66		N
	KPPAQPLGPA	32			
P11717	RNPACSGANICQV	43	274.31	Type I membrane	N
Q9Y5Y9	KYYNAM	53	220.57	Integral	N

	PGDKIHCL	42		membrane	
Q86UK5	ELSASEMLTKS	71	147.95		N
Q8NF91	KEDVSSIVMSTLRE	39	1011.0	Type IV	N
	KATLDTALS	77		membrane	
	QELGMEG	57			
Q96F46	GAAVRLALAG	92	96.13	Type I	N
	RQSVQSDQ	83		membrane	
P20585	QVDPGAAAAAAAAAAAAPPA P	53	127.46		N
	QVDPGAAAAAAAAAAAAPPA PPAPA	60			
	QVDPGAAAAAAAAAAAAPPA PPA	68			
Q8WUU4	GEGAGAAATAGVQEPGAPGS GAQAGPG	46	50.81		N
	GAGAAATAGVQEPG	39			

Table 2.2 List of proteins detected in both trypsin and MAAH digests

Name	Access ID	Unique Peptide Sequence (trypsin digest)	Unique Peptide Sequence (MAAH)
Desmin	P17661	FANYIEK GTNDSLMR LLEGEESR AQYETIAAK ADVDAATLAR TSGGAGGLGSLR VSDLTQAANK KLLEGEESR QVEVLTNQR DNLLDDLQR EYQDLLNVK VAELYEEELR IESLNEEIAFLK FASEASGYQDNIAR	SPLSSPVFPRAGFGSKG SQAYSSSQRVSSYRRTFG SPVFPRAGFGSKG SSPVFPRAGFGSKG
Putative phosphoglycerate mutase 3	P18669	HYGGLTGLNK VLIAAHGNSLR KAMEAVAAQGK ALPFWNEEIVPQIK AMEAVAAQGK	SPAGHEEAKRG IKEGKRVLIAAH
Voltage-dependent anion-selective channel protein 1	P21796	WNTDNTLGTEITVEDQLAR TDEFQLHTNVNDGTEFGGSI YQK	GHKLGLGLEFQA HKLGLGLEFQA

Creatine kinase, M chain	P06732	ELFDPIISDR	PIISDRHG
		EQQQLIDDHFLFDKPVSPLL	PIISDRHGGY
		LAS GMAR	PIISDRHGGYKPT
		GGDDLDPNYVLSSR	PIISDRHGGYKPTD
		GTGGVDTAAVGSVFDVSNA	PNYVLSSRVR
		DR	SYEVFKELFD
		GYTLPPHCSR	PNYVLSSRVRTG
		GQSIDDMIPAQK	PFGNTHNK
		KLEK	
		LGSSEVEQVQLVVDGVK	
		LNYKPEEEYPDLSK	
		LSVEALNSLTGEFK	
		PFGNTHNK	
		PFGNTHNKFK	
		SFLVWVNEEDHLR	
		TDLNHENLK	
		VLTLELYKK	
Serum albumin precursor	P02768	AACLLPK	DAHKSEVAHRF
		AEFAEVSK	DAHKSEVAHRFK
		LCTVATLR	HKSEVAHRFKDLG
		DDNPNLPR	AHKSEVAHRFKDL
		FQNALLVR	DAHKSEVAHRFKD
		QTALVELVK	AHKSEVAHRFKDLG
		TYETTLEK	DAHKSEVAHRFKDL
		LDELRDEGK	DAHKSEVAHRFKDLG
		CCTESLVNR	DAHKSEVAHRFKDLGE
		LVNEVTEFAK	DAHKSEVAHRFKDLGEE
		FKDLGEENFK	
		AVMDDFAAFVEK	
		AAFTECCQAADK	
		ETYGEMADCCAK	
		YICENQDSISSK	
		TCVADESAENCDK	
		ADDKETCFAEEGK	
		CCAAADPHECYAK	
		ADDKETCFAEEGKK	
		KVPQVSTPTLVEVSR	
		QNCELFEQLGEYK	
		QEPERNECFLQHK	
		VFDEFKPLVEEPQNLIK	
		SHCIAEVENDEMPADLPSLA	
		ADFVESK	
		EFNAETFTFHADICTLSEKE	
		R	
		HPYFYAPELLFFAK	
		KQTALVELVK	
		LKECCEKPLLEK	
		MPCAEDYLSVVLNQLCVLH	
		EK	
		RPCFSALEVDETYVPK	
Troponin I	P19429	AKESLDLR	SSDAAREPRPA
		AYATEPHAK	AAREPRPAPA
		ISADAMMQALLGAR	
		KNIDALSGMEGR	

Alpha-actin 2	P62736	AGFAGDDAPR AVFPSIVGRPR DIKEK DLYANNVLSGGTTMYPGIA DR DSYVGDEAQSK DSYVGDEAQSKR EITALAPSTMK FRCPETLFQPSFIGMESAGI HET HQGVMVGMGQK KDLYANNVLSGGTTMYPGIA DR	PPERKYSVWIG PPERKYSVWIGG SIVGRPRHQG APRAVFP APRAVFPS DDAPRAVFP DEAQSKRGIL VTHNVPIYEG SYVGDEAQSKR SYVGDEAQSKRG RVAPEEHPTLL TFYNELRVAPEEH LVKAGFAGDDAPR GQVITIGNERFR) SGLVKAGFAGDDAPR SGLVKAGFAGDDAPRA SGLVKAGFAGDDAPRAVF YALPHAIMRLD YALPHAIMRLDLA YALPHAIMRLDLAG
Myosin light polypeptide 4	P12829	ALGQNPTNAEVLR HVLATLGEK IDFTADQIEEFK IDFTADQIEEFKEAFSLFDR MLDFETFLPILQHISR NKEQGTYEDFVEGLR SVKIDFTADQIEEFK VFDKESNGTVMGAELR VLGKPKPEEMNVK	FVEGLRVFDKE TVMGAELRHVLATLGEK APAPAPAPAPAPEAPKEPAF D
Myosin light chain2a	Q01449	GKVAATK GVVNKDEFK DGIICK ETYSQLGK GVVNKDEFK EAFSCIDQNR LNGTDPEEAILSAFR VSVPEEELDAMLQEGK GSSNVFSMFEQAQIQEFK ETYSQLGKVSVPEEELDAM LQEGK	KVAATKQAQRG TRGKVAATKQAQR TRGKVAATKQAQRG PEEAILSAFRMFD GTRGKVAATKQAQRG PSGKGVVNKDEFKQLLL NIDYKSLCYIITHGDEKEE PSGKGVVNKDEFKQLLLTQ ADKF PEEAILSAFRMF GTDPEEAILSAFRMFD
Myoglobin	P02144	HGATVLTALGGILK HGATVLTALGGILKK HLKSEDEMK KDMASNYK VEADIPGHGQEVLIR GHHEAEIKPLAQSHATK	HHEAEIKPLA HGQEVLIRLF HHEAEIKPLAQ HGQEVLIRLFKG HHEAEIKPLAQSHA IPGHGQEVLIRLFK AMNKALELFRKDMA HGQEVLIRLFKGHPE
Heat-shock protein beta-1	P04792	AQLGGPEAAK YTLPPGVDPTQVSSSLSPE GTLTV EAPMPK KYTLPPGVDPTQVSSSLSPE GTLTVEAPMPK VSLDVNHFAPDELTVK	TERRVPFSLL TERRVPFSLLR TERRVPFSLLRGP GPEAAKSDETAAK

Glyceraldehyde -3-phosphate dehydrogenase , liver	P04406	VIHDNFGIVEGLMTTVHAITA TQK AGAHLQGGAKR GALQNIIPASTGAAK	PSKIKWGDA PSKIKWGDAG GPSGKLWRD GKVKVGVNGFGRIG SNRVVDLMAHMASKE PITIFQERD NPITIFQERD GAKRVIISAPSAD PSKIKWGDAGAEYVVE KLVINGNPITIFQERD
NADH-ubiquinone oxidoreductase 15 kDa subunit	O43920	ECKIEYDDFVECLLR	TPPPHHIGKGEPRP PFLDIQKRF
Serum amyloid P-component precursor	P02743	IVLGQEQDSYGGKFDR DNELLVYK AYSDLSR	HTDLSGKVFVFPR HTDLSGKVFVFPRE YVIIKPLVWV KVFVFPRE
Creatine kinase, sarcomeric	P17540	EVENVAITALEGLK GTGGVDTAAVADVYDISNID R HTTDLDASK ITQGQFDEHYVLSSR LIDDHFLFDKPVSPLLTCAG MAR LSEMTEQDQQR RGTGGVDTAAVADVYDISNI DR	PVIKLRHN PVIKLRHNG IKVPPPLPQFGKK PVIKLRHNGY PVIKLRHNGYD
Cytochrome c oxidase subunit IV	P13073	AHESVVK DHPLPEVAHVK HLSASQK SEDFSLPAYMDRR	AHESVVK AHESVVKSE AHESVVKSED AHESVVKSEDF
Alpha crystallin B chain	P02511	APSWFDTGLSEMR HEERQDEHGFISR HFSPEELK KQVSGPER QDEHGFISR VLGDVIEVHGK VLGDVIEVHGKHEER RPFFPFHSPSR	REEKPAVTAAPKK TIPITREEKPAVTAAPKK
Fructose-bisphosphat e aldolase	P04075	ADDGRPFPQVIK ALANSLACQGK FSHEEIAMATVTALR GILAADESTGSIAK IGEHTPSALAIMENANVLAR YTPSGQAGAAASESLFVSN HAY ELSDIAHR AAQEEYVKR ELSDIAHR	GRPFPQVIKSKG RPFPQVIKSKG
Peroxiredoxin 2	P32119	KEGGLGPLNIPLLADVTR LSEDYGVLKTDEGIAYR	SDTIKPNVDDSKEYFSKHN SKEYFSKHN

Peroxiredoxin 1	Q06830	TIAQDYGVLKADEGISFR	SDTIKPD SDTIKPDVQKKSKEYFSKQK
Hemoglobin alpha chain	P69905	TYFPHFDLSHGSAQVK VGAHAGEYGAEALER MFLSFPTTK VLSPADKTNVK	AHKLRVD TVLTSKYR HAHKLRVD SHGSAQVKGH FPHFDLSHG LHAHKLRVD DLHAHKLRVD YFPHFDLSHG SVSTVLTSKYR TYFPHFDLSH SDLHAHKLRVD TYFPHFDLSHG LSDLHAHKLRVD KTYFPHFDLSHG TNVKAAWGKVGAHAG ALSDLHAHKLRVD SAQVKGHGKKVADAL KTNVKAAWGKVGAHA TKTYFPHFDLSHG KTNVKAAWGKVGAHAG TTKTYFPHFDLSHG LSALSDLHAHKLRVD HAGEYGAEALERMFL SFPTTKTYFPHFDL AHAGEYGAEALERMFL VLSPDKTNVKAAWGKVG SLDKFLASVSTVLTSKY PNALSALSDLHAHKLRVD SPADKTNVKAAWGKVGAHA G SFPTTKTYFPHFDLSHG SLDKFLASVSTVLTSKYR VLSPADKTNVKAAWGKVGA H VLSPADKTNVKAAWGKVGA HA MPNALSALSDLHAHKLRVD VLSPADKTNVKAAWGKVGA HAG DMPNALSALSDLHAHKLRV D VDDMPNALSALSDLHAHKL RVD VLSPADKTNVKAAWGKVGA HAGEY
Myosin XVIIIB	Q8IUG5	AGVISRLEK	EGQSIVGKGLGTPK KAEKTRTGGL
ATP synthase B chain	P24539	HVVQSISTQQEK LAQLEEAK SQQALVQK	PVPPLPEYGGKVRY PVPPLPEYGGKVRYG

Elongation factor Tu	P49411	HYAHTDCPGHADYVK DKPHVNVGTIGHVDHGK	AVEAKKTYVRD AVEAKKTYVRDKPHVNVG
Malate dehydrogenase	P40926	KGEDFVK SQETECTYFSTPLLLGK LTLYDIAHTPGVAADLSHIET K VAVLGASGGIGQPLSLLLK VDFPQDQLTALTGR	EDFVKTLK SIKKGEDFVKTLK PARVNVPVIGGHAG SPLVSRLTLYDIAH
Thioredoxin-dependent peroxide reductase	P30048	HLSVNDLPVGR	PAVTQHAPYFKG PNGVIKHL
Hemoglobin beta chain	P68871	VHLTPEEK VNVDEVGGEALGR SAVTALWGK LLVVYPWTQR FFESFGDLSTPDAVMGNPK	ALAHKYH VHLTPEEK NALAHKYH HGKKVLGAF NPKVKAHGK AHGKKVLGAF VANALAHKYH GVANALAHKYH PKVKAHGKKVL PKVKAHGKKVLG NPKVKAHGKKVL VAGVANALAHKYH GNPKVKAHGKKVLG NPKVKAHGKKVLGA VVAGVANALAHKYH NPKVKAHGKKVLGAF GNPKVKAHGKKVLGAF PKVKAHGKKVLGAFSD VHLTPEEKSAVTALWG PKVKAHGKKVLGAFSDG NPKVKAHGKKVLGAFSD NPKVKAHGKKVLGAFSDG YQKVVAGVANALAHKYH AYQKVVAGVANALAHKYH AAYQKVVAGVANALAHKYH VQAAYQKVVAGVANALAHK YH PVQAAYQKVVAGVANALAH KYH

2.4 Conclusions

The application of combined methods to the analysis of proteins from human heart tissue was demonstrated in this study. Peptides were generated by both in-solution trypsin digestion and MAAH from SDS-soluble and -insoluble protein mixtures, respectively.

Strong cation exchange chromatography was used to fractionate the peptide mixtures generated from the two methods in order to simplify the complex peptide mixtures further. LC-ESI MS/MS was used to analyze the fractions collected from the trypsin digestion. However, LC-MALDI MS/MS was used to analyze the peptides collected from the MAAH

Using this combination of two methods, a wide range of proteins, in terms of GRAVY and molecular weight were identified. Significantly lower and higher molecular weight proteins were obtained from the SDS-insoluble protein mixture than the SDS-soluble protein mixture. Therefore, some of the problems encountered in the analysis of shotgun proteomics could be solved by generating peptides by MAAH from SDS-insoluble proteins.

2.5 Literature Cited

1. Chien, K. R. *Nature* **2000**, 407, 227-232.
2. Tan, F. L.; Moravec, C. S.; Li, J.; Apperson-Hansen, C.; McCarthy, P. M. et al. *Proc. Natl. Acad. Sci.* USA **2002**, 99, 11387-11392.
3. Boheler, K. R.; Volkova, M.; Morrell, C.; Garg, R. et al. *Proc. Natl. Acad. Sci.* USA **2003**, 100, 2754-2759.
4. McGregor, E.; Dunn, M. Hum. Mol. Genet. **2003, 12, 135**-144.
5. Jager, D., Jungblut, P. R., Muller-Werdan, U., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2002**, 771, 131-153
6. Thiede, B; Otto, A; Zimny-Arndt, U; Müller, E C; Jungblut, P. *Electrophoresis* **1996**, 17, 588-599.

7. Muller E.-C.; Thiede B.; Zimny-Arndt U.; Scheler C.; Prehm J.; Muller-Werdan U.; Wittmann-Liebold B.; Otto A.; Jungblut P. *Electrophoresis* **1996**, 17, 1700-1712.

8. Corbett, J. M., Why, H. J., Wheeler, C. H., Richardson, P. J. et al., Electrophoresis **1998**, 19, 2031-2042.

9. Link, A. J., Eng, J., Schieltz, D. M. et al., Nat. Biotechnol. **1999**, 17, 676-682.

10. Washburn, M. P., Wolters, D., Yates, J. R. 3rd, *Nat. Biotechnol.* **2001**, 19, 242-247.

11. Rusel, C.; Tan, T.; Kinter, M.; Bond, M. *Proteomics* **2004**, 4, 1505-1516.

12. Zhong, H.; Marcus, S.; Li, L. J. Am. Soc. Mass Spectrom, **2005**, 16, 471-481.

13. Nandakumar, M. P.; Shen, J.; Raman, B.; Marten, M.R. *J. Proteome Res.* **2003**, 2, 89-93.

14. Kingston, H. M.; Haswell, S. J. (Editors), Microwave-Enhanced Chemistry: Fundamentals, Sample Preparation, and Applications. ACS: Washington, D.C., **1997.**

15. Zhang, B.; McDonald, C.; Li, L. *Anal. Chem.* **2004**, 76, 992-1001.

Chapter 3

Identification and Relative Quantification of Protein Mixtures from Squamous Carcinoma Cells and Human Heart Tissue Samples

3.1 Introduction

One area of focus in proteomics research is identification and relative quantification of complex protein mixtures from cells, tissues or body fluids. Proteomics research on quantification is mostly relative quantification or profiling proteomics. It is used for the identification of proteins that are differentially expressed between different samples. This approach holds promise for identifying disease markers that could be important in early detection and diagnosis, and in monitoring the efficacy of a treatment, and eventually could lead the way to the design of novel treatments [1]. One application of this approach is in the comparison of changes in protein expression between normal and transformed squamous carcinoma cell lines. Using similar approaches, it is possible to characterize global alterations in protein expression associated with processes of human heart disease.

The traditional method for quantitative analysis of protein mixtures was two-dimensional gel electrophoresis (2-DE) in combination with an MS based technique, like peptide mapping or MS-sequencing (MS/MS) [2-7]. Although widely used, the 2-DE/MS technique has limitations when dealing with very large or small proteins, proteins having extreme pIs, and membrane or low–abundance proteins. The chromatography-based mass spectrometric method, multidimensional protein identification technology (MudPIT) [8], has been shown to overcome the drawbacks in 2-DE and thus represents a powerful alternative to 2-DE. The methods include digestion of complex protein mixtures and data-dependent

LC/MS/MS analysis of the resulting peptides, followed by protein identification through database searching [2-10]. This approach can be complemented by quantitative profiling of the complex protein digests such as isotope-coded affinity tags (ICAT) [11-14]. ICAT is based on modifying the side chain of cysteine residues with either a light or heavy isotope-labeled biotin tag. These reagents enrich cysteine containing peptides and also provide a relative measurement of the amount of protein present in two different samples. One of the drawbacks of this technology is that it fails to quantify cysteine-free proteins in the given sample. Apart from targeting the thiol group of cysteine, different tagging strategies for proteome analysis by mass spectrometry have been reported[15-34]. For example, a different isotope coding procedure has been described by Hsu et al.[35], who labeled amino groups of the N-termini and lysine side chains by reductive amination using d_0-and d_2-formaldehyde, resulting in a mass shift of 4 amu. In this labeling strategy, dimethylamino groups are formed, incorporating two isotopicaly labeled moieties. The application of the method towards digests of standard protein mixtures and cell lysates are reported[35-36]. The compatibility of this method with two-dimensional liquid chromatography has been evaluated recently by our group and reported[37]. A chemical labeling strategy that specifically labels the N-terminus of all peptides in a digested sample with either light or heavy tagging reagent group can also be done. In this method, to prevent the incorporation of multiple labels, amino groups of lysine side chains are blocked by guanidation[38-39] prior to N-terminal labeling[40]. In the first work of this study, stable dimethyl isotope labeling of the amino groups of N-termini and lysine groups was applied to quantify and determine the proteins that are differentially expressed between an E-cadherin-deficient human carcinoma cell line (SCC9) and E-cadherin- and plakoglobin-expressing SCC9 transfectants. A squamous

carcinoma cell line (SCC9), which lacks plakoglobin (Pg) and E-cadherin and doesn't assemble desmosomes, was used as a control in this study. SCC9 cells have both α-and β-catenins but display unusual expression of N-cadherin, which is not normally expressed in epithelial cells. Plakoglobin is a multifunctional cytoplasmic protein. It is a major component of both desmosomes and adherens junctions. Desmosome and adherin junctions connect cells together and anchor the cytoskeleton to the plasma membrane. Loss of cell adhesion and alterations in the expression of cadherin are common features of malignant cells and markers for aggressive tumor growth and poor prognosis. The two other cell lines used in this study were made by introducing E-cadherin and plakoglobin expression and they are referred to as SCC9-E-cad and SCC9-Pg, respectively. As reported [41-42], transfection of SCC9 cells with E-cadherin cDNAs induces a morphologic transformation from fibroblast to epidermoid which coincides with downregulation of the endogenous N-cadherin and increased synthesis and stability of the catenins, but introduction of Pg cDNA into SCC9 cells enables them to form desmosomes and induces a fibroblast to epidermoid transition [42-43]. However, unlike the effects of E-cadherin expression, these Pg-induced changes coincide with increased stability and a steady-state level of N-cadherin and decreased level and stability of β-catenin, without any significant effects on α-catenin. Part of this work is published in The Journal of proteome research [44].

The second part of this work is the relative quantification of protein mixtures extracted from human heart tissue samples. In chapter 2, it was stated that full characterization of human heart proteome is essential in understanding the protein expression alterations in failing heart tissues. A significant number of proteins from human heart tissue

have been detected and the goal of this work was to establish a comprehensive method to profile as many proteins from heart tissue as possible.

The other strategy for comparing changes in protein expression between normal and diseased heart tissue is to apply global stable isotope labeling combined with multidimensional liquid chromatography and mass spectrometry to complex protein mixtures from two control tissue extracts. In this work three different heart tissue samples from the same patient, taken at different stages of surgical intervention, are compared. Stable isotope dimethyl labeling strategy, similar to that used for squamous carcinoma cell line extracts but with a slightly different protocol, was used. Although trypsin is expected to cleave with equal propensity at lysine and arginine, it is reported that signals from arginine-containing peptides are generally stronger. It is postulated that the higher ionization efficiency of arginine-containing peptides is due to the very high basicity of the guanidine functionality in the side chains of arginine residues. Based on this, a recent focus of research has been directed toward increasing mass spectral signal intensities from lysine-containing peptides [45-47]. In this work, all the lysine side chains of the digested peptides of the tissue extracts were first blocked by guanidation and then followed by dimethyl labeling with formaldehyde reagent.

3.2 Experimental Section

3.2.1 Materials and Reagents

2, 5-dihydroxybenzoic acid (DHB), bovine trypsin, dithiothreitol (DTT), iodoacetamide, trifluoroacetic acid (TFA), and sodium dodecylsulfate (SDS) were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). d_0-formaldhyde (37% wt. % solution

in H_2O) and sodium cyanoborohydride were purchased from Sigma-Aldrich (Oakville, ON, Canada). d_2-formaldehyde (~20% w/w solution in deuterated water) was obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). HPLC grade acetone and acetonitrile were purchased from Fisher Scientific Canada (Edmonton, Canada). Water used in this experiment was obtained from a Milli-Q plus purification system (Millipore, Bedford, MA).

3.2.2 Protein Extraction and Sample Preparation

3.2.2.1 Squamous Carcinoma Cells

The human squamous carcinoma cells SCC9 are derived from carcinomas of the tongue (American Type Culture Collection (ATTC), Rockville, MD) and provided by Dr. M. Pasdar (Department of Cell Biology and Anatomy, University of Alberta, Edmonton, Alberta, Canada). The cell lines were maintained in minimum essential medium (MEM) (Sigma, St. Lous, MO) supplemented with 10% fetal bovine serum (FBS). The cells were prepared by Dr. M. Pasdar's research group. SCC9 cells were grown to about 80% confluency in 150 mm tissue culture dishes. The cells were washed three times with 150 mM sorbitol in 10 mM phosphate buffer (pH 7.0) prior to scraping with a rubber policeman and the cells pellets were stored in aliquots at -80 °C. Two transformations of the SCC9 cells were performed by Dr. Padar's group. SCC9 cells transfected with E-cad cDNA are referred to as SCC9-E-cad in this report. The SCC9-E-cad cells show decreased N-cadherin and increased catenin expression (compared to the control cell line SCC9). SCC9 cells transfected with Pg cDNA are referred to as SCC9-Pg cells in this study. This cell line shows increased N-cadherin and decreased β-catenin expression, the opposite of that observed for SCC9-E-cad.

The sample preparation of the squamous carcinoma cells for 2D liquid chromatography and MALDI MS are explained below. CytoBuster Protein Extraction Reagent (Novagen) was used to extract proteins from the cell lines. The amount of extraction reagent added was based on 150 μL/ 10^6 cells. The mixture was incubated while shaking at room temp. for 5 min. The extracts were centrifuged for 5 min at 15,000 x g (4 °C) and the supernatant was transferred into a new test tube. Protein concentration was estimated using the Bradford assay.

3.2.2.2 Human Heart Tissue

Human heart tissue samples were obtained from a person with heart disease and provided by Dr. Shaohua Wang (Cardiac Surgery Division, University of Alberta Hospital). Three different samples were obtained from the same patient. The first sample (sample 1) was taken prior to cardiopulmonary bypass (CPB) surgery. The second (sample 2) one was taken during CPB surgery, and the third (sample 3) after CPB surgery. The tissue was frozen in liquid nitrogen and stored at -80 °C.

Protein extraction was made with CelLytic^{MT} reagent. The tissue was cut into pieces and homogenized in CelLytic MT reagent (1 g of tissue/ 20 mL of reagent). The lysed sample was centrifuged for 10 min at 14,000 x g (4 °C). The protein-containing supernatant was transferred into a chilled test tube. Protein concentration was estimated using the Bradford assay. Reduction and alkylation was done using standard procedures. Salts and other small molecules were removed using a 3000 Da molecular weight cut-off membrane.

3.2.3 Protein Digestion

The digestion of the protein extracts from the squamous carcinoma cell lines was made using trypsin with a procedure similar to that described in chapter 2. In the case of protein extracts from heart tissue, two approaches to digestion were used. While attempting to denature the protein mixture at 95 °C, a small amount of protein precipitate was observed. For this reason, MAAH was used to digest proteins in addition to trypsin. The protein solution was reduced with dithiothreitol (DTT) and then alklyated with iodacetamide using standard procedures. The pH of the mixture was adjusted with 1 mM $NaHCO_3$ to ~8.5. 20 mM $CaCl_2$ was added to the mixture to a final concentration of 2 mM $CaCl_2$. Finally, trypsin was added to the protein solution at a protein: enzyme ratio of 40:1 by weight and the solution was incubated at 37 °C overnight. The undissolved part was suspended in aqueous solution followed by the standard reduction and alkylation steps. The mixture was then divided into three vials and 25 % TFA was added and microwave-assisted acid hydrolysis was carried out using the procedures described in chapter 2. The hydrolyzates were collected into one vial and dried in a vacuum centrifuge to remove the acid and then the dried sample was dissolved in 0.1% TFA. Finally both the peptides from the trypsin and MAAH digestions were mixed together.

3.2.4 Isotope Labeling

3.2.4.1 Dimethyl Labeling of Peptide Mixtures from Squamous Carcinoma Cell Lines

The same amount (~200 μg protein mixture) of each sample of the peptide mixtures obtained from the squamous carcinoma cell lines was labeled as follows. The sample volume was reduced to about 150 μL. 150 μL 0.2M acetate buffer was added into each vial

(pH = 6) containing SCC9, SCC9-E-cad and SCC9-Pg respectively. The mixtures were vortexed and 1M sodium cyanoborohydride was added to each vial (4 x volume of formaldehyde used). The mixture was vortexed and mixed either with d_0-formaldehyde or d_2-formaldehyde (4% (w/w) in water, 5 μL). In this particular experiment SCC9 was labeled with d (0) and SCC9-E-cad and SCC9-Pg were labeled with d (2). The mixtures were vortexed and incubated at 37 °C for 3 h while shaking. 15 μL 1M $NaHCO_3$ was added to each vial and vortexed for 1 h at room temp. 10 μL 1M NH_4HCO_3 was added to each vial, vortexed and incubated at room temp for 30 minutes. Finally each sample mixture was acidified by adding 10% TFA.

3.4.4.2 Dimethyl Labeling of Peptide Mixtures from Tissue Extracts

Only the tissue samples were studied under this category. The first sample which was taken prior to cardiopulmonary bypass (CPB) surgery was considered as sample 1 in this study. The second, taken during CPB surgery, as sample 2, and the third, after CPB surgery, as sample 3. About 250 μg of protein mixture of each sample was taken for the analysis. The ε-amino groups of all lysine were blocked by adding 100 μL 2M O-methylisourea in 100 mM $NaHCO_3$, adjusting the pH to over 10.5 with 2M sodium hydroxide and incubating the resulting mixture at 65 °C for 10 min (guanidination). The reaction was stopped and the pH adjusted to 8 by adding 10% TFA. Then, the guanidinated peptide solution was mixed with 10 μL 1M sodium cyanoborohydride. The mixture was vortexed and mixed with d_0, ^{12}C-formaldehyde or d_2, ^{13}C-formaldehyde (4% (w/w) in water, 3 μL). The mixtures were vortexed, incubated at 37 °C for 2 h and acidified by adding TFA. In this study, sample 1 was labeled with d_2, and samples 2 and 3 were labeled with d_0.

3.2.5 2D LC MS and MS/MS

For the squamous carcinoma cell lines study, labeled SCC9 peptide mixtures were combined with the same amount of SCC9-E-cad and SCC9-Pg cell lines, respectively. The labeled tissue samples were mixed as explained below. Labeled peptide mixtures of Sample 1 (considered as a control) were combined with the same amount of sample 2 and sample 3, respectively. For the tissue samples, the combined sample was further desalted using a cartridge. The mixture was acidified with 0.1% TFA and ACN was added to a final concentration of 20% ACN. Each combined labeled peptide mixture was loaded onto a strong cation exchange column (2.1 x 150 mm). 10 fractions were collected from the squamous carcinoma cell lines and the tissue samples when the column was eluted with the solvent gradient described in the 2DLC-MALDI MS/MS section of chapter 2 above. All collected fractions were reduced in volume to ~0.01 ml to remove ACN. Each fraction was injected onto a capillary C18 column (Vydac, 1.0 x 150 mm). Gradient elution was performed with solvent A (0.1% TFA and 4% ACN in water) and B (0.1% TFA in ACN) at a flow rate of 40 μL/min. Each fraction was collected every one minute and directly deposited onto a 100 well gold MALDI plate using the heated droplet interface. 1μL of 1M DHB in 50% acetonitrile was deposited onto each sample. Peptides were analyzed by a QSTAR MALDI MS/MS mass spectrometer (MDS SCIEX, Concord, Canada).

3.2.6 Data Processing

Peptide sequences were automatically identified by database searching of the MS/MS spectra against the Swiss-Prot database using the MASCOT algorithm

(http://www.matrixscience.com). For the proteins from the squamous carcinoma cell lines, MS/MS data were searched twice; in one case with d_0 dimethyl labeling and in the other with d_2 dimethyl modification, using trypsin as the enzyme requirement. In the case of tissue extracts, MS/MS data were searched four times as follows: with no enzyme requirement (no enzyme), with trypsin as enzyme and for each case with d_0 and d_6 dimethyl-N-terminus modification.

3.3 Results and Discussion

Figure 3.1 shows the workflow for protein quantification and identification using stable isotope labeling, LC-MALDI MS and MS/MS. Two control protein mixture extracts were initially digested with trypsin or MAAH and then labeled with either the light or heavy forms of formaldehyde for dimethylation (in the case of the tissue samples, the ε–amino group of the lysine group was first blocked by guanidination). The labeled peptide mixtures were then combined. SCX chromatography was used to fractionate the complex labeled peptide mixtures (but the labeled tissue samples were desalted before fractionation). Each fraction was further separated by RP-LC and collected onto a gold MALDI target using the heated droplet LC-MALDI interface. A matrix (DHB) was added to each spot on the target. Peptide pairs with significant differences were selected from the MALDI MS spectra and MALDI MS/MS analysis and sequence database searching was subsequently done on the peptide pairs to identify the protein pairs with the significant difference.

3.3.1 Dimethyl Labeling

Formaldehyde reacts with both the N-terminus and the ε–amino groups of lysine residues of a peptide to form Schiff bases, which are then reduced with sodium cyanoborohydride to form secondary amino groups that are relatively more reactive than their original primary amino groups. Subsequently, each of these more-reactive species reacts with another formaldehyde unit and is then reduced to form the end product that bears dimethyl-substituted tertiary amino groups [35-36]. The reaction mechanism is shown in Figure 3.2.

As mentioned in the experimental section, the tissue samples were labeled using a different procedure. In this case the ε–amino groups of lysine were first blocked by guanidination by reaction with O-methylisourea, and then followed by reductive amination of the N-terminal. However, d_2, ^{13}C-formaldehyde was used as heavy isotope reagent instead of d_2, ^{12}C-formaldehyde reagent. As a result the value of m/z for this isotopic pair differs by 6 mass units.

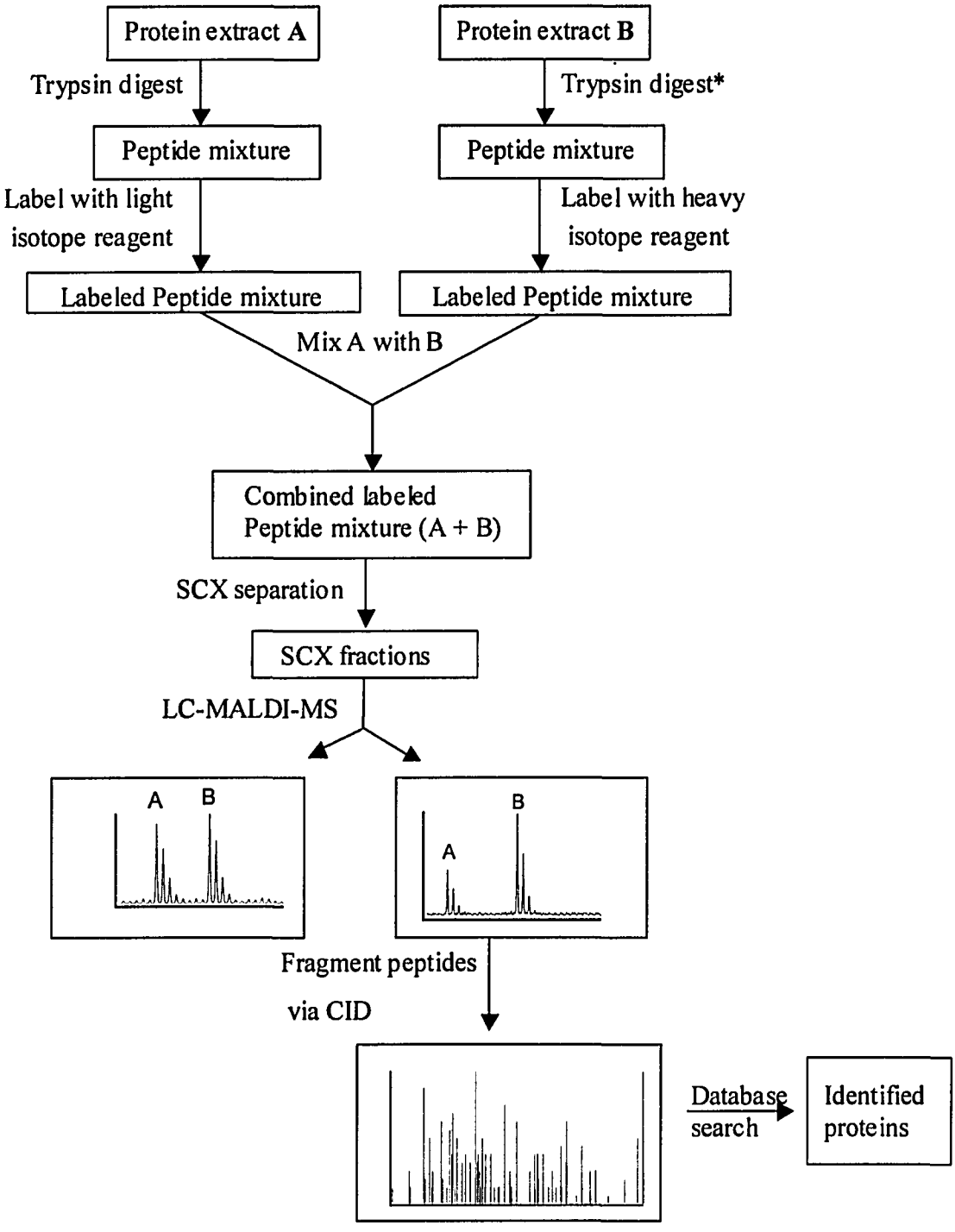

Figure 3.1 A flow chart of the experiment. * incase of the tissue sample, MAAH also used for digestion

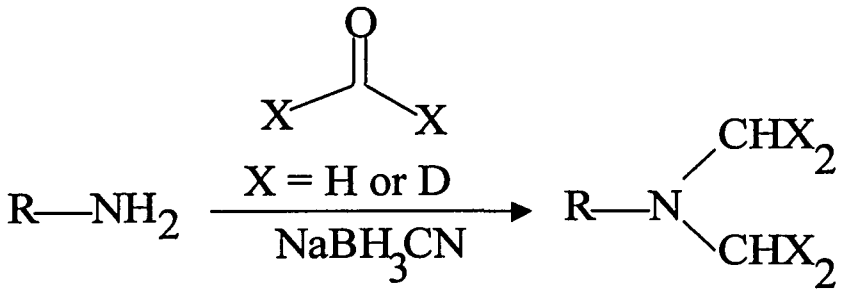

Figure 3.2 Dimethyl labeling reaction scheme.

3.3.2 HPLC Separation of Peptide Mixtures

Global isotope labeling of protein mixtures produces a large number of peaks, which requires a relatively pure sample or greater separation power to resolve these peaks. In this experiment, two dimensional liquid chromatography was used to simplify the complex nature of the sample mixtures. The combined labeled peptide mixtures were first fractionated using SCX. which also served to remove SDS which had been used to solubilize squamous carcinoma cell line extracts. In each combination, ten fractions were collected every one minute during a 40 min gradient as described in chapter 2.

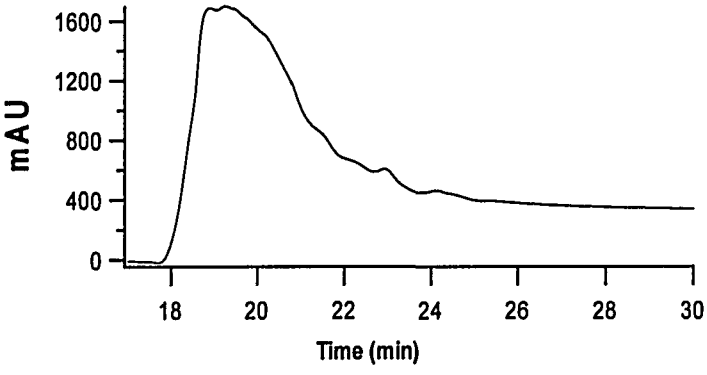

Figure 3.3 UV Chromatogram of cation exchange separation of labeled tryptic digest of SCC9 + SCC9-E-cad protein mixture.

The UV-absorbance chromatogram of the cation exchange separation of the tryptic digest from the squamous carcinoma cell line extract (SCC9 and SCC9-E-cad) is shown in Figure 3.3.

The heart tissue protein extracts were digested using trypsin and MAAH. The peptide mixtures from the tryptic and MAAH digestions were combined and labeled after a guanidation step using the dimethyl labeling procedure explained in the experimental section, and then fractionated using SCX. The UV-absorbance chromatogram of the cation exchange separation of the tryptic and MAAH digest from the tissue extract (sample 1 & 2) is shown in Figure 3.4. Each fraction collected from the SCX column was injected into an RP-LC column and deposited onto a gold MALDI target every one minute at a flow rate of 40 μL/min. As an example, the first three reversed-phase HPLC chromatograms of the squamous carcinoma cell lines (SCC9 & SCC9-E-cad, SCC9 & SCC9-Pg) obtained from SCX fractions 2, 3 and 4 respectively are shown in Figure 3.5 and 3.6 respectively. The HPLC chromatograms of the tissue extracts (sample 1&2 and sample 1&3) obtained from SCX fractions of 1, 2 and 3 respectively are also shown in Figure 3.7 and Figure 3.8, respectively.

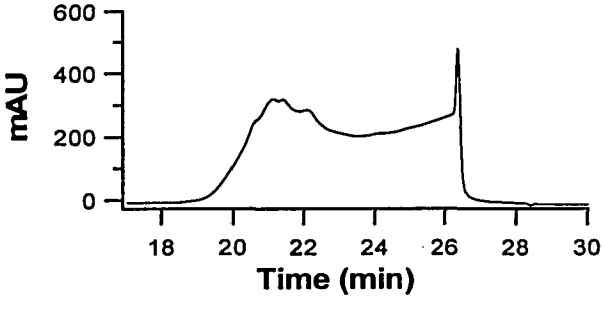

Figure 3.4 UV chromatogram of cation exchange separation of the labeled tryptic and MAAH digest of sample 1 +2 protein mixture.

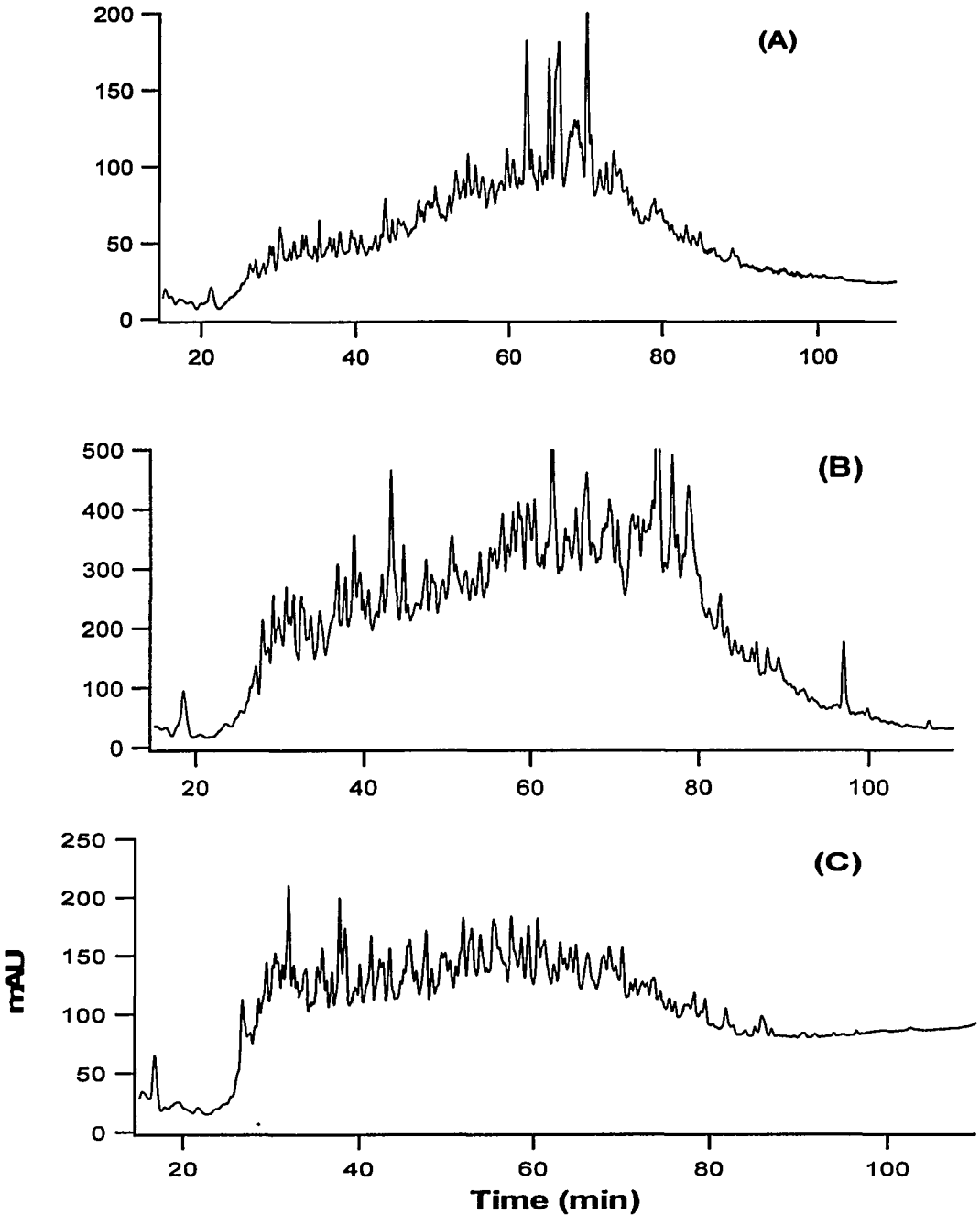

Figure 3.5 Reversed phase HPLC chromatogram of SCC9 + SCC9-E-cad obtained from SCX fractions 2 (A), 3 (B) and 4 (C), respectively.

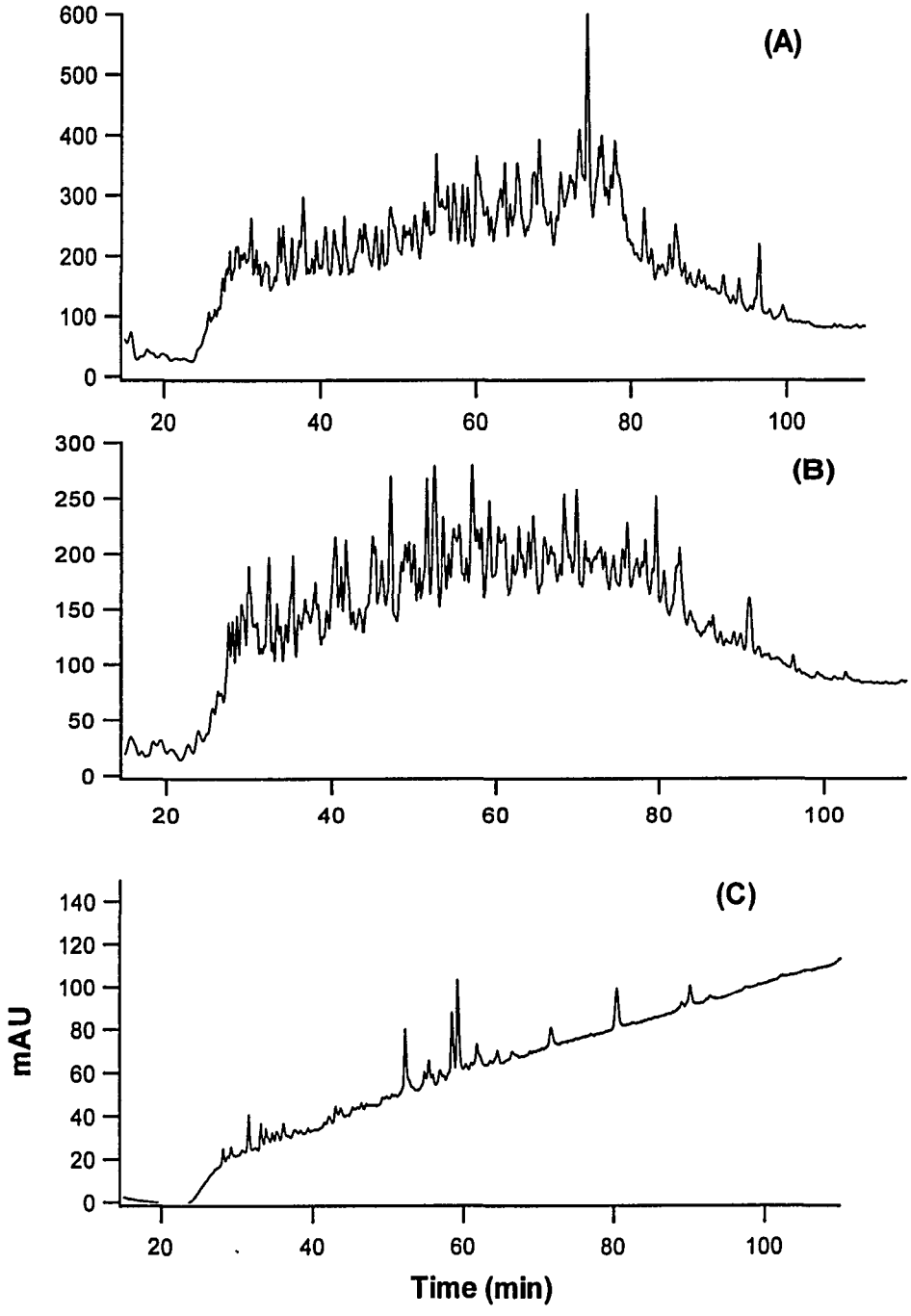

Figure 3.6 Reversed phase HPLC chromatogram of SCC9 + SCC9-Pg obtained from SCX fractions 2 (A), 3 (B) and 4 (C), respectively.

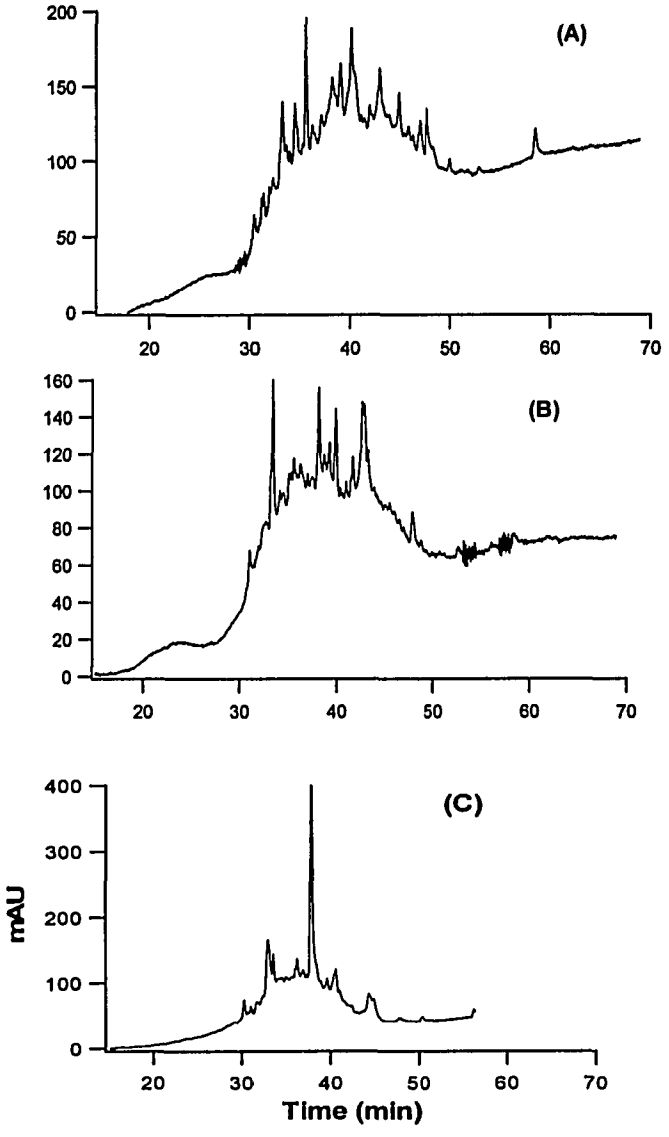

Figure 3.7 Reversed phase HPLC chromatogram of tissue sample 1 + sample 2 obtained from SCX fraction 1 (A), 2 (B) and 3 (C), respectively.

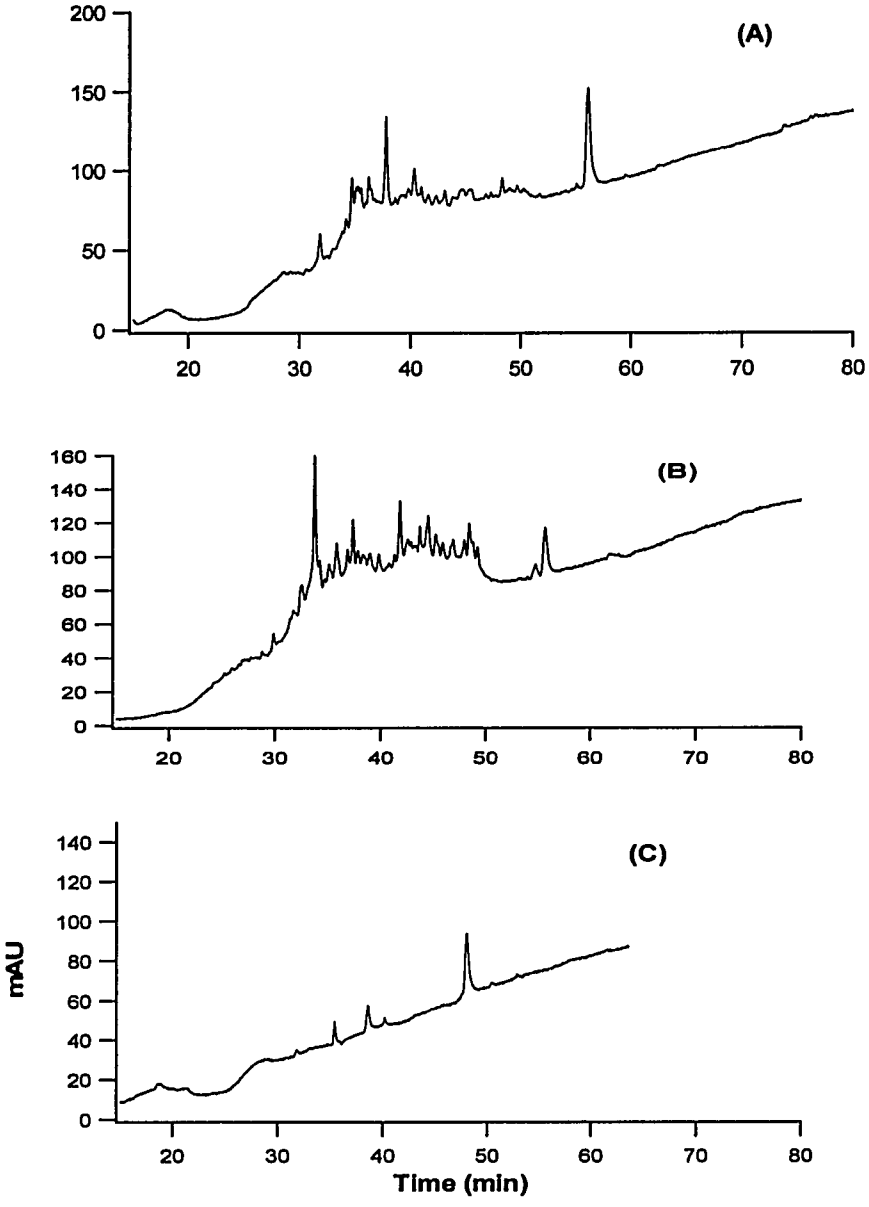

Figure 3.8 Reversed phase HPLC chromatogram of tissue sample 1 + sample 3 obtained from SCX fraction 1 (A), 2 (B) and 3 (C), respectively.

3.3.3 Quantification of Proteins in Squamous Carcinoma Cells using MALDI MS

The importance of the dimethyl labeling method for quantitative analysis using MALDI-MS, based on the presence of stable isotopes, has been reported in the literature[35-37, 44].

In this work, the dimethyl labeling method has been applied to quantify proteins in squamous carcinoma cell lines. More than five thousand pairs were detected from all MALDI-MS spectra collected. However, only about 300 pairs showed relative abundance ratios of greater than 2. In this method, the expected mass difference between the pairs is 4n (where n could be 1, 2, 3, 4 etc) depending on the number of lysine present in the peptide. If the peptide has no lysine residues, the expected m/z difference should be 4 as shown in Figure 3.9A and with a single lysine residue the m/z difference is 8 (Figure 3.9 B). Relative quantification determination of protein levels in the mixture is made by comparing the intensity ratios between the mono isotopic pair.

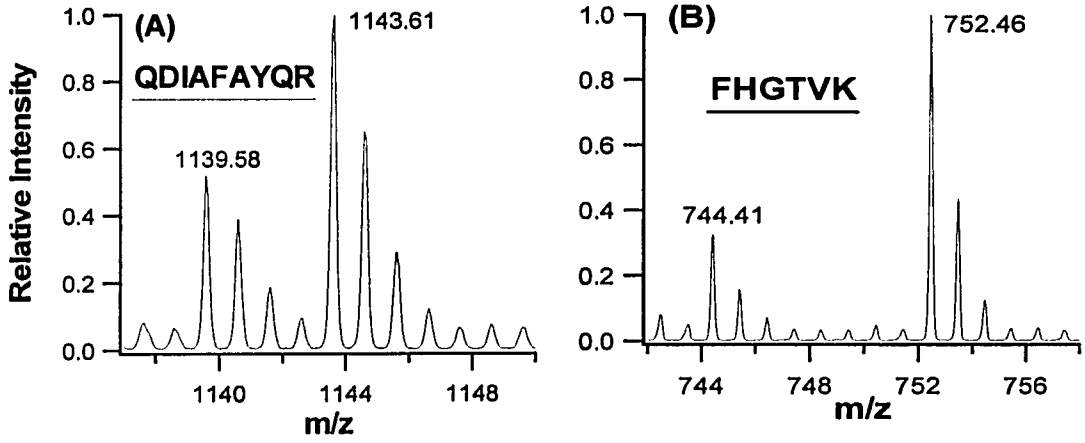

Figure 3.9 MALDI MS Spectra showing 4 Da (A) and 8 Da (B) mass differences respectively.

3.3.4 Quantification of Proteins in Tissue Samples using MALDI MS

The compatibility of dimethyl labeling with 2D-LC was reported recently by our group[44]. The reagents used in dimethyl labeling are inexpensive and also commercially available. The derivatization procedure for dimethyl labeling is relatively fast and simple when compared to some of the other methods. The ionic state is not changed significantly by dimethyl modification, and so the ionization efficiency of the fragment is more likely to be conserved. With this method the dimethyl modification is a global labeling that labels not only lysine residues but also the N-terminus of the peptide, without significant isotopic effects. However, as a result of the multiple labeling, searching the peptide pairs with mass differences of 4, 8, 12, 16, 20, etc in the MALDI spectrum manually could be time consuming.

However, the focus of the second work is on the relative quantification of protein mixtures extracted from heart tissue samples. The procedure used in this method is similar to that used for the relative quantification of squamous carcinoma cell lines. In this method, lysine side chains were blocked by guanidation prior to N-terminal labeling to prevent the incorporation of multiple labels. The N-terminus of all peptides in a digested sample was specifically labeled either with light or heavy formaldehyde reagent through the reductive amination reaction. The expected mass difference between the peptide pairs is only 6 Da. The labeling procedure of this experiment was carried out by Chengjie Ji (from our group).

Trypsin was not the only digestion technique used in this work, microwave-assisted acid hydrolysis (MAAH) was also used to digest some of the insoluble proteins. Figure 3.10 shows the MALDI spectrum of peptide pair obtained as the result of the MAAH.

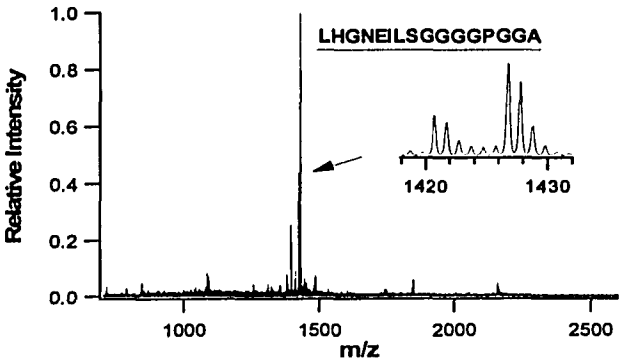

Figure 3.10 MALDI MS spectrum obtained from SCX fraction 1 of combined sample 1 and 2 at 37 min. as the result of the MAAH. The inset shows the expanded view of the peptide pair.

Overall, over 2000 peptide pairs were detected from all the SCX fractions of a single combination. Of these, 160-216 pairs show significant differences, that is, in excess of a two-fold difference.

3.3.5 Protein Identification Using MALDI QqTOF TOF MS/MS

Once the peptide pairs with significant differences were detected from the MALDI spectrum, the next step was to fragment those peptides using MALDI MS/MS to identify the proteins. Only peptides that showed a difference in relative abundance were targeted for identification by tandem MS. Each MS/MS spectrum was collected manually using MALDI Qq-TOF and searched against the database. The introduction of E-cad and Pg expression produced significant changes in the protein expression in SCC9-E-cad and SCC9-Pg cells. Table 3.1 shows the list of proteins identified from the combination of SCC9 and SCC9-E-cad cell lines. E-cadherin expression in SCC9 cells led to the downregulation of 10 (ratio >2) and upregulation of 43 proteins (ratio <1), as shown in Table 3.1. Among the down regulated proteins are prothymosin alpha, DNA replication licensing factor MCM7, and

deoxycytidine kinase and nucleosome assembly protein 1-like 1. These proteins are directly involved with cell proliferation and have been shown to be up regulated in various tumors or carcinoma cell lines [44]. Some of the other down regulated proteins such as peroxiredoxin 2, T-complex protein 1, epithelial protein lost in neoplasm, and long-chain-fatty-acid-CoA ligase 4 are involved in actin polymerization, protein folding or metabolic and oxidative pathways, and exhibit increased expression in tumors and transformed cells. The reduced expression of these proteins is consistent with decreased growth and the induction of an epidermoid morphology in SCC9 cells expressing E-cadherin. The upregulated proteins include collagen or collagen-binding precursor proteins involved in cell-matrix adhesion, a number of cytoskeletal and actin binding proteins, such as cortactin, transgelin, and caldesmon. The list of all the up regulated proteins with a ratio of SCC9/SCC9-E-cad of less than 1 is shown in Table 3.1. The expression of six proteins between the two cells lines was validated by immunoblotting and immunofluorescence microscopy[44] (result not shown).

Introduction of plakoglobin plays a role in mediating cell-cell adhesion and is associated with transcription factors that induce changes in the expression of genes involved in cell fate determination and proliferation. In this work, plakoglobin expression in the SCC9 cells led to the down regulation of 50 and up regulation of 32 proteins, respectively, as shown in Table 3.2. Unlike the E-cadherin expression, most of the proteins are found to be down regulated proteins. However, four of the down regulated proteins: epithelial protein lost in neoplasm, DNA replication licensing factor MCM7, prothymosin alpha and nucleosome assembly protein 1-like 1, which are listed as down regulated proteins in E-cadherin expression cells are also observed as down regulated proteins in the combination of SCC9 and SCC9-Pg cell lines.

Some significant protein changes were detected in the initial work on the relative quantification of tissue samples. The tissue samples were taken from the same patient at different surgical operation stages (as explained in the experimental section). About 32 common proteins with significant changes were detected between the tissue sample taken prior to CPB and that taken during CPB. Also, changes in 24 proteins were detected between the tissue sample taken prior to CPB and that taken after CPB. Tables 3.3 and 3.4 show the list of proteins and ratios of isotope pairs of peptides between sample 1 and the two other samples (2 and 3).

Table 3.1 List of Identified Proteins Differentially Expressed between SCC9 and SCC9-E-cad Cells

Name	Access ID	Unique Peptide Sequence	SCC9/ SCC9-E-cad ratio	Average ratio
Annexin A2	P07355	SVPHLQK	0.492	0.503
		AYTNFDAER	0.601	
		QDIAFAYQR	0.516	
		SALSGHLETVILGLLK	0.457	
		LSLEGDHSTPPSAYGSVK	0.501	
		AYTNFDAERDALNIETAIK	0.449	
Glyceraldehyde-3-phosphate dehydrogenase, liver	P04406	FHGTVK	0.306	0.345
		LTGMAFR	0.35	
		VVDLMAHMASKE	0.353	
		LISWYDNEFGYSNR	0.33	
		VIISAPSADAPMFVMGVNHEK	0.389	
4F2 cell-surface antigen heavy chain	P08195	HWDQNER	0.401	0.364
		WWHTGALYR	0.349	
		IKVAEDEAEAAAAAK	0.335	
		IGDLQAFQGHGAGNLAGLK	0.392	
		VKDALEFWLQAGVDGFQVR	0.354	
		SLLHGDFHAFSAGPGLFSYIR	0.357	
Collagen-binding protein 2 precursor	P50454	GVVEVTHDLQK	0.151	0.165
		LQIVEMPLAHK	0.18	
		LFYADHPFIFLVR	0.165	
		TGLYNYYDDEKEK	0.164	
		LSSLIILMPHHVEPLER	0.152	

		QHYNCEHSK	0.206	
		KPAAAAAPGTAEK	0.101	
		TDGALLVNAMFFKPHWDEK	0.163	
		AVLSAEQLRDEEVHAGLGELL R	0.202	
Glyceraldehyde-3-phosphate dehydrogenase, muscle	P00354	FHGTVK	0.304	0.288
		LTGMAFR	0.357	
		VGVDGFGR	0.14	
		VVDLMAHMASKE	0.358	
		LEKPAKYDDIKK	0.279	
Ribosome-binding protein 1	Q9P2E9	KAEGAPNQGR	0.352	0.291
		KAEGTPNQGK	0.259	
		KVEGAQNQGK	0.261	
Clathrin heavy chain 1	Q00610	LEKHELIEFR	0.502	0.495
		RKDPELWGSVLLESNPYR	0.488	
Epithelial protein lost in neoplasm	Q9UHB6	GNYDEGFGHRPHK	2.96	2.73
		SKGNYDEGFGHRPHK	2.51	
Transferrin receptor protein 1	P02786	HVFWGSGSHTLPALLENLK	0.471	0.463
		EAGSQKDENLALYVENQFR	0.456	
Putative RNA-binding protein 3	P98179	YYDSRPGGYGYGYGR	0.459	0.423
		GFGFITFTNPEHASVAMR	0.387	
Calreticulin precursor	P27797	KVHVIFNYK	0.471	0.490
		KIKDPDASKPEDWDER	0.509	
Peroxiredoxin 2	P32119	KEGGLGPLNIPLLADVTR	2.88	
Ribonucleoprotein A0	Q13151	EDSARPGAHAK	0.383	
Myosin heavy chain, type A	P35579	THEAQIQEMR	0.29	0.41
		TRLQQELDDLLVDLDHQR	0.53	
Procollagen-lysine	Q02809	LTHYHEGLPTTR HTLGHLLSLDSYR	0.365 0.36	0.362
Catenin delta-1	O60716	GIPVLVGLLDHPK	0.539	
Heat shock protein HSP 90-alpha	P07900	QKAEADKNDK	2.29	5.86
		RAPFDLFENRK	9.44	
Caldesmon	Q05682	EAEGAPQVEAGKR	0.242	0.265
		QKEFDPTITDASLSLPSR	0.288	
Ubiquitin-like protein NEDD8	Q15843	ILGGSVLHLVLALR	0.492	
Triosephosphate isomerase	P60174	TATPQQAQEVHEK	0.468	
Treacle protein	Q13428	ELLPLIYHHLLR	0.467	
DNA replication licensing factor MCM7	P33993	LAQHITYVHQHSR	2.36	
T-complex protein 1	P50990	ELEVQHPAAK	2.11	
L-type aminoacid transporter 1	Q01650	ALAAPAAEEKEEAR	0.343	
Prothymosin alpha	P06454	AAEDDEDDDVDTKK	2.09	
Prolyl 4-hydroxylase alpha-2 subunit precursor	O15460	KGTAVFWYNLLR	0.104	0.126
		SQVLDYLSYAVFQLGDLHR	0.149	
Tubulin beta-4 chain	Q13509	LHFFMPGFAPLTAR	0.45	
Caveolin-1	Q03135	YVDSEGHLYTVPIR	0.485	

Deoxycytidine kinase	P27707	HESWLLHR	2.49	
EH-domain containing protein 2	Q9NZN4	LLEALDEMLTHDIAK	0.444	
Alpha-1 catenin	P35221	EKQDETQTK	0.294	
Enabled protein homolog	Q8N8S7	VHIYHHTGNNTFR	0.323	
Transducin â-like 2 protein	Q9Y4P3	EKPQQHNFTHR	0.391	
Long-chain-fatty-acid--CoA ligase 4	O60488	AKPTSDKPGSPYR	2.37	
Golgi apparatus protein 1 precursor	Q92896	EAEEREPK	0.248	
Mitotic checkpoint serine/threonine-protein kinase BUB1	O43683	MGPSVGSQQELR	12.9	
Collagen alpha 1(I) chain precursor	P02452	GDKGETGEQGDR YHDRDVWKPEPCR	0.018 0.103	0.06
Carbonic anhydrase IX precursor	Q16790	QLHTLSDTLWGPGDSR FPAEIHVVHLSTAFAR	0.23 0.045	0.137
Transgelin	Q01995	HVIGLQMGSNR KYDEELEER	0.21 0.108	0.159
Kinectin	Q86UP2	TAEHEAAQQDLQSK	0.227	
Peroxiredoxin 4	Q13162	IPLLSDLTHQISK	0.465	
Prolyl 4-hydroxylase alpha-1 subunit precursor	P13674	FILAPAKQEDEWDKPR	0.276	
Collagen alpha 1(V)	P20908	GVQGPPGPAGKPGR	0.139	
Lysyl hydroxylase 2	O00469	VVFAADGILWPDKR	0.296	
Collagen alpha 1(III)	P02461	KHWWTDSSAEKK	0.202	
Annexin A6	P08133	GLGTDEDTIIDIITHR	0.200	
Src substrate cortactin	Q14247	HESQQDYSK	0.345	
Nucleophosmin	P06748	MTDQEAIQDLWQWR DSKPSSTPR		0.392
Tropomyosin beta chain	P09493	AEQAEADKK	0.433	
Tropomyosin 1 alpha chain	P07951			
Nucleosome assembly protein 1-like 1	P55209	AKIEDEKKDEEKEDPK	2.32	
Protein kinase C, alpha type	P17252	EHAFFR	0.464	
Protein kinase C, beta type	P05771			

IgG receptor FcRn large subunit p51	P55899	LREHLEK	0.309	
Zinc finger CCHC domain containing protein 2	Q9C0B9			

Table 3.2 List of Identified Proteins Differentially Expressed between SCC9 and SCC9-Pg Cells

Name	Access ID	Unique Peptide Sequence	SCC9/ SCC9-Pg ratio	Average ratio
Ezrin	P15311	KENPLQFK	2.542	2.380
		EKEELMLR	2.462	
		DDRNEEKR	2.044	
		ITEAEKNER	2.434	
		FVIKPIDKK	2.372	
		IQVWHAEHR	2.284	
		APDFVFYAPR	2.116	
		AQEEAERLEADR	2.563	
		KEDEVEEWQHR	2.167	
		RKPDTIEVQQMK	2.020	
		RKEDEVEEWQHR	2.579	
		AKEELER	2.974	
Tubulin beta-1 chain	P07437	YLTVAAVFR	1.999	2.367
		KLAVNMVPFPR	2.890	
		LHFFMPGFAPLTSR	2.345	
		EIVHLQAGQCGNQIGA K	2.235	
Talin 1	Q9Y490	KIFQAHK	2.646	2.469
		QKLHTDDELNWLDHG R	2.567	
		LASEAKPAAVAAENEE IGSHIK	2.193	
Epithelial protein lost in neoplasm.	Q9UHB6	EDKPAETKK	2.677	2.737
		ASSQQEKEDKPAETK	2.976	
		SKGNYDEGFGHRPHK	2.901	
		HEVEKSEISENTDASG K	2.392	
Brain acid soluble protein 1	P80723	EKPDQDAEGK	0.115	0.188
		AEPEKTEGAAEAK	0.205	
		EGEKDAAAAKEEAPK	0.206	
		KTEAPAAPAAQETK	0.214	
		KAEGAATEEEGTPK	0.223	
		EKPDQDAEGKAEEK	0.163	

Radixin	P35241	KENPLQFK	2.542	2.263
		FVIKPIDKK	2.372	
		APDFVFYAPR	2.116	
		RKPDTIEVQQMK	2.020	
Tropomyosin alpha 3 chain	P06753	ALKDEEKMELQEIQLK	2.670	2.472
		YSQKEDKYEEEIK	2.274	
Annexin A1	P04083	KALTGHLEEVVLALLK	2.948	2.507
		AAYLQETGKPLDETLK K	2.066	
Alpha-actinin 1	P12814	KTFTAWCNSHLR	2.413	2.318
		ILAGDKNYITMDELR	2.129	
		KAGTQIENIEEDFRDG LK	2.413	
Vimentin	P08670	FANYIDKVR	2.420	2.391
		ILLAELEQLKGQGK	2.362	
Nucleosome assembly protein 1-like 1	P55209	GIPEFWLTVFK	2.578	3.014
		KYAVLYQPLFDKR	2.824	
		AKIEDEKKDEEKEDPK	3.640	
Annexin A2	P07355	SVPHLQK	0.252	0.320
		SALSGHLETVILGLLK	0.279	
		WISIMTER	0.431	
Clathrin heavy chain 1	Q00610	EHLELFWSR	0.206	0.272
		KVGYTPDWIFLLR	0.262	
		KFDVNTSAVQVLIEHIG NLDR	0.347	
Early endosome antigen 1	Q15075	AAVEQEKR	2.785	2.789
		HNEESVSKK	2.792	
Rho GDP-dissociation inhibitor 1	P52565	EGVEYR	4.279	4.693
		IDKTDYMVGSYGPR	5.139	
		YIQHTYR	4.659	
Nucleophosmin	P06748	ADKDYHFK	0.393	0.417
		MTDQEAIQDLWQWR	0.440	
60S ribosomal protein L6	Q02878	AGEKVEKPDTK	2.903	2.903
Filamin A	P21333	KIQQNTFTR	1.99	2.094
		DAPQDFHPDRVK	2.199	
H14_HUMAN	P10412	AASGEAKPK	3.796	3.434
		KAASGEAKPK	3.072	
Colligin 1 Colligin 2	P29043 P50454	LFYADHPFIFLVR	0.155	0.179
		TGLYNYYDDEKEK	0.226	
		LSSLIILMPHHVEPLER	0.157	
Calreticulin precursor	P27797	FYGDEEKDKGLQTSQ DAR	2.506	2.506
Transferrin receptor protein 1	P02786	HVFWGSGSHTLPALL ENLK	0.335	0.377
		RLYWDDLK	0.419	
60S ribosomal protein L18a	Q02543	SRFWYFVSQLK	0.445	0.463
		SRFWYFVSQLKK	0.482	
Heat shock cognate 71 kDa protein	P11142	MVQEAEKYKAEDEK	3.699	

Neutral alpha-glucosidase AB precursor	Q14697	YSLLPFWYTLLYQAHR	0.353	0.296
		ALLDSLQLGPDSLTVH LIHEVTK	0.253	
		DGDKPEETQGK	0.281	
Filamin C	Q14315	SSHTYTR	3.372	
60S ribosomal protein L10 60S ribosomal protein L10	P27635	GPLDKWR	2.828	3.289
		LHPFHVIR	3.751	
Elongation factor 1-alpha 1 EF12_HUMAN	P04720 Q05639	TIEKFEK	2.167	2.565
		EHALLAYTLGVK	2.963	
		GSFKYAWVLDKLKAE RER		
ACTA_HUMAN ACTH_HUMAN ACTS_HUMAN ACTG_HUMAN	P62736 P12718 P02568 P02571	RGILTLK	2.871	3.053
		AVFPSIVGRPR	2.769	
		DSYVGDEAQSKR	3.519	
Prolyl 4-hydroxylase alpha-2 subunit precursor	O15460	KGTAVFWYNLLR	0.410	0.430
		YHHGNR	0.451	
40S ribosomal protein S10	P46783	HFYWYLTNEGIQYLR	0.373	
40S ribosomal protein S20	P60866	DTGKTPVEPEVAIHR	2.3434	
28 kDa heat- and acid-stable phosphoprotein	Q13442	AREEEEQKEGGDGAA GDPK	2.386	
Splicing factor arginine/serine-rich 11	Q05519	KSESDKDVK	3.425	
Nonhistone chromosomal protein HMG	P05204	LSAKPAPPKPEPKPK	2.654	
Myosin heavy chain nonmuscle type A	P35579	AGKLDPHLVLDQLR	2.870	
Ubiquitin thiolesterase L3	P15374	VTHETSAHEGQTEAP SIDEK	2.105	
Treacle protein	Q13428	ELLPLIYHHLLR	0.382	
Prothymosin alpha	P06454	AAEDDEDDDVDTKK	2.710	
Myosin regulatory light chain 2	P24844	NAFACFDEEASGFIHE DHLR	0.341	
HLA class I histocompatibility antigen	P10316	HKWEAAHVAEQLR	0.461	
Complement component 1	Q07021	AFVDFLSDEIKEER	0.395	
Dynein heavy chain, cytosolic	Q14204	IWAHEALR	0.435	
Calnexin precursor	P27824	EIEDPEDRKPEDWDE RPK	2.198	
Thymosin beta-4	P62328	KTETQEKNPLPSKETI EQEK	4.925	

Actin-like protein 3	P61158	LKPKPIDVQVITHHMQR	3.366	
Triple functional domain protein	O75962	IVFGNIHQIYDWHR	0.289	
Alpha enolase	P06733	KLNVTEQEKIDK	2.479	
Eukaryotic translation initiation factor 2 subunit 1	P05198	YKRPGYGAYDAFK	4.127	
Eukaryotic translation initiation factor 3 subunit 6	P60228	HLVFPLLEFLSVK	2.210	
Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	EDTEEHHLRDYFEEYGK	3.404	
26S proteasome non-ATPase regulatory subunit 2	Q13200	EKEEDKDKK	3.534	
Translationally controlled tumor protein	P13693	GKLEEQRPER	2.438	
ERO1-like protein alpha precursor	Q96HE7	MLLLEILHEIK	0.237	
Thioredoxin	P10599	MIKPFFHSLSEK	2.278	
40S ribosomal protein S4, X isoform	P62701	ECLPLIIFLR	0.389	
DNA replication licensing factor MCM7	P33993	LAQHITYVHQHSR	2.665	
Eukaryotic translation initiation factor 5A	P10159	EDLRLPEGDLGKEIEQK	2.479	
DNA replication licensing factor MCM4	P33991	THIDVIHYR	2.310	
80 kDa nuclear cap binding protein	Q09161	LFVWEILHSTIR	0.393	
Cyclin-dependent kinases regulatory subunit 2	P33552	QIYYSDKYFDEHYEYR	3.693	
ARP2/3 complex 34 kDa subunit	O15144	ARPDAEKK	2.009	
Ornithine aminotransferase, mitochondrial precursor	P04181	FAPPLVIKEDELR	3.084	
Lupus La protein	P05455	YKETDLLILFKDDYFAK	2.468	
T-complex protein 1, eta subunit	Q99832	KADKVEQR	2.637	
4F2 cell-surface antigen heavy chain	P08195	VKDALEFWLQAGVDGFQVR	0.380	
MR11_HUMAN	P49959	DIIHFFR	0.359	
Citrate synthase, mitochondrial precursor	O75390	ALGVLAQLIWSR	0.348	

apurinic or apyrimidinic site	P27695	QRWDEAFR	0.357	
26S proteasome non-ATPase regulatory subunit 4	P55036	DKKEEDKK	2.731	
Glutathione transferase omega 1	P78417	NKPEWFFK	2.665	
Laminin beta-1 chain precursor	P07942	KYEDNQR	2.804	
Cell division control protein 2 homolog	P06493	AFGIPIR	2.227	
Ubiquitin carboxyl-terminal hydrolase 11	P51784	QAWSGHHR	0.461	
Golgi-associated plant pathogenesis-related protein 1	Q9H4G4	AHNEYR	0.433	
General vesicular transport factor p115	O60763	LEVGIQAMEHLIHVLQTDR	0.265	
Splicing factor 3 subunit 1	Q15459	KKEEEEKEKER	3.240	
Protein transport	Q15436	FGEYHKDDPSSFR	0.334	
HLA class I histocompatibility antigen	P30481	THVTHHPISDHEVTLR	0.233	
TBA8_HUMAN TBA6_HUMAN TBA1_HUMAN TBA4_HUMAN TBA2_HUMAN	Q9NY65 Q9BQE3 P05209 P05215 Q13748	QLFHPEQLITGKEDAANNYAR	3.881	
Nucleoside diphosphate kinase B	P2239	DRPFFPGLVK	0.227	

Table 3.3 List of Identified Proteins Differentially Expressed between Heart Tissue Samples of 1 and 2

Name	Access ID	Unique Peptide Sequence	Sample 2/ Sample 1 ratio	Average ratio
Hemoglobin alpha chain	P69905	TYFPHFDL	0.369	0.423
		LRVDPVNFK	0.445	
		TYFPHFDLSHG	0.353	
		AHAGEYGAEALER	0.425	
		VGAHAGEYGAEALER	0.421	
		TYFPHFDLSHGSAQVK	0.519	
		MFLSFPTTK	0.405	

		TYFPHFDLSH	0.451	
Hemoglobin beta chain	P68871	SAVTALWGK	0.505	0.389
		VHLTPEEK	0.45	
		LLVVYPWTQR	0.286	
		VVAGVANALAHKYH	0.316	
10 kDa heat shock protein	P61604	FLPLFDR	2.10	
Serotransferrin	P02787	YLGEEYVK	0.402	0.394
		HQTVPQNTGGK	0.387	
Glyceraldehyde -3-phosphate, liver	P04406	LISWYDNEFGYSNR	0.456	0.453
		AGAHLQGGAK	0.367	
		LVINGNPITIFQER	0.512	
		GALQNIIPASTGAAK	0.478	
Smooth muscle gamma actin	P63267	AVFPSIVGRPR	0.478	0.444
		AGFAGDDAPR	0.423	
		EITALAPSTMK	0.432	
Cellular myosin heavy chain, type A	P35579	SMAVAAR	0.422	0.410
		QASMPDNTAAQKV	0.398	
Chromodomain -helicase-DNA-binding protein 7	Q9P2D1	IGQSKSVK	0.469	0.437
		SLLIGVFK	0.406	
Aftiphilin	Q6ULP2	SSGTGTEPVAKLK	0.079	
Probable G-protein coupled receptor 135	Q8IZ08	EAGAAVR	0.490	
Citron Rho-interacting kinase	O14578	LPAGAVR	0.460	0.428
		KHAMLEMNAR	0.397	
Creatine kinase M-type	P06732	YYPLK	0.540	0.425
		LMVEMEKK	0.457	
		LSVEALNSLTGEFK	0.346	
		LGSSEVEQVQLVVDGV K	0.356	
Integrin beta-5	P18084	LAEEMRK	0.410	
Importin beta-1 subunit	Q14974	ELITILEK	0.376	
Glycine receptor alpha-1 chain precursor	P23415	SPEEMR	0.349	0.390
		SPEEMRK	0.432	
Bullous pemphigoid antigen 1	O94833	AELSRQLEGILK	0.412	0.406
		QRGEEMIAR	0.400	
Notch 2	Q04721	ADAAKR	0.395	0.424
		CTCKKGFK	0.453	
Transcription factor NRF	O15226	DRATELAVKLLQK	0.331	
Myosin Va	Q9Y4I1	MAASELYTK	3.756	4.816

		YLRMR	4.816	
Procollagen-lysine	Q02809	SEDYVDIVQGR	2.230	
Regulator of G-protein signaling 16	O15492	TLAAFPTTCLER	3.923	
Ankyrin repeat domain protein 11	Q6UB99	HDRDHFK	3.516	
Protein MICAL-3	Q7RTP6	GFLAAMDSAWMVR	3.258	
Acetyl-CoA acetyltransferase	P24752	TPIGSFLGSLSLLPATK	0.334	
Zinc finger and BTB domain containing protein 9	Q96C00	LHGNEILSGGGGPGGA	0.423	
DNA ligase IV	P49917	DTDLNQLK	0.362	
Acidic mammalian chitinase	Q9BZP6	HCVNGVTYQQNC	0.432	
Forkhead box protein D4B	Q8WXT5	KGNYWSLDPASQDMF	2.323	2.211
		QLTPGAHLPHPF	2.101	
		GAHLPHPF	2.210	
Golgi autoantigen	Q8TBA6	NASNIYSK	3.020	2.847
		AVAAKDSQLAVLK	2.182	
		EPDDELLFDFLNSSQ	3.341	
Protocadherin 16	Q96JQ0	NEHAPAFAR	2.450	2.395
		QDGGSPPRSTT	2.340	
Filamin C	Q14315	DAGEGLLTVQILGPEG	3.05	
Serine/threonine-protein kinase receptor R3	P37023	RLAADPVLSGLAQMM	3.35	

Table 3.4 List of Identified Proteins Differentially Expressed between Heart Tissue samples of 1 and 3

Name	Access ID	Unique Peptide Sequence	Sample 3/ Sample 1	Average ratio
Serum albumin precursor	P02768	QTALVELVK	0.54	0.454
		FKDLGEENFK	0.434	
		YLYEIARR	0.448	
		FPKAEFAEVSK	0.477	
		LSQRFPK	0.375	
Vimentin	P08670	ILLAELEQLK	0.356	0.423
		FANYIDKVR	0.415	

		FLEQQNK	0.498	
Hemoglobin beta chain	P68871	LHVDPENFR	0.348	0.416
		VAHHFGK	0.485	
10 kDa heat shock protein	P61604	FLPLFDR	2.10	
Citrate synthase, mitochondrial precursor	O75390	VVPGYGHAVLR	0.435	
Kinesin-like protein KIF12	Q96FN5	LLADSLGGR	0.25	
Mitogen-activated protein kinase kinase kinase kinase 6	Q8N4C8	QEINMLK	0.375	
Acylamino-acid-releasing enzyme	P13798	SFNLSALEK	0.295	
Phosphoglycerate kinase 1	P00558	AEPAKIEAFR	2.33	
Phosphoglycerate mutase 1	P18669	VLIAAHGNSLR	0.456	0.444
		KAMEAVAAQGK	0.501	
		AMEAVAAQGK	0.375	
Isocitrate dehydrogenase	P48735	LIDDMVAQVLK	0.398	0.467
		GRPTSTNPIASIFAWTR	0.567	
		TIEAEAAHGTVTR	0.436	
ADAM 29 precursor	Q9UKF5	HIIHIK	0.494	
NF-kappa-B essential modulator	Q9Y6K9	EVEHLK	2.141	
Red protein	Q13123	KISAIIEK	0.425	
Haptoglobin	P00738	FTDHLK	0.484	
Proactivator polypeptide precursor	P07602	LVGYLDR	2.271	2.201
		QEILAALEK	2.130	
Thioredoxin reductase 1, cytoplasmic	Q16881	NGPEDLPK	0.414	
Potassium voltage-gated channel subfamily B member 1	Q14721	RNGSIVSMNMK	0.256	
Zinc finger protein HRX	Q03164	KGRGNLEK	0.445	
Keratin, type II cytoskeletal 2 epidermal	P35908	SRGRGGGGGGFR	2.34	

Low-density lipoprotein receptor-related protein 2 precursor	P98164	ERATLGGNFR	0.453	
Cell division protein kinase 4	P11802	GPRPVQSVVP	0.324	
Angiopoietin 1 receptor precursor	Q02763	PRGLNLLPK	0.245	
Peroxiredoxin 2	P32119	KEGGLGPLNIPLLADVTR	0.235	

3.4 Conclusions

Comparing the relative abundance of each protein present in two or more complex samples can be accomplished using isotope-coded tags incorporated at the peptide level. The ideal strategy to introduce a stable isotope label for quantitative proteomics should be able to label proteins from any source: cell culture, tissue, or biological fluid. It should be capable, if possible, to introduce only one label per peptide to simplify data analysis. It should be relatively inexpensive, simple to perform, and go to completion under mild reaction conditions.

The dimethyl labeling method coupled with two dimensional liquid chromatography and mass spectrometry has been found to be effective for profiling expressed proteins in tumor cells. With this method, all the amino groups, both from the N-terminus and lysines, are labeled. More than 5000 peptide pairs have been detected for example between SCC9 and SCC9-E-cad cells. However, only about 300 pairs showed significant differences, that is, over a 2 fold change. Overall, 53 proteins were detected from the combination of SCC9 and SCC9-E-cad cell lines. However, from the combination of SCC9 and SCC9-Pg cell lines, about 82 proteins were detected. From the single peptide ratio and average ratio, it shows that most of the protein expression changes between the combined cell lines are over 2 fold. These studies could be useful to assess changes in a disease state or to distinguish specific state during disease progression.

Dimethyl labeling is a powerful chemical labeling strategy for protein mixtures extracted not only from cell lines but also proteins from tissue extracts. It proved useful in a comparison of proteins from diseased and healthy heart tissues and for protein extracts from the same patient taken at different stages of surgical procedure. However, in this protocol, the lysine side chains are blocked by guanidation prior to N-terminal labeling to prevent the incorporation of multiple labels. Therefore, the mass difference between the labeled peptides is always 6 mass units. More than 2000 pairs from a single combination and 160-216 pairs showed significant differences. The number of proteins detected from the combination of sample 1 and 2 and sample 1 and 3 were 32 and 24 respectively.

3.5 Literatures Cited

1. Wu, W.; Hu, W.; Kavanagh, J. J. Int. J. Gynecol. Cancer **2002**, 12, 409-423
2. Chelius, D.; Zhang, T.; Wang, G.; Shen, R. *Anal. Chem.* **2003**, 75, 6658 -6665.
3. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. *Nat. Biotechnol.* **1999**, *17*, 676-82.
4. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14440-45.
5. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. *Mol. Cell Biol.* **1999**, *19*, 1720-30.
6. Garrels, J. I.; McLaughlin, C. S.; Warner, J. R.; Futcher, B.; Latter, G. I.; Kobayashi, R.; Schwender, B.; Volpe, T.; Anderson, D. S.; Mesquita-Fuentes, R.; Payne, W. E. *Electrophoresis* **1997**, *18*, 1347-60.

7. Boucherie, H.; Sagliocco, F.; Joubert, R.; Maillet, I.; Labarre, J.; Perrot, M. *Electrophoresis* **1996**, *17*, 1683-99.

8. Washburn, M. P.; Wolters, D.; Yates, J. R., III. *Nat. Biotechnol.* **2001**, *19*, 242-7.

9. Spahr, C. S.; Davis, M. T.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Courchesne, P. L.; Chen, K.; Wahl, R. C.; Yu, W.; Luethy, R.; Patterson, S. D. *Proteomics* **2001**, *1*, 93-107.

10. Davis, M. T.; Spahr, C. S.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Yu, W.; Luethy, R.; Patterson, S. D. *Proteomics* **2001**, *1*, 108-17.

11. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-9.

12. Smolka, M. B.; Zhou, H.; Purkayastha, S.; Aebersold, R. *Anal. Biochem.* **2001**, *297*, 25-31.

13. Griffin, T. J.; Gygi, S. P.; Rist, B.; Aebersold, R.; Loboda, A.; Jilkine, A.; Ens, W.; Standing, K. G. *Anal. Chem.* **2001**, *73*, 978-86.

14. Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. *Anal. Chem.* **2001**, *73*, 5142-9.

15. Leitner, A.; Lindner, W. *J. Chromatogr. B* **2005**, 813, 1-26.

16. Peters, E.C.; Horn, D.M.; Tully, D.C. Brock, A. *Rapid Commun.Mass Spectrom.* **2001**, 15, 2387.

17. Peters, E.C.; Brock, A.; Horn, D.M.; Phung, Q.T.; Ericson, C.; Salomon, A.R.; Ficarro, S.B.; Brill, L.M. LC–GC Eur. **2002**, 15, 423.

18. Beardsley, R.L.; Reilly, J.P. *J. Proteome Res.* **2003**, 2, 15-21.

19. Kuyama, H.; Watanabe, M.; Toda, C.; Ando, E.; Tanaka, K.; Nishimura, O. *Rapid Commun. Mass Spectrom.* **2003**, 17, 1642.

20. Chelius, D.; Shaler, T.A. *Bioconjugate Chem.* 2003, 14, 205.

21. Gevaert, K.; Goethals, M.; Martens, L.; J. van Damme,; Staes, A.; Thomas, G.R.; Vandekerckhove, J. *Nat. Biotechnol.* 2003, 21, 566.

22. Staes, A. Demol, H.; J. van Damme, Martens, L.; Vandekerckhove, J.; Gevaert, K.; *J. Proteome Res.* **2004**; 3, 786-791.

23. Kuhn, K.; Thompson, A.; Prinz, T.; M¨uller, J.; Baumann, C.; Schmidt, G.; Neumann, T.; Hamon, C. *J. Proteome Res.* **2003**, 2, 598-609.

24. Prinz, T.; M¨uller, J.; Kuhn, K.; Sch¨afer, J.; Thompson, A.; Schwarz, J.; Hamon, C.; *J. Proteome Res.* **2004**, 3, 1073-1081.

25. Chakraborty, A.; Regnier, F.E.; *J. Chromatogr. A* **2002**, 949,173.

26. Geng, M.; Ji, J.; Regnier, F.E.; *J. Chromatogr. A* **2000**, 870 295.

27. Ji, J.; Chakraborty, A.;Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F.E.; *J. Chromatogr. B* **2000**, 745 197.

28. Kindy, J.M.; Taraszka, J.A.; Regnier, F.E.; Clemmer, D.E. *Anal. Chem.* **2002**, 74, 950-958.

29. Liu, P.; Regnier, F.E. *J. Proteome Res.* **2002**, 1, 443-450.

30. Liu, P.; Regnier, F.E. *Anal. Chem.***2003**, 75, 4956-4963.

31. Xiong, L.; Andrews, D.; Regnier, F. *J. Proteome Res.* **2003**, 2, 618-625.

32. Ren, D.; Penner, N.A.; Slentz, B.E.; Regnier, F.E. *J. Proteome Res.* **2004**, 3, 37-45.

33. Mason, D.A.; Liebler, D.C. *J. Proteome Res.* **2003**, 2, 265-272.

34. M¨unchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, 72, 4047-4057.

35. Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. *Anal. Chem.* **2003**, 75, 6843-6852.

36. Hsu, J.-L.; Huang, S.-Y.; Shiea, J.-T.; Huang, W.-Y.; Chen, S.-H. *J. Proteome Res.* **2005**, 4, 101-108.

37. Ji, C.; Li, L. *J. Proteome Res.* **2005**, 4, 734-742.

38. Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, 74, 1884-1890.

39. Kimmel, J. R. Guanidination of proteins. *Methods Enzymol.***1967**, 11, 584–589.

40. Zappacosta, F.; Annan, R. S. *Anal. Chem.* **2004**, *76*, 6618-6627.

41. Li, Z.; Gallin, W. J.; Lauzon, G.; Pasdar, M. *J. Cell Sci.* **1998**, 111, 1005-1019.

42. Parker, H. R.; Li, Z.; Sheinin, H.; Lauzon, G.; Pasdar, M.; *Cell Motil. Cytoskeleton* **1998**, 40, 87-100.

43. Hakimelahi, S.; Parker, H. R.; Gilchrist, A. J.; Barry, M.; I, Z.; Bleackley, R. C.; Pasdar, M. J. Biol. Chem. **2000**, 275, 10905-10911.

44. Ji, C.; Li, L.; Gebre, M.; Pasdar, M.; Li, L. *J. Proteome Res.* **2005**, 4, 1419-1426.

45. Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2000**, 14, 2070-2073.

46. Beardsley, R. L.; Karty, J. A.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **2000**, 14, 2147-2153.

47. Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. *Anal. Biochem.* **2000**, *287*, 110-117.

Chapter 4

Conclusions and Future Work

Mass spectrometry (MS) combined with different separation techniques has proven to be a powerful method in proteomic research. Two types of MS are used for most work in proteomics: matrix-assisted laser-desorption-ionization time-of-flight (MALDI-TOF) instruments and electrospray ionization tandem mass spectrometry (ESI-tandem MS) instruments. Recently, mass spectrometry has become a useful technique in the quantitative analysis of differential expression. Some stable-isotope-based methods have been developed and applied extensively in this comparative analysis because they are well suited to mass spectrometric analysis. The overall analysis depends on several factors, such as extraction, sample preparation, separation methods and choice of mass spectrometry.

In chapter 2, two combined methods for the identification of proteins from human heart tissue are demonstrated. Protein extraction, sample preparation, digestion and peptide fractionation can have a major impact on the identification process. Some alternative protein extraction reagents that are applicable for heart tissues are commercially available. Trizol reagent was found to be efficient at extracting proteins from heart tissue. However, one of the challenging steps in the extraction process and sample preparation is to solublize the protein pellets obtained from the Trizol extraction. While surfactants, such as SDS, are commonly used in solubilizing protein extracts, their presence and concentration in a protein sample can potentially affect both the enzymatic digestion process and the subsequent analysis of the resulting peptides by mass spectrometry. For this reason, 1% SDS was used to solubilize the protein pellets and then finally diluted to about 0.05% SDS. However, a

significant proportion of protein pellets was observed to be insoluble in 1% SDS. A new protein digestion method, microwave -assisted acid hydrolysis (MAAH), developed by our group, was used to analyze the SDS-insoluble proteins. This method has become a promising technique to solve some of the problems that are encountered in shotgun proteomics techniques. MAAH is detergent-free and digests proteins without enzymes or leaving residual chemicals.

Due to the complex nature of the peptide mixtures obtained by trypsin digestion or microwave-assisted acid hydrolysis, two-dimensional liquid chromatography has been used to fractionate the complex mixtures. The use of strong-cation exchange chromatography not only fractionates the peptide mixtures but also serves to remove detergents, such as SDS. Before identification of the protein by mass spectrometry, the intensity of SCX UV-absorbance chromatographs indicates the presence of a large amount of peptides in the mixture. It was demonstrated that reversed-phase chromatography is very useful to further separate the fractions collected from the SCX HPLC.

Protein identification has been analyzed using two different mass spectrometry instruments. Analyzing the SDS-soluble protein mixture collected from SCX fractions has been carried out using an electrospray ionization tandem mass spectrometry (ESI-tandem MS) instrument. Although an additional step was required to remove salts and detergents from the fractions, LC-ESI-tandem MS system was found to be fast and convenient process. MALDI is known to be tolerant to detergents and salts and also sensitive. High mass accuracy and high resolution is expected from the MALDI-QTOF instruments. Therefore, SCX fractions collected from MAAH digests were analyzed with this instrument. In this experiment the heated droplet LC-MALDI interface was found to be very useful to collect

the effluents from RP-HPLC. There are many factors that may limit the high throughput identification of the proteins analyzed using MALDI-QTOF. The first one is the selection of peptide masses from the MALDI spectrum. Only peptide masses that have high intensity were selected for fragmentation using CID. Those peptides with low intensity are missed due to the low quality of MS/MS spectra collected. The other factor is related to the quality of the spectra collected because all the MS/MS spectra are collected manually. The collected MS/MS spectra are searched. A significant number of proteins are reported in this work using the two different digestion methods followed by LC-ESI ion trap and LC-MADI-QTOF. The proteins with at least two matched peptides obtained in both methods gave high confidence in the identification of proteins.

In chapter 3, global isotope labeling strategy was applied to compare the protein expression changes between carcinoma cell lines and also heart tissue samples. Quantitative analysis of the relative abundance of expressed proteins is an essential issue in comprehensive proteomics. Using global-isotope labeling, coupled with mass spectrometry based techniques, to quantify changes in protein abundance between two samples is a simple and practical way for profiling biological differential regulation.

Since the most popular labeling method, ICAT fails to identify peptides lacking cysteine residues, dimethyl isotope labeling, which labels both the N-termini and lysine residues of the peptides generated by trypsin or MAAH digestion, was used. This method has several advantages, including the use of the inexpensive and commercially available reagent, formaldehyde. The dimethyl labeling procedure is also fast and simple when compared to some of the other methods. The labeled peptide mixtures were first fractionated using SCX HPLC and salts were removed by RP-HPLC followed by the deposition of the

effluents onto the MALDI plate with the help of heated droplet interface. Peak pairs separated by mass difference of 4n in the MALDI MS spectrum were selected, where n is the number of dimethyl labeling sites in the peptide. For the sample combinations of SCC9 and SCC9 E-cad, about 5480 pairs were detected. However, Only 320 peptide pairs showed significant changes, that is, 2-fold or above, and were selected for fragmentation or CID to identify the proteins. All the collected MS/MS spectra were searched against the Swiss Prot database using the MASCOT program. The purpose of applying this method to the carcinoma cell lines was to determine proteins differentially expressed between an E-cadherin-deficient human carcinoma cell line, SCC9, and E-cadherin and plakoglobin-expressing SCC9 transfectants. Therefore, better understanding of the cadherin-catenin system in the regulation of cell proliferation, invasion and intracellular signaling during cancer can be obtained.

In chapter 3, using a similar method to that reported, the quantification and identification of heart tissue proteins is reported. The objective was to identify the proteins that are differentially expressed in the different tissue samples. The samples were obtained from the same patient at different surgical steps. The first sample was taken prior to cardiopulmonary bypass (CPB). The second was taken during CPB, and the third, after CPB. While attempting to denature the protein mixture at 95 ^{o}C, a small amount of protein precipitate was observed. For this reason, MAAH was also used to digest proteins, in addition to trypsin. Although trypsin is expected to cleave with equal tendency at lysine and arginine, it is reported that signals from arginine-containing peptides are generally stronger. It has been reported that the higher ionization efficiency of arginine-containing peptides is due to very high basicity of guanidine functionality in side chain of the arginine residue.

Based on this, a recent focus of research has been directed toward increasing mass spectral signal intensities from lysine-containing peptides. In this procedure, all the lysine side chains of the digested peptides of the tissue extracts were first blocked by guanidation and then followed by dimethyl labeling using formaldehyde reagent. The advantage of this method is that the mass difference between the peptide pairs is only 6, unlike the dimethyl labeling method mentioned above. From the MALDI spectra, more than 2000 peptide mass pairs were selected from a single combination and 160-216 pairs showed at least 2-fold differences. From these pairs, 24 and 32 proteins were identified from samples 1+3 and 1+2, respectively. The comparison of protein expression and detection of the differentially expressed proteins by dimethyl isotope labeling could be useful for biomarker discovery.

Although a significant number of proteins were detected from the combination of trypsin and MAAH digests of proteins coupled with mass spectrometry, an effective protein isolation method, fast sample preparation and automated peptide separation and high-throughput identification techniques still need to be established in order to identify all possible proteins in heart tissue. In particular, an effort should be made to automate the collection of MS/MS spectra in MALDI-QTOF instruments.