TOWARDS THE DEVELOPMENT OF A PROTEOMICS WORKFLOW FOR HIGH-THROUGHPUT PROTEIN BIOMARKER DISCOVERY

by

Mark James Wall

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
May 2010

DALHOUSIE UNIVERSITY

Department of Chemistry

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "TOWARDS THE DEVELOPMENT OF A PROTEOMICS WORKFLOW FOR HIGH-THROUGHPUT PROTEIN BIOMARKER DISCOVERY" by Mark James Wall in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Dated:

Supervisor: _____

Readers: _____

_____

_____

DALHOUSIE UNIVERSITY

DATE: May 17, 2010

AUTHOR: Mark James Wall

TITLE: TOWARDS THE DEVELOPMENT OF A PROTEOMICS WORKFLOW FOR HIGH-THROUGHPUT PROTEIN BIOMARKER DISCOVERY

DEPARTMENT OR SCHOOL: Department of Chemistry

DEGREE: PhD CONVOCATION: October YEAR: 2010

Table of Contents

List of Tables

List of Figures

Abstract

Two popular workflows exist for quantitative proteome analysis: two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), with staining to visualize proteins, and multidimensional solution phase separations of isotopically labelled peptides coupled to mass spectrometry (MS). However, the development of an alternative strategy, which combines easy-to-read differential profiling as seen in 2D-PAGE, with the sensitivity of MS for detection and identification, is needed. This thesis presents work towards the development of a workflow for high-throughput protein biomarker discovery.

Multidimensional separations are vital to obtain sufficient fractionation of complex proteome mixtures. As a first dimension of separation, ion exchange chromatography (IEC) is a common choice, though it has yet to be thoroughly evaluated in terms of its effectiveness as a proteome prefractionation tool. This study used a defined set of protein standards to establish the resolution and proteome yield obtained through IEC. The evaluation uncovered significant bias in terms of protein recovery and separation.

To improve throughput of a multidimensional separation strategy, a multiplexed (8-column) reversed phase liquid chromatography (RPLC) platform was constructed. The system design allowed for even distribution of flow across all columns with limited cross-loading during sample loading. This system was directly coupled to matrix-assisted laser desorption/ionization (MALDI) through a novel well plate device. The Teflon wells allowed for high recovery and no cross-contamination during collection/spotting, improved throughput, and greatly reduced the number of sample manipulation steps.

An evaluation of MALDI MS, using the ThermoFisher vMALDI LTQ, for quantitative profiling was performed, employing the multiplexed LC-MALDI platform. The use of MALDI MS allowed for fast (< 5.5 hours) acquisition of quantitative data from isotopically differentiated samples partitioned over 640 fractions from two-dimensional LC. Proteins comprising 0.1% of the proteome were detected and quantified using this method.

Finally, the effects of varying concentrations of acetonitrile (ACN) upon the products generated from tryptic digestions were explored. Poor enzymatic efficiency in 80% ACN was found to be responsible for an increased concentration of peptides containing missed cleavage sites. These peptides often contained unique amino acid sequences, which were not detected from complete digestions, resulting in improved protein sequence coverage following MS analysis.

List of Abbreviations and Symbols Used

ΔCN	Delta correlation value
%C	Percent cross-linker (*w*/*w*)
%T	Percent total acrylamide (*w*/*v*)
%RSD	Percent relative standard deviation
µg	Microgram
µL	Microlitre
1D	one dimensional
2D	Two dimensional
2DLC	Two-dimensional liquid chromatography
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
Å	Angstrom
AC	Alternating current
ACN	Acetonitrile
amu	Atomic mass unit
BAEE	N-benzoyl-L-arginine ethyl ester
BSA	Bovine serum albumin
CE	Capillary electrophoresis
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate
CHCA	α-Cyano-4-hydroxycinnaminic acid
CID	Collision induced dissociation
Da	Dalton
DC	Direct current
DDA	Data dependent acquisition
DIGE	Differential gel electrophoresis
DTT	Dithiothreitol
ECD	Electron capture dissociation
ESI	Electrospray ionization
ETD	Electron transfer dissociation
EtOH	Ethanol
ExPASy	Expert Proteomics Analysis System
FA	Formic acid
fmol	Femtomole
FTICR	Fourier transform ion cyclotron resonance
g	Gram
GELFrEE	Gel-eluted liquid fraction entrapment electrophoresis
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
ICAT	Isotope-coded affinity tags
ID	Internal diameter
IEC	Ion exchange chromatography
IEF	Isoelectric focusing
IPG	Immobilized pH gradient

iTRAQ	Isobaric tag for relative and absolute quantitation
kDa	Kilodalton
L	Litre
LC	Liquid chromatography
LIT	Linear ion trap
M	Molarity
MALDI	Matrix-assisted laser desorption/ionization
MDLC	Multidimensional liquid chromatography
min	Minutes
mL	Millilitre
mm	Millimetre
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MudPIT	Multidimensional protein identification technology
m/z	mass to charge ratio
NCBI	National Center for Biotechnology Information
Nd:YAG	neodymium-doped yttrium aluminium garnet
nm	nanometer
NMR	Nuclear magnetic resonance
NSI	Nanospray ionization
PDMS	Polydimethylsiloxane
pI	Isoelectric point
PMF	Peptide mass fingerprinting
pmol	Picomole
ppm	Parts per million
PPS	Sodium 3-(4-(1,1-bis(hexyloxy)ethyl)pyridinium-1-yl)propane-1-sulfonate
PTM	Post-translational modification
QTOF	Quadrupole time-of-flight
RF	Radio frequency
RPLC	Reversed phase liquid chromatography
RSD	Relative standard deviation
Rsp	Ranked preliminary score
s	Seconds
SAX	Strong anion exchange
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
S/N	Signal to noise
SEC	Size exclusion chromatography
SILAC	Stable isotope labelling with amino acids in culture
SRM	Selected reaction monitoring
STD	Standard deviation
TEAB	Triethylammonium bicarbonate
TOF	Time-of-flight
TPCK	Tosyl phenylalanyl chloromethyl ketone

TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
WAX	Weak anion exchange
WCX	Weak cation exchange
Xcorr	Cross-correlation value

Acknowledgements

I would first like to thank my supervisor, Dr. Alan Doucette, for his encouragements and helpful criticisms over the course of my time at Dalhousie. His infectious enthusiasm and keen insights have been a continuous source of inspiration.

I thank all of my committee members, Dr. Robert White, Dr. Heather Andreas, and Dr. Robert Guy, for their advice over the years. Thank you to Dr. Neil Burford, Dr. Mark Stradiotto, Dr. Peter Wentzell, Dr. Amares Chatt, and Dr. Louis Ramaley, for their advice and help. I extend my thanks to Dr. Ken Yeung for reviewing this thesis.

Thank you to the office staff of the Department of Chemistry, particularly Giselle Andrews, for their incredible patience and help over the years. Thank you to Machinists Mike Boutilier and Rick Conrad, of the Chemistry Department, for their invaluable help with many aspects of the work presented in this thesis. Also, I would like to thank all of the lab instructors, especially Cathy Byers, during my time as a teaching assistant.

I especially would like to thank all of my current group members, Diane Botelho, Fang Liu, Shayla Fitzsimmons, Dennis Orton, and Douglas Vieira, for continuing to make my time at Dalhousie enjoyable. Thanks to my former group members, Dr. John Tran, Leanne Lucas, Dr. Aaron Steevensz, Jason Smith, Simon Larose, Patrick Murphy, Graham Cary, Gordon Simms, and Harmon Claire. Thanks to Ken Chisolm and Pat Murphy of the NRC-IMB for their valuable help and time. Thank you to Chris McDonald for his help with construction of the MALDI well plate device. Again, a special thank you to both John and Diane for their invaluable friendship and help over the years.

I would like to thank my entire family; especially my mom and dad, Vicki and John, as well as Freda and Gerard, who have given unwavering support and love throughout my education. To my brother Michael, for his indispensible camaraderie, and my sister Julie, for all her help over the years, thank you so much.

Finally, I thank the most important person in my life. Mary Beth, your support, love, and especially patience have been more valuable to me than you could imagine.

Chapter 1

Introduction

1.1 Overview

Proteomics is a field of study which depends upon effective sample preparation strategies and sensitive instrumentation for the characterization of complex biological samples. Among the many objectives of proteomics, the discovery of novel protein biomarkers has been a major driving force in the development of proteome characterization methodologies [1-3]. Mass spectrometry (MS), in particular, has become indispensable for proteome characterization, owing to its speed, selectivity, and sensitivity [4, 5]. Also, given the complexity of the proteome, extensive sample manipulation, including proteome prefractionation, must be performed prior to protein identification through MS [6]. The set of tools and techniques available to the proteomics researcher is referred to as the proteomics toolbox [7]. Expanding the proteomics toolbox not only necessitates an optimization of current technologies, but also the development, evaluation, and optimization of novel technologies. This need is particularly pressing when considering the throughput of a modern proteomics analysis experiment. Shorter analysis times than what is currently achieved with existing technologies is critical to advancing proteomics as a routine experiment, especially in the context of protein biomarker discovery. In this introductory chapter, the goals and challenges of modern proteomics will be introduced. An overview of the current proteomics workflows employed for comparative quantitative proteomics analysis is presented, highlighting separation technologies for fractionation of the proteome and the basic principles of MS-

based proteomics.

1.2 Proteomics

1.2.1 The Unique Challenges of Proteomics

The completion of the human genome project [8, 9], along with the genomes of an increasing number of other species [10-13], revealed a genome that was much simpler than originally predicted. Thus, the underlying mechanisms of cellular function are more complex than imagined and a direct study of gene products may establish the functional role of proteins in the cell, and further our understanding of how biological systems work. However, the analysis of organisms at the protein level presents a variety of unique analytical challenges, many of which still need to be overcome.

A proteome contains a wide diversity of unique proteins ranging in physical and chemical properties and are present over a large concentration dynamic range [14, 15]. A typical proteome may contain thousands, to tens of thousands of unique proteins. In more complex eukaryotic organisms, such as humans, which contain around 20,000 to 25,000 genes, alternative splicing of RNA precursors following transcription results in the translation of several different protein isoforms, further increasing the complexity of the sample [16]. Combined with a large number of possible post-translational modifications (PTMs) [17], the number of unique proteins which may comprise the proteome of complex organisms may be upwards of one million [18]. This incredible number of proteins expressed in the proteome presents a difficult analytical challenge as not all proteins are amenable to separation or analysis. Hydrophobic proteins, for example, being critical components of cellular membranes, are considered more difficult to manipulate in

solution due to their poor solubility in water [19].

1.2.2 The Goals of Proteomics

In general, proteomics is the holistic study of the proteins expressed by an organism, referred to as the proteome [20]. This includes characterizing the proteome in terms of structure, function, and expression level, be that over time or under particular conditions of biological stress [21]. On a practical level, one of the major goals of proteomics is to provide useful information to medical practitioners about the mechanisms of human diseases (e.g., cancer), with a long term objective of developing improved methods for disease diagnosis and treatment. Although easy to generalize, the realization of this objective often requires an accumulation of knowledge derived from various aspects of proteomics analysis, with each having its own specific objectives.

The areas of study within proteomics are fairly diverse in scope. For instance, considering the function of proteins, the study of protein-protein interactions as well as their role in metabolic and signal pathways is critical to understanding how proteins carry out biological processes [22-24]. Structural analysis of proteins through X-ray crystallography, nuclear magnetic resonance (NMR) [25-33], or mass spectrometry [34-37] allows insights into how the higher order folding of a protein is correlated to its function. The identification and localization of protein PTMs is critical to understanding the changing function of a protein as these are often responsible for regulating metabolic pathways as well as a host of other biological functions [38, 39, 39-42].

1.2.3 Biomarker Discovery

Expression proteomics is an area of considerable interest in the field as it has direct implications in the development of disease diagnostics and treatment [43-54].

Expression proteomics involves monitoring the changing level of protein concentration which manifests as a consequence, or as causality, to a change in the environment of the organism, a stage of growth or, most importantly, a disease state. Expression proteomics attempts to identify those proteins whose expression level correlates to the physiological state of the organism, and as such, may be useful indicators of this state. Proteins which can be used as identifiers of a given state are referred to as protein biomarkers, and may include a single protein or multiple proteins.

Considering the complexity of the proteome, improved methods of analysis are of critical importance to realize the goals of proteomics. As already mentioned, the methodologies employed in proteomics vary depending on the objectives of the experiment. The analysis scheme used to study the proteome is referred to as the proteomics workflow. The following sections will highlight the two most common proteomics workflows used in the context of expression proteomics.

1.3 The Proteomics Workflow

The proteomics workflow refers to the collection of experimental techniques and instruments used as a means of acquiring knowledge related to a particular goal of proteomics analysis. Expression proteomics entails the determination of protein concentration and is typically conducted in a comparative scheme. Thus, one may compare the proteome of a given organism under two physiological conditions; the healthy (control) state versus the diseased (test) state of the organism. The goal of this experiment is to determine the change in expression of proteins between the two states. Several workflows have been devised for conducting expression proteomics analysis. In

general, the majority of these proteomics workflows incorporate one or two stages of protein prefractionation followed by tandem mass spectrometry for protein identification (peptide sequencing). Enzymatic digestion of the proteome, as well as staining or labelling techniques which permit relative comparison of the two proteome states, are critical aspects of these workflows. A schematic outlining the stages of two distinct workflows for comparative proteomics analysis is provided in Figure 1.1.

Workflow "A" uses a gel-based approach which involves the multidimensional separation of proteins through two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), the details of which will be discussed later in this chapter. Visualization of the proteins within the gel through various staining methods is used to compare the proteome profiles of a control and test sample. Proteins of interest can be enzymatically digested and extracted from the gel for identification using MS-based peptide mass fingerprinting or peptide sequencing. In contrast to this gel-based workflow, workflow "B" employs a solution-based, peptide level analysis which involves the enzymatic digestion of the proteome prior to multidimensional liquid chromatography (MDLC) separations and MS peptide sequencing. Relative quantitation between and test and control sample is afforded by MS detection and can be performed through differential isotopic labeling at either the protein or peptide level.

The workflows outlined above for comparative proteome profiling clearly make use of distinct technologies to acquire knowledge about the biological system. It is noted, however, that the role of mass spectrometry is indispensable to both comparative workflows. Due to its importance, a review of mass spectrometry instrumentation and strategies for protein identification through MS will first be presented. Following this, a

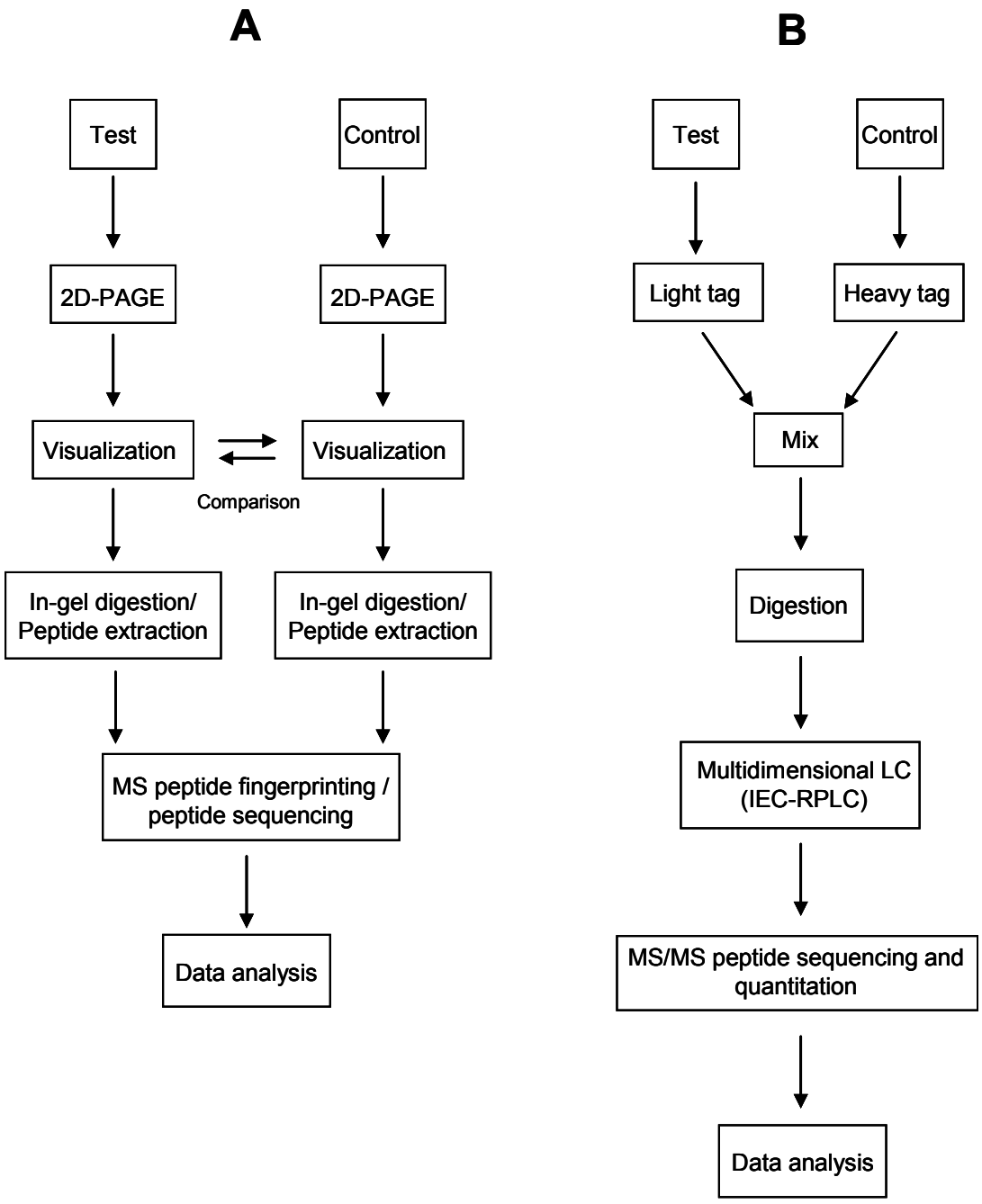

Figure 1.1 Two workflows used for expression proteomics analysis. They employ either (A) a gel-based approach using protein level separation followed by visualization and comparison of test and control samples, with in-gel digestion and LC-MS/MS analysis of the extracted peptides, or (B) a solution-based approach, with peptide level separations, coupled directly to MS/MS peptide sequencing following a complete digestion and combination of differentially mass labelled samples.

description of the currently used methods for protein sample preparation prior to MS is provided.

1.4 Mass Spectrometry

Mass spectrometry has been, without a doubt, the greatest contributing technology to the field of proteomics in the last two decades. Its speed and sensitivity make it an ideal tool for the high-throughput analysis of large numbers of proteins from complex mixtures. The design and function of these instruments plays an important part in dictating the type of proteomics experiment that can be performed. A review of the different types of mass analyzers and a brief discussion of their advantages and disadvantages for MS-based proteomics strategies is given.

1.4.1 The Mass Analyzer

Mass spectrometers come in a variety of instrumental configurations. The heart of a mass spectrometer is the mass analyzer. Mass analyzers achieve ion separation by taking advantage of the differences in behaviour between two ions of different mass to charge ratio (*m/z*) within an electric and/or magnetic field. This can be achieved in several different ways and the varying strategies have birthed a range of commercial mass spectrometers utilizing these different mass analyzers. These include: the quadrupole, ion trap (2D and 3D), time-of-flight (TOF), and high performance instruments like Fourier Transform ion cyclotron resonance (FTICR) and the Orbitrap. In addition, hybrid MS instruments are available which combine two or more mass analyzers into a single instrument allowing for greater experimental flexibility, such as tandem MS analysis. The ion trap instrument employed in all studies presented in this

thesis is the ThermoFisher LTQ linear ion trap mass analyzer. The features of this particular mass analyzer will be discussed in detail along with brief descriptions of other MS instrumentation.

The ion trap mass analyzer is a popular instrument for proteomics analysis. Ion traps come in two different varieties: the 3D trap, or Paul traps, and the 2D trap, also known as a quadrupole or linear ion trap (LIT). A diagram of these analyzers is provided in Figure 1.2. The ion trap mass analyzer is particularly popular for proteomics analysis due to its improved sensitivity over quadrupole instruments and its ability to perform several stages of tandem mass spectrometry, referred to as MS^n, where n indicates the stage of MS analysis. The low cost, simple maintenance and small footprint of these instruments also add to their popularity. In both designs, ions are retained, or trapped, within the mass analyzer by establishing stable ion trajectories over a given *m/z* range through the use of an alternating current (AC) electric field with radio frequency (RF) and a direct current (DC) electric field. In the 3D trap, these fields are generated when voltage is applied through the hyperbolic metal end caps and a hyperbolic ring electrode.

The design of the linear ion trap is slightly modified from that of the 3D trap, resembling more closely to the configuration of a quadrupole mass analyzer, with the exception that end cap electrodes are used to trap the ions along the z axis. One of the main advantages of a linear ion trap over 3D traps in that it can trap a greater number of ions before space charge effects begin to impact negatively upon the resolution and mass accuracy. Linear ion traps are available with either axial or, in the case of the ThermoFisher LTQ linear ion trap, radial detection (see Figure 1.2). With this instrument, mass spectra are typically acquired with a base peak resolution of 1 amu and provide a

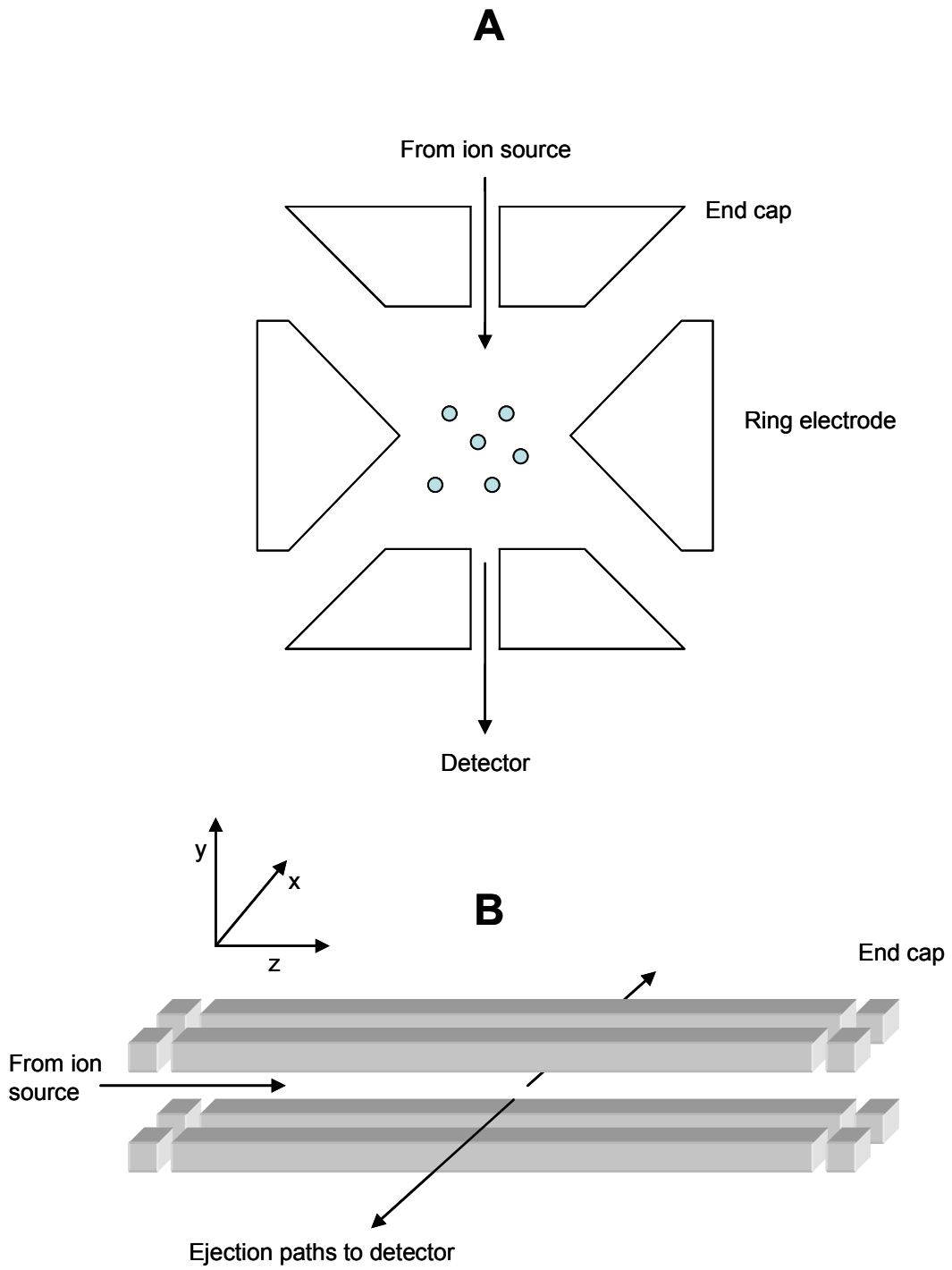

Figure 1.2 Diagrams of both the (A) 3D ion trap (Paul trap) and (B) linear ion trap. Both analyzers trap ions with the use of a RF AC voltage applied to the ring or quadrupole electrode and DC voltage applied to end cap electrodes. While the operation of each analyzer is very similar the linear ion trap offers greater sensitivity due to the increased number of ions which may be stored in the trap without space-charge effects.

mass accuracy of 100 ppm. Higher resolution scans can be performed through the use of "zoom scans". These higher resolution scans come at the expense of sensitivity or, more typically, scan speed and thus are usually performed over a narrow mass window. The combination of the LTQ with the Orbitrap, referred to as the LTQ-Orbitrap, is a very common combination for MS-based proteomics, and has been responsible for the best coverage of the proteome in terms of protein identifications to date [55, 56].

Other MS instruments are available for proteomics studies. Given the requirement for tandem MS analysis, the triple quadrupole instrument can be used for proteome investigations. This instrument simply comprises three quadrupole mass analyzers strung together in series (referred to as Q1, Q2, and Q3). A popular use for triple quadrupole instruments has been for the unambiguous identification and accurate quantitation of small drug molecules, toxins, or metabolites [57]. This strategy, referred to as selected reaction monitoring (SRM), has been applied to targeted approaches for protein quantitation through selection of a parent ion in Q1, fragmentation of the molecule in Q2, and isolation/detection of one of the resulting fragment ions in Q3. Multiple reaction monitoring (MRM) scans apply this strategy to more than one target molecule in a single scan. For applications which require greater resolution and mass accuracy, orthogonal injection time-of-flight (TOF) mass analyzers are available. Such instruments typically provide mass resolution on the order of 20,000 or greater and an accuracy of 5 ppm or better. Tandem MS is possible through hybrid instruments, such as the TOF/TOF instrument and has been applied to proteomics analysis [58].

More recently, a generation of high resolution mass analyzers have become increasingly prevalent amongst proteome facilities. These include the aforementioned

FTICR mass analyzer [59, 60] and Orbitrap mass analyzer [61, 62]. The FTICR can provide high resolution (500,000) and accuracy (<1 ppm). In the proteomics workflow the conventional strategy to protein identification has involved MS analysis of peptide fragments generated through enzymatic cleavage of the sample. This "bottom-up" approach to proteomics analysis, described further in section 1.5, can be conducted on low resolution instruments. Top-down analysis involves MS of intact protein molecules and requires higher resolution and mass accuracy to interpret the complex spectra arising from this type of analysis. FTICR systems are expensive and require a high level of maintenance due to the powerful magnet which must be supercooled for operation. The Orbitrap mass analyzer retains some of the qualities of the quadrupole ion trap as well as FTICR. The resolving power of an Orbitrap is around 200,000 and its mass accuracy is very comparable to the FTICR analyzer. Additionally, by replacing the magnet with an electrostatic form of trapping ions, the Orbitrap is more cost effective and easier to maintain.

Although MS instrumentation is advancing rapidly, its application to biomolecule analysis would not have been possible without the development of two particular ionization methods. These two technologies, namely electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), permitted the ionization of large biomolecules, and as such, enabled these molecules to be analyzed using mass spectrometry.

1.4.2 Ion Sources

The ion source is the first stage of the MS instrument prior to *m/z* determination and serves two main purposes. First, it brings molecules into the gas phase. Second,

molecules are ionized so that they can be manipulated by the applied magnetic and/or electric fields of the mass analyzer. The vaporization and ionization of proteins proved historically to be the most difficult aspect of MS analysis. Large molecules, such as proteins, undergo significant fragmentation through classical ionization processes due to their thermal lability. However, the development of the soft ionization techniques of ESI by John Fenn [63] and MALDI by Franz Hillenkamp and Michael Karas [64, 65], subsequently applied to whole proteins by Koichi Tanaka and coworkers [66], changed the course of mass spectrometry and proteomics. This pivotal development was recognized by partly awarding the 2002 Nobel Prize in Chemistry to both John Fenn and Koichi Tanaka for their contribution to the development of soft ionization methods for mass spectrometric analyses of biological macromolecules.

1.4.2.1 Electrospray Ionization

Electrospray ionization is a relatively simple ion source which permits the direct coupling of liquid samples to a mass spectrometer. A diagram of the ESI source is shown in Figure 1.3. The ESI source consists of an emitter tip through which sample is passed. A voltage ranging from 1.5 to 5 kV is applied across the emitter and the entrance to the mass spectrometer. The applied voltage results in the generation of charge within the solvent droplets emitting from the ESI emitter tip. The repulsion forces of the charge present in the droplet causes the formation of a Taylor cone and a subsequent series of desolvation and fission stages which result in the formation of analyte ions, which then enter the mass analyzer. The evaporation process is often assisted by applying nitrogen gas, especially at higher flow rates. The formation of multiply charged ions is typical

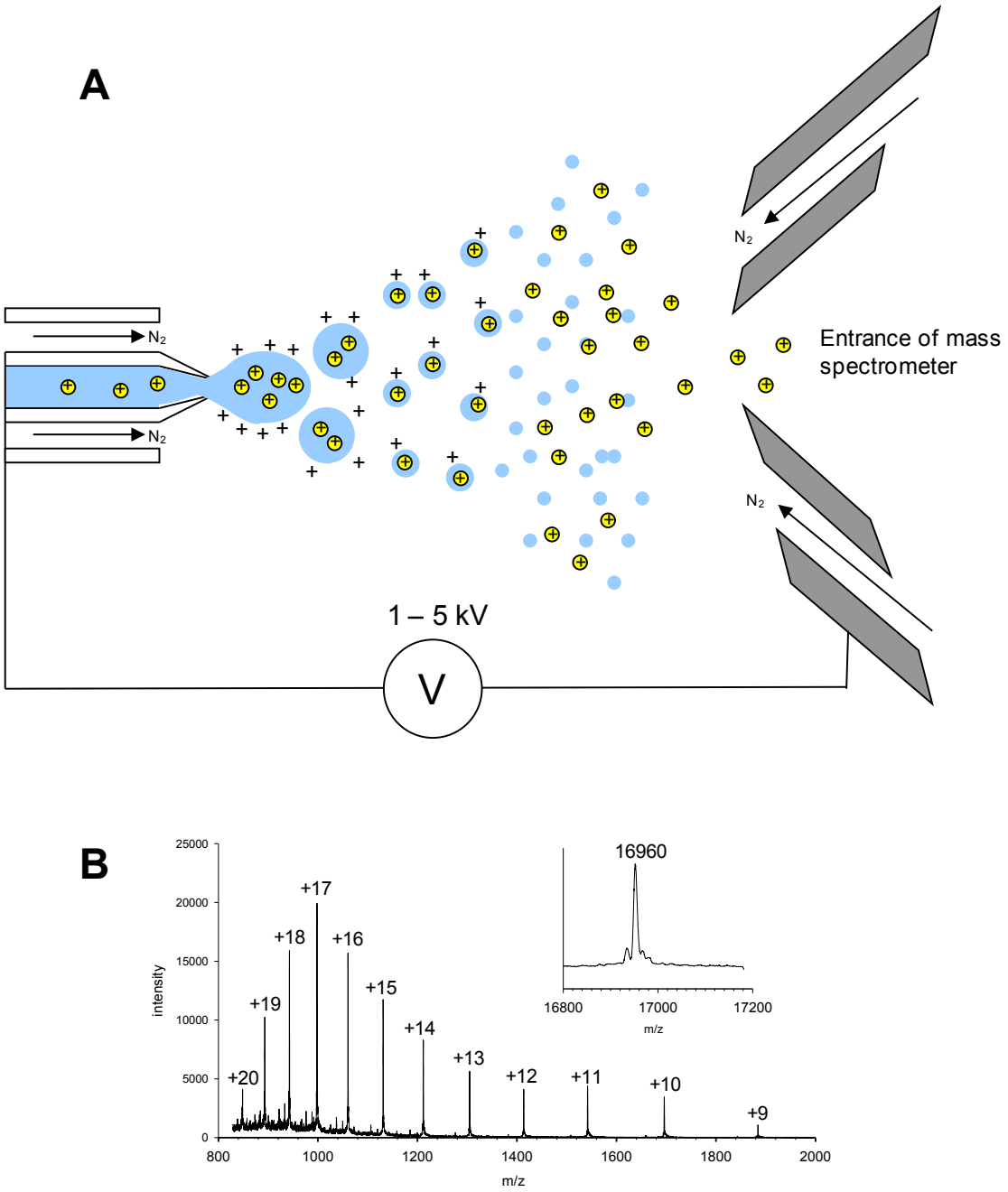

Figure 1.3 Example of the (A) electrospray process and (B) charge envelope observed for myoglobin resulting from the distribution of charge states along with the deconvoluted parent mass (shown in insert). Deconvolution of the charge envelope is performed to obtain the molecular weight of the protein. Electrospray operates by applying a voltage across an emitter tip in the range of 1 to 5kV which results in the formation of charge to the eluting solvent. Ideally the solvent forms a Taylor cone at the end of the tip which emits a liquid jet which then breaks into smaller droplets due to coulomb repulsion. Desolvation due to vacuum and nitrogen gas results in individually charged analyte particles entering the mass spectrometer. The mass spectrum presented was collected by the author.

of the electrospray process. While this can complicate the interpretation of protein spectra, as they are observed as a charge envelope (example provided in Figure 1.3), they provide more information-rich fragmentation spectra for peptide ions, which will be discussed in section 1.6. If low flow rates are used (<2 μL/min), as in the case when incorporating capillary column separations, an alternative form of ESI is often applied. This form is referred to as nanospray ionization (NSI), referring to the small volumes of sample eluting from the emitter tip. Dynamic NSI permits the coupling of capillary flow liquid chromatographic systems which supply a continuously changing sample composition over time. This dynamic coupling is the basis for the majority of MS-based proteomics analysis.

1.4.2.2 Matrix-assisted Laser Desorption/ionization

A second soft ionization method for introducing protein and peptides to the mass spectrometer is known as matrix-assisted laser desorption/ionization (MALDI). This technique was developed near the same time as ESI. However, unlike ESI which directly couples liquid samples to the mass spectrometer, the MALDI technique introduces solid samples through laser ablation. The co-crystallization of small organic molecules with the protein sample to form a solid matrix allows ionization through the use of a laser, resulting in desorption and ionization of the protein or peptide. A diagram of the MALDI processes is given in Figure 1.4. Several organic molecules are commonly used as MALDI matrices. The most common are sinapinic acid (SA) for proteins and α-cyano-4-hydroxycinammic acid (CHCA) for peptides. Both of these matrices absorb UV light supplied through a low repetition nitrogen laser (1-20 Hz), delivering a wavelength of 337 nm. Alternatively, a 355 nm high repetition (kHz) Nd:YAG laser can

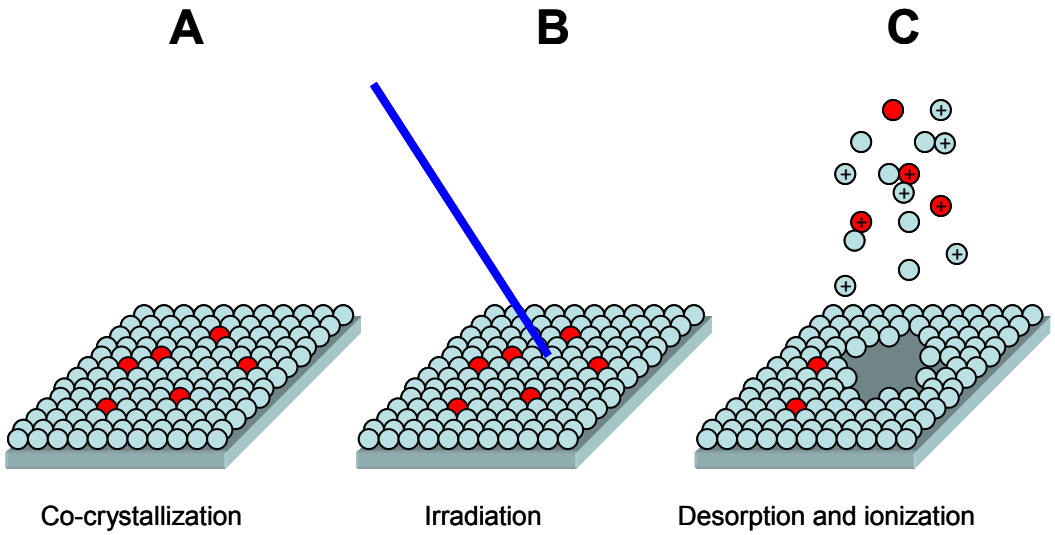

Figure 1.4 Mechanism of matrix-assisted laser desorption/ionization. First, (A) analyte is co-crystallized with an organic matrix. Then, (B) the crystal is irradiated with laser light with wavelength matching an absorbance wavelength of the matrix. Finally, (C) desorption of the matrix causes analyte to enter the gas phase where proton transfer ionizes the analyte prior to entering the mass spectrometer for analysis.

be used, as is seen for coupling some MALDI TOF instruments. There are several benefits to using MALDI over ESI for peptide and protein analysis, such as its higher tolerance to sample contaminants which would be detrimental to ESI analysis [67], as well as the formation of predominantly singly charged ions, making the interpretation of the mass spectrum of proteins much simpler. Perhaps most significantly is that MALDI allows greater freedom in the time spent for the analysis of sample, as it is static, while in LC-ESI experiments, online analysis of LC effluent entails a dynamic sample, limiting the time allowed for analysis of any given component. The advantages of MALDI over ESI for proteome profiling is a main component of the proposal of this thesis, which is presented in section 1.11. One of the disadvantages of MALDI is that, compared to LC-ESI, it is considered to be more difficult to couple MALDI to separation platforms such as liquid chromatography, although attempts have been made to improve upon this [68-70]. The coupling of MALDI to separation platforms is another major research theme of this thesis.

1.5 MS-based Protein Identification Strategies

As a tool for protein and peptide analysis, mass spectrometry has been applied in several different ways; the two major approaches are top-down and bottom-up MS. Top-down MS is the direct analysis of proteins at the intact level. This would seem the most obvious method to do MS-based proteomics. However, as will be discussed, this approach is fraught with difficulty. Conversely, bottom-up MS identifies proteins following enzymatic or chemical digestion of the sample, resulting in the formation of much smaller peptide fragments. These smaller protein segments are much easier to

analyze with low resolution MS instruments and therefore bottom-up peptide analysis is currently the most popular MS-based proteomics approach. A summary of the various strategies for conducting proteomics analysis is provided in Figure 1.5.

1.5.1 Bottom-up Proteomics Analysis

There are two main methods for bottom-up analysis of peptide fragments with MS, either using protein mass fingerprinting (PMF) or through peptide sequencing using tandem mass spectrometry (MS/MS).

1.5.1.1 Peptide Mass Fingerprinting

In PMF, a protein is enzymatically digested which produces a set of peptides which have predictable masses. The most common enzyme employed in this strategy is trypsin. Trypsin is an endoprotease which cleaves the amide bond on the carboxyl side of lysine and arginine residues, unless followed by proline. A MALDI MS spectrum of a tryptic digest of bovine serum albumin (BSA) is shown in Figure 1.6 as an example of PMF analysis. Other proteases such as Lys-C and chymotrypsin are popular choices depending upon the application. The popularity of trypsin is due to the guaranteed inclusion of at least one basic amino acid residue into the created peptides, ensuring there is a location for positive ionization through protonation. Due to the predictable nature of enzymatic digestion, the peptide products, and more specifically the masses of these products, are reproducibly generated from a given protein digestion. This list of peptide masses is unique (or nearly unique) to a particular protein, and can be used to identify the protein if its amino acid sequence is known. PMF utilizes the knowledge of protein sequences contained within computer databases to identify proteins based simply on the masses of the peptides observed following enzymatic digestion.

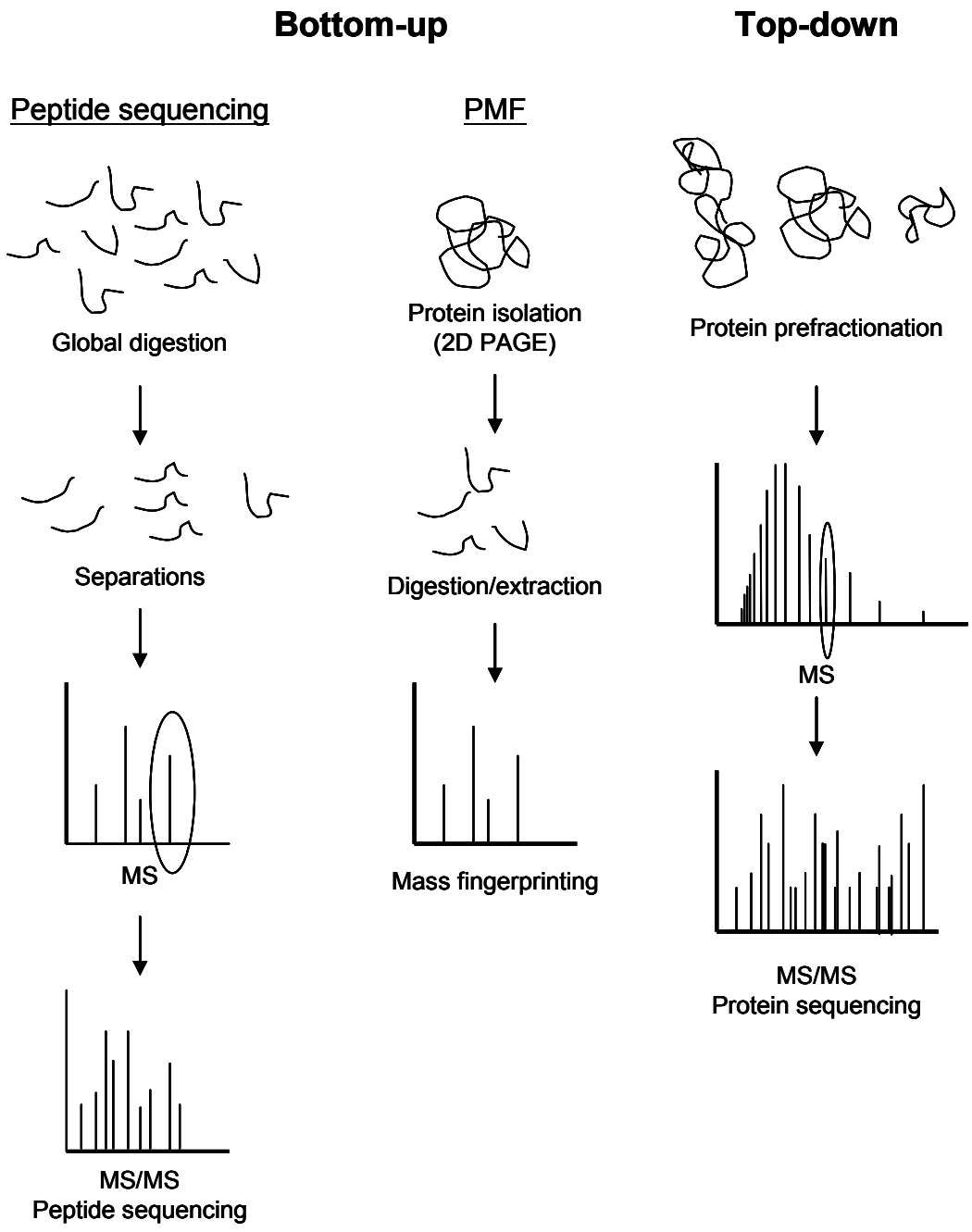

Figure 1.5 Two strategies for MS-based proteomics analysis. A bottom-up strategy uses either peptide sequencing or peptide mass mapping to identify proteins from their peptide fragments following enzymatic or chemical digestion. Top-down involves the direct analysis of intact proteins using mass spectrometry and requires high-performance MS instrumentation for high resolution and involves the interpretation of very complex tandem MS spectra.

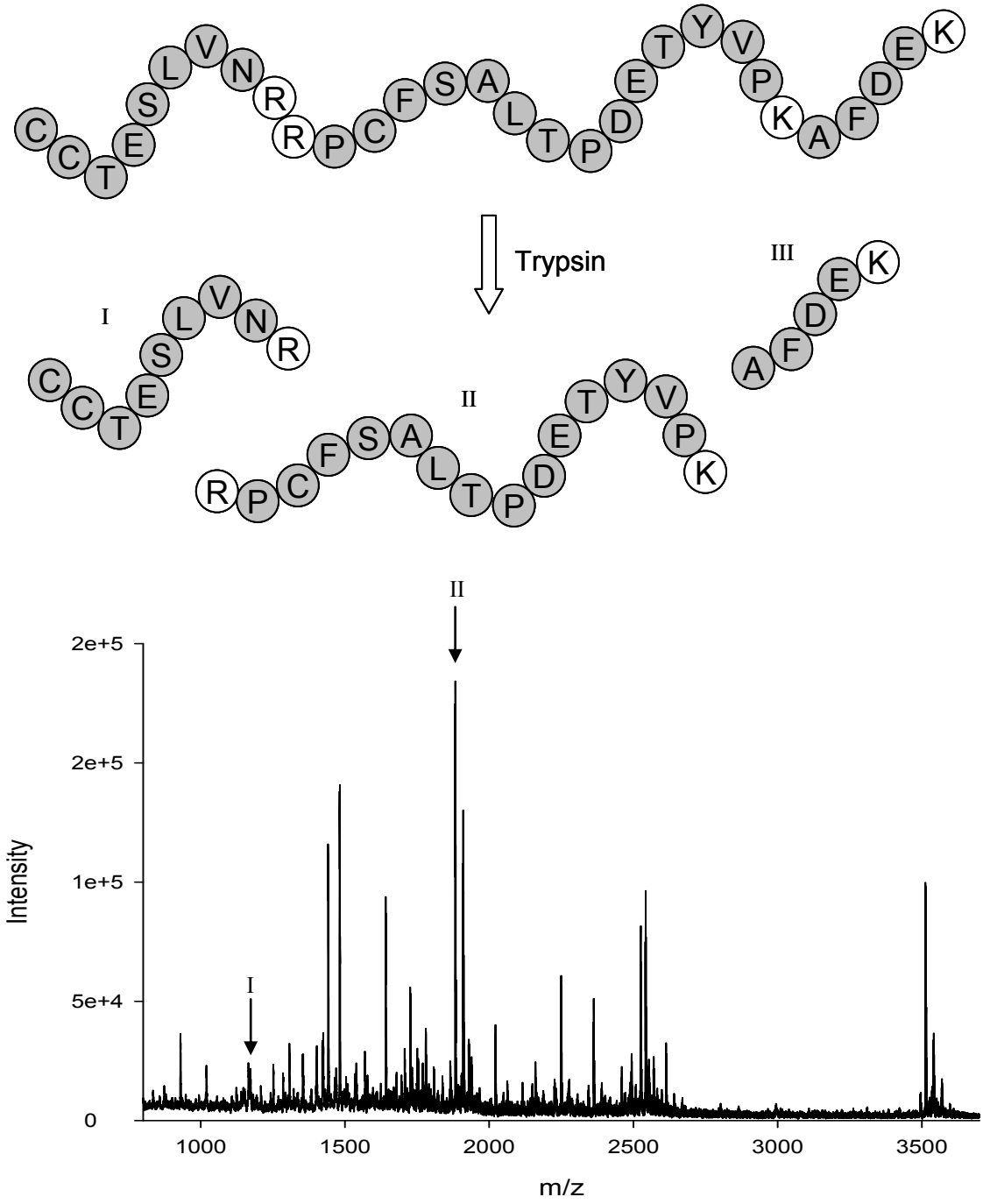

Figure 1.6 An example of a peptide mass fingerprint generated from the tryptic digestion of bovine serum albumin (BSA). The peptide sequence shown at top is a portion of the BSA amino acid sequence. Since trypsin cleaves on the carboxyl side of arginine and lysine, except if followed by proline, three peptides are expected to be produced following digestion. Accounting for a +57 amu mass shift from the carbamidomethylation of cysteine, peptides I and II are detected at masses of 1165.36 and 1881.91 respectively (highlighted by the arrows in the MALDI MS spectrum. The third peptide (predicted *m/z* 609.3) was below the mass range and not observed in this particular spectrum. The mass spectrum presented was collected by

There are, however, some important limitations to the PMF method. The protein must be isolated, otherwise, peptides generated from contaminating proteins would make accurate identification of the protein difficult. Thus, adequate separation of the sample is required. This is usually conducted through one or two dimensional gel electrophoresis, described in section 1.6. Also, unknown PTMs would result in the miscalculation of expected peptide masses, as would any mass accuracy error in the mass spectrometer. An alternative strategy which does not rely on extensive proteome purification is to identify peptides using tandem mass spectrometry, referred to as peptide sequencing.

1.5.1.2 Peptide Sequencing

In peptide sequencing, peptide ions detected in the mass spectrometer are isolated and then fragmented, typically through collision induced dissociation (CID). The process involves applying energy to promote collisions with an inert gas, such as helium, within a collision chamber, which is, in the case of ion trap MS, the mass analyzer itself. In peptide CID, MS/MS of positively charged ions produces a predictable fragmentation pattern produced from cleavage along the amide bond of the peptide (though other fragments are possible). The naming convention for these ions refers to b ions as the fragment containing the N-terminus, and y ions for the fragments containing the C-terminus [71]. A diagram illustrating the naming convention for peptide fragmentation is shown in Figure 1.7. Through calculation of the observed mass differences between fragmentation ions, along with its parent mass, the amino acid sequence of the peptide can be determined in a *de novo* fashion. Based on this amino acid sequence, it is possible to identify the protein that it was generated from, even in the presence of many other peptides. An example of the process of MS/MS sequencing of peptides using the triple

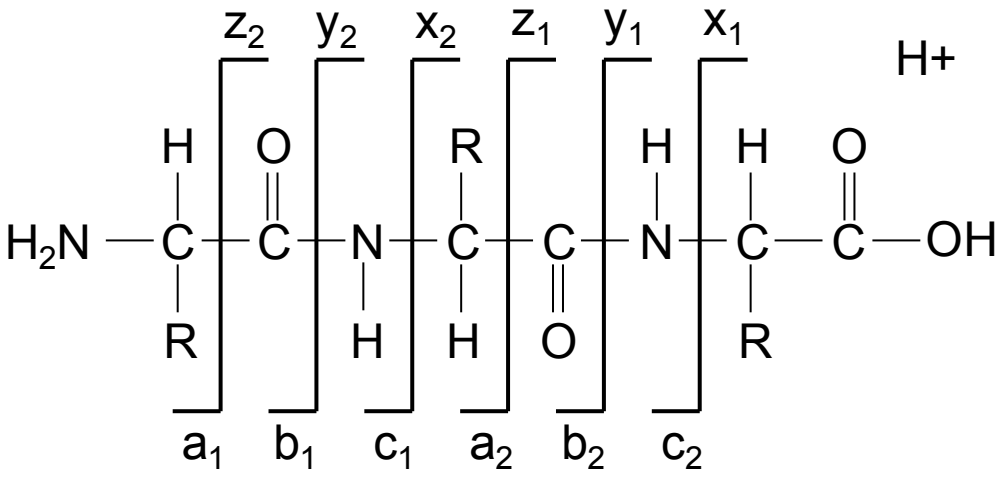

Figure 1.7 Naming convention for peptide fragment ions. The black lines indicate the location of fragmentation along the peptide. Fragments with charge on the N-terminal side to be either a, b, or c and peptide fragments with charge on the C-terminal side to be z, y, and z ions. CID fragmentation typically used in ion trap mass spectrometry produces predominantly b and y ions, which form along the peptide bond, producing fragments which have mass shifts related to the mass of the amino acid residues.

play method utilized for the peptide sequencing analysis presented in this thesis is given in Figure 1.8. Triple play refers to the three step process of: ion selection, charge state determination, and tandem MS fragmentation. Peptide sequencing does depend, however, upon having knowledge of the amino acid sequence of all proteins which may be expressed in a proteome. It is here where the reliance upon genomic databases becomes exceedingly valuable.

Using genomic information, as well as experimental, a proteome for an organism can be constructed. These databases (e.g., Swiss-Prot, PDB) are freely available on the web through servers such as the Expert Proteomics Analysis System (ExPASy) proteomics server (www.expasy.org) or the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). Current proteomics databases contain hundreds of thousands of non-redundant proteins from a wide range of organisms. The development of searching algorithms which can quickly analyze the thousands of MS/MS peptide spectra generated from one LC-MS/MS analysis have been developed, with popular search engines including SEQUEST [72] and MASCOT [73]. These algorithms match the experimentally collected fragmentation spectra to computationally generated spectra of potential peptide matches from proteomics databases. Scoring systems are employed to assess the quality of matches, in order to remove false positives matches. In the case of Sequent, Xcorr values represent the degree of cross correlation between the experimentally collected fragmentation spectrum and the computationally generated theoretical spectrum. Other scoring parameters are available, with a combination of these values used as filtering criteria to remove poorly correlated spectra. Decoy searches using databases which contain nonsensical proteins, such as a reversed proteome sequence,

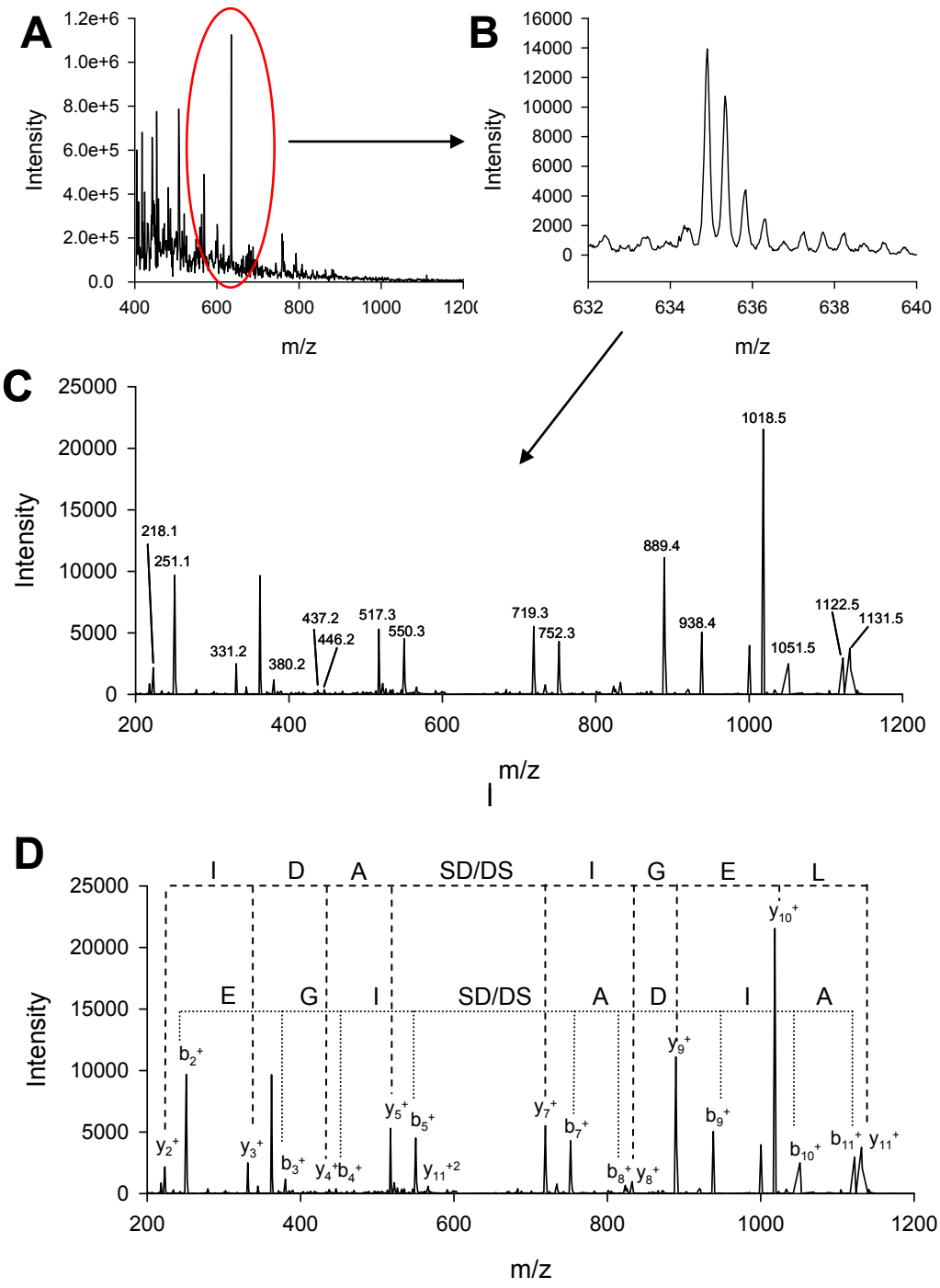

Figure 1.8 The triple play protocol for peptide sequencing by ESI-MS/MS. First, a peptide ion is selected from a general MS scan (A) and subjected to a higher resolution zoom scan (B) for charge state determination. The peptide is then isolated and fragmented by CID (C). The peptide sequence is determined based on the fragment ions produced and matched to its parent protein (D). In the example here, the peptide HLEGISDADIAK was found to be a sequence present in the protein Phosphoglycerate 1 mutase 1 from *S. cerevisiae*. The mass spectra presented were collected by the author.

have become a mandatory feature to assess false discovery rates [74].

1.5.2 Top-down Proteomics Analysis

Though bottom-up MS is currently the most popular approach to proteomics analysis, the use of top-down MS approaches is an increasingly active area of research [75-79]. The largest barriers to the direct analysis of intact proteins through MS is the lower sensitivity achieved from analysis of multiply charged ions (see Fig 1.3) and the difficulty associated with interpretation of fragmentation spectra. However, top-down MS is a promising approach to PTM analysis and reveals a more complete picture of the identified protein than bottom-up MS. Top-down analysis has progressed following developments in high resolution MS instrumentation, improved strategies for fragmentation of large molecules, namely electron capture dissociation (ECD) and electron transfer dissociation (ETD), and software platforms for MS/MS spectral interpretation. Of course, the need for analysis of intact proteins redirects all front-end manipulations towards intact molecules meaning peptide level separation can no longer be incorporated into a top-down workflow.

1.6 Separations in the Proteomics Workflow

Separations are performed prior to MS to simplify the mixture and improve the information gained from the proteome. Of the available methods for proteome separation, the two most widely used are two-dimensional gel electrophoresis (2D-PAGE) and multidimensional liquid chromatography (MDLC), with two-dimensional liquid chromatography (2DLC) being the most popular. 2D PAGE is a gel-based method which was first introduced by O'Farrell in 1975 and remains a popular choice for proteomics

analysis [80]. It involves the electrophoretic separation of proteins within a polyacrylamide gel with a first dimension of isoelectric focusing (IEF) following by size-based separation using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Although MS is used to identify selected proteins from the gel, staining methods are first used to visualize the proteins. 2D-PAGE is an excellent method for easy differential mapping of two proteomes for comparative analysis. This makes it a popular method for determining expression differences between two proteomes.

The latter method, 2DLC, is typically applied to the separation of peptides following enzymatic digestion of the proteome. In this case, ion exchange chromatography (IEC) is often used as a first dimension of charge-based separation followed by reversed phase liquid chromatography (RPLC), which is based on peptide hydrophobicity. The second dimension of RPLC separation is coupled to MS (typically through ESI) for high-throughput peptide sequencing. This method was first demonstrated by Anderegg and coworkers in 1997 [81], then refined by Yates and coworkers in 2001 [82]. This method has been given the name Multidimensional Protein Identification Technology (MudPIT) [83]. Unlike the gel-based approach, MudPIT is a comprehensive analysis technique, meaning that an attempt is made for all components to be identified through MS. This differs from the 2D-PAGE approach in which only selected protein spots of interest are subjected to MS analysis. This means that proteins which may not be detected by 2D PAGE, such as low abundant or extremes in MW or pI, may be detected by MS. Though the approach takes advantage of the sensitivity of MS for protein detection, it does not generate easy-to-read differential maps, meaning that it can be more difficult to detect differences between proteomes.

1.6.1 Two-dimensional Polyacrylamide Gel Electrophoresis

High resolution protein fractionation is most commonly achieved through 2D-PAGE as it affords high resolution separation of proteins, being capable of resolving thousands of proteins [84, 85]. This technique combines isoelectric focusing, as a first dimension of separation, and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as a second dimension.

1.6.1.1 Isoelectric Focusing

The general principle of isoelectric focusing in 2D-PAGE first involves the extraction and solubilization of proteins using a buffer containing a variety of protease inhibitors and solubilizing agents such as urea, 3-[(3-Cholamidopropyl) dimethylammonio]-2-hydroxy-1-propanesulfonate (CHAPS), dithiothreitol (DTT), and thiourea, to enhance yields and resolution [84, 85]. Proteins are then loaded onto a polyacrylamide or agarose gel strip, containing an immobilized pH gradient (IPG) [86, 87], cast into the gel prior to polymerization, with the use of immobilines. These immobilines consist of a small series of no more than eight acrylamido derivatives which contain both weakly acid carboxyl and weakly basic tertiary amino group as well as a strong basic or acidic group. Classically, the use of carrier ampholytes was employed for generation of the pH gradient [88, 89]. Combinations of thousands of species with a wide range of pI and buffering pH values would establish a pH gradient in solution following the application of an electric field. However, cathodic drift results in a non-linear pH gradient and decreases the pH range, reducing the resolving power and reproducibility.

After sample is loaded an electric field is applied across the gel strip causing proteins to migrate in a direction based on their charge. Protein charge is dependent upon

the pH; proteins at a pH below their pI have a net positive charge and migrate towards the cathode, while proteins at a pH above their pI are net negative and migrate towards the anode. The protein will eventually encounter a local portion of the gel strip in which the pH matches its pI, at which point its net charge becomes zero and no longer migrates through the gel. This process is illustrated in Figure 1.9.

Although this method provides high-resolution separations, proteins which have pI values greater or lower than the maximum or minimum pH of the IEF strip are lost during the focusing. Lost proteins cannot be visualized and subsequently cannot be analyzed using mass spectrometry. Also, hydrophobic proteins, such as integral membrane proteins, are more difficult to analyze using IEF due to their poor solubility.

1.6.1.2 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

As a protein separation technique, SDS-PAGE is very common not only as the second dimension of 2D-PAGE [90] but is frequently used as a stand alone method of protein separation [56]. In SDS-PAGE proteins are denatured using the anionic surfactant sodium dodecyl sulfate (SDS), which binds at a constant SDS to protein ratio of 1.4:1, and imparts a similar negative charge density to all proteins in the sample [91]. In 2D-PAGE, the focused IPG strip is first equilibrated with an SDS solution and then the strip is placed onto an SDS gel slab. An electric field is applied with the positive electrode at the bottom and the negative at the top, as shown in Figure 1.10. SDS bound proteins migrate through the gel at different rates, which are dependant upon their molecular weights. The distance a protein migrates in SDS-PAGE is based on the logarithm of its molecular weight, with heavier proteins traveling a much smaller distance than lighter

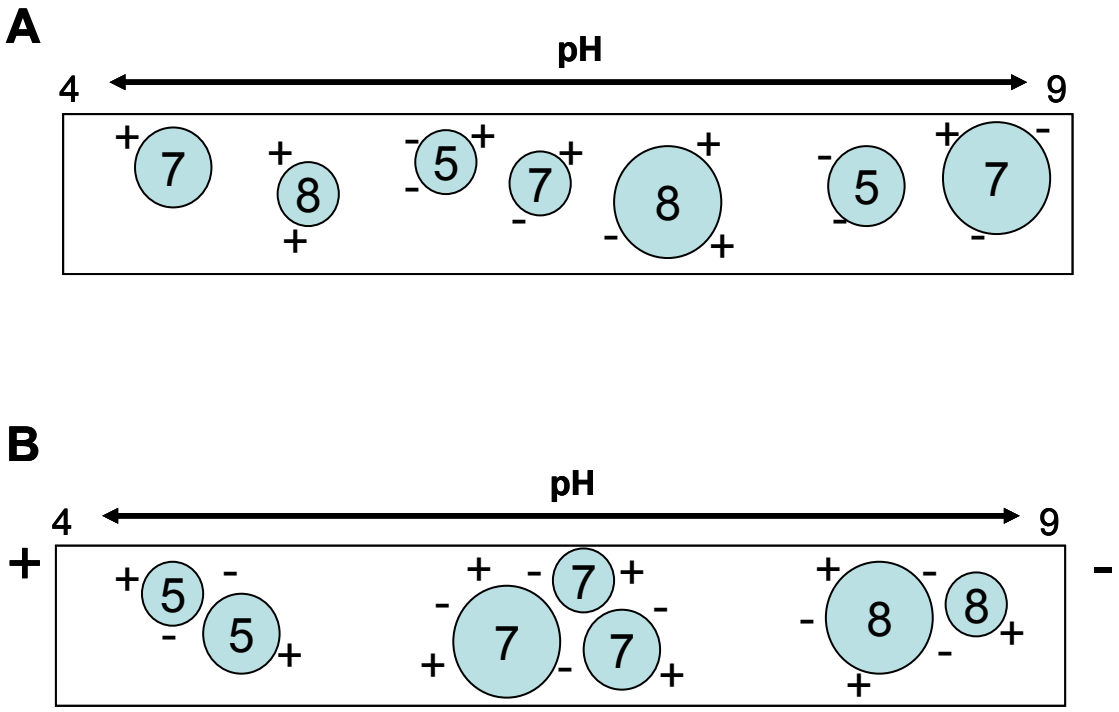

Figure 1.9 Illustration demonstrating isoelectric focusing. (A) Protein is loaded onto a polyacrylamide strip containing a pH gradient generated either through the use of immobilines or carrier ampholytes. (B) When voltage is applied, the proteins migrate. Proteins in a region with a pH above their pI will gain a net negative charge and migrate towards the positively charged anode. Proteins in a region with a pH below their pI will have a net positive charge and will migrate towards the negatively charged cathode. Proteins will continue to migrate until they reach a region with a pH matching their pI, at which point they become zwitterionic and no longer migrate within the electric field.

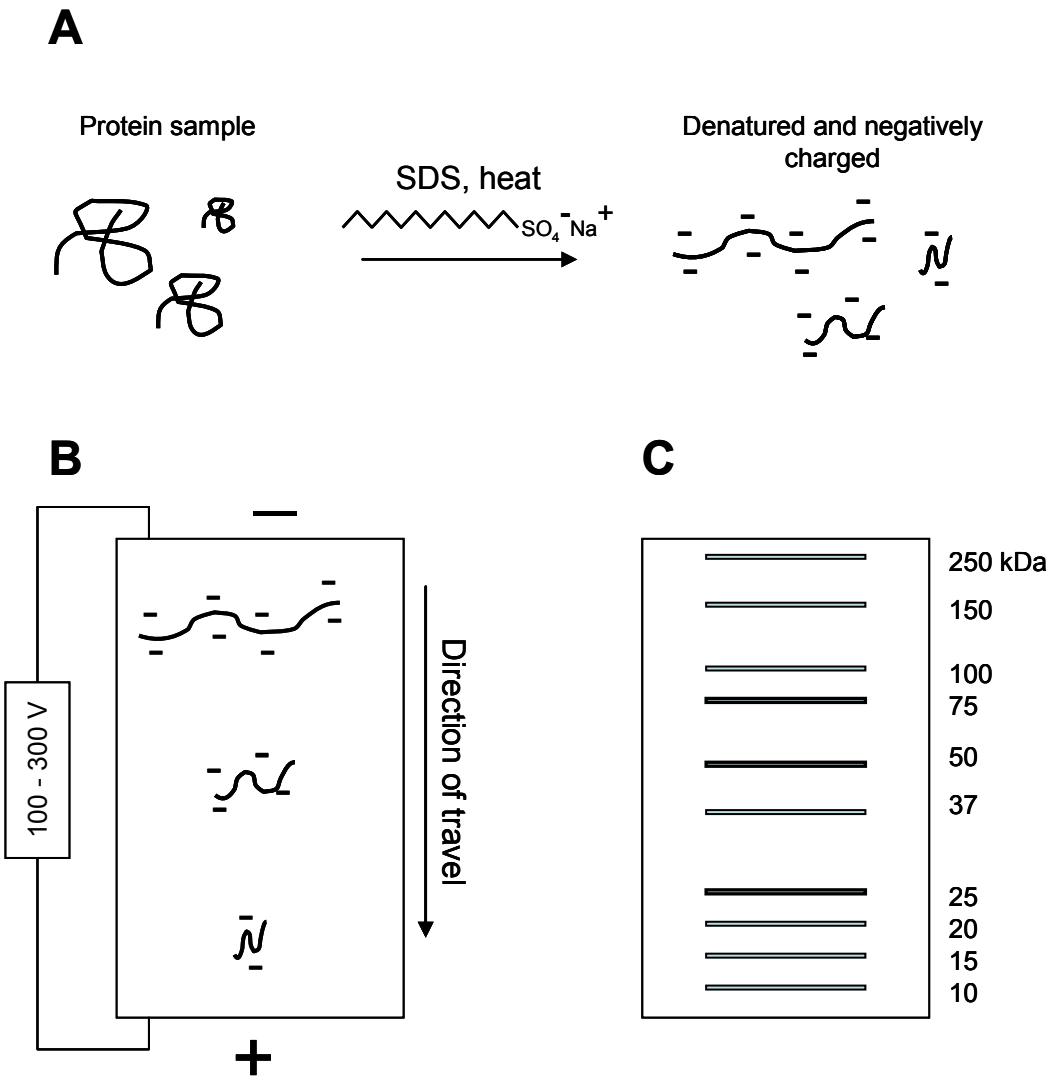

Figure 1.10 Diagram illustrating the principles of SDS-PAGE. (A) SDS is added to protein samples which bind at a 1.4:1 ratio causing the proteins to denature and acquire a net negative charge. (B) The sample is loaded onto the gel and a voltage is applied which causes the proteins to migrate with a velocity related to the log of their molecular weight. Small proteins migrate faster while larger proteins migrate slower. (C) Typical molecular weight range of a 15% T SDS-PAGE separation as shown from a standard protein molecular weight ladder.

proteins, given the same amount of migration time. The effective molecular weight range for proteins on SDS-PAGE can be tailored by adjusting the pore size of the gel. This is done by changing the concentration of acrylamide and cross-linking bis-acrylamide prior to polymerization. The total percentage by weight of acrylamide is referred to as %T. Of this, the percentage of the mass of acrylamide which is composed of cross-linker is given as %C. Typical concentrations are 15% T with 2.67% C for most SDS-PAGE separations with effective molecular weight separation between 10 and 250 kDa.

There are two different methods to performing SDS-PAGE, continuous or discontinuous, though the latter is the only method used throughout this thesis. In discontinuous PAGE [92, 93], the gel is in two parts, the upper part in which sample is loaded is cast with a much lower %T (typically 4%), referred to as the stacking gel. The lower portion, the resolving gel, has a %T between 10 to 15% for protein separation. The pH of the resolving gel is about 8.8, while the pH of the stacking gel is a lower at 6.8, buffered using Tris-HCl. During loading of the protien sample, a technique known as isotachophoresis is used to improve separation resolution. Variation in the migration speed of glycine in the stacking gel compared to the resolving gel causes the protein sample to compress into a thin layer before migrating through the resolving gel improving resolution.

1.6.1.3 In-gel Visualization and Mass Spectrometry

Following separation using 2D PAGE, proteins are visualized through staining techniques. These include Commassie staining [94], silver staining [95], fluorescence [96, 97], as well as autoradiography [98]. Of these, silver staining is likely the most popular with detection limits of low nanograms quantities of protein. However, the

dynamic range displayed in detection is poor, only spanning about one order of magnitude. Fluorescence and autoradiography staining methods provide similar sensitively as the silver staining method, but have a 1000 fold dynamic range . However, they are both expensive, and have health implications or additional safety requirements which are needed to perform autoradiography detection [98]. Other methods for improving gel staining include differential gel electrophoresis (DIGE) [99], which uses two different fluorescent tags to label two or more samples which are then combined and separated on the same 2D-PAGE gel. Since each sample on the gel is tagged with a different fluorescent dye, the proteins from each sample can be individually visualized. This process greatly reduces variability due to run to run differences and also improves throughput as more samples can be separated on a single 2D-PAGE experiment.

Mass spectrometry as a protein detection tool following 2D-PAGE is very common in proteomics. PMF was developed to use MS for detection of proteins following 2D-PAGE separation [100]. Although MS is used as the detection and identification tool; visualization using one of the methods described above is required beforehand in order to determine which protein spots are of interest. Once a protein spot is selected it must be cut away from the rest of the gel and is subject to an in-gel digestion procedure, with extraction of the resulting peptides.

1.6.2 Multidimensional Protein Identification Technology

The greatest advantage of MudPIT over the 2D-PAGE approach is the speed of protein identification. An entire proteome digest can not only be separated in a comparable amount of time to 2D-PAGE, but proteins are also identified immediately through online LC-ESI-MS/MS analysis. MudPIT employs a charge-based separation

strategy for the first dimension, ion exchange chromatography in this instance, which is then followed by reversed phase liquid chromatography (RPLC) as an orthogonal second stage of separation. With MudPIT, cation exchange (CX) is the most common choice, although anion exchange (AX) is certainly possible, and more recently, combinations of mixed-bed cation/anion exchange resins are used to improve the resolution of peptides on ion exchange [101]. IEC works by first loading the sample onto the chromatographic column while in an aqueous solution of low ionic strength. The solvent pH is critical. Acidic pH is required for cation exchange separations while basic pH is required for anion exchange separations. Once peptides are bound to the column they are eluted over time with stepwise or linear gradients of increasing salt concentration. Typically, volatile buffers, such as ammonium formate or ammonium acetate, are used for better MS compatibility, as well as to maintain buffer pH and elute peptides from the column.

The second stage of separation, RPLC, is based on the hydrophobicity of the protein. Typical stationary phases used are composed of linear hydrocarbon chains, such as C_4 (protein separations) to C_{18} (peptide separations), bonded to silica beads. Other reversed phase columns employ polymers such as polystyrene as the chromatographic material. Protein and peptide separations performed on reversed phase columns involve linear gradients from mostly water to increasing amounts of a more non-polar solvent, such as methanol (MeOH), or more commonly, acetonitrile (ACN). The more polar analytes elute earlier in the run while more hydrophobic ones elute later. A low concentration of acid (e.g., trifluoroacetic acid (TFA), formic acid (FA), acetic acid) in the range of 0.1% to 1% is used as an ion-pairing reagent to improve chromatographic performance. In the case of LC-MS, the addition of formic acid and acetic acid also

serves to improve ionization by increasing the concentration of protons in solution.

Although MudPIT allows for comprehensive analysis of all the peptides in the sample, it does not allow for the creation of easy-to-read differential maps between two proteomes in comparison to 2D-PAGE, as it is difficult to achieve comprehensive identification of the proteomes MudPIT cannot easily differentiate between protein isoforms and post-translational modifications, as only a portion of the protein is used to elucidate its identity. Also, the peptides which are generated from the high-abundant proteins in the sample can easily overwhelm the entire analysis, making it difficult to detect and identify low abundant proteins in the mixture.

In an attempt to resolve the issues surrounding both the 2D-PAGE and MudPIT approach to proteomics analysis, solution-based methods which separate the proteome at the protein level rather than the peptide level have been utilized and are gaining in popularity [102]. These electrophoretic and chromatographic approaches include solution isoelectric focusing, IEC, RPLC, size-exclusion chromatography (SEC), capillary electrophoresis (CE), as well as a variety of gel-based separation methods which involve collection in solution phase, with many possible combinations using each. A brief description of each method will be given.

1.7 Liquid Phase Protein Prefractionation

1.7.1 Ion Exchange Chromatography

Ion exchange has been used for decades for a variety of purification applications, ranging from the isolation of hormones [103] to viruses [104]. As mentioned previously, ion exchange comes in two forms, cation and anion exchange. In addition, both strong

and weak acidic or basic functional groups are available, referred to as strong and weak ion exchange respectively. Strong cation exchange (SCX) typically uses sulfonic acid functionality while weak cation exchange (WCX) uses weakly acid carboxylic acids. For strong anion exchange (SAX), quaternary amines are used, and for weak (WAX), primary, secondary, or tertiary amines. Protein separation strategies incorporating ion exchange for prefractionation are still often employed [105, 106].

Variations on IEC are available in the form chromatofocusing [107, 108], as well as pH gradient IEC [109]. These techniques employ the use of buffering species to produce pH gradients on column, or in the solvent, to elute proteins based on their pI. This has been particular effective for the high resolution separation of phosphorylated proteins [110, 111].

1.7.2 Solution Isoelectric Focusing

Many solution phase isoelectric focusing strategies are available for use in proteomics workflows. Early designs using combinations of carrier ampholytes and/or membranes with immobilines have been commercialized for preparative scale protein prefractionation. High resolution IEF can be achieved through the use of capillary isoelectric focusing [112, 113], although it is limited by a low loading capacity.

Larger volume designs for IEF separation are available, such as the Bio-Rad Rotofor, which uses a carrier ampholyte system, or combinations such as the ZOOM IEF (Invitrogen). More recently, the OFFGEL system was commercially developed by Agilent, which uses gel-based separation similar to the first dimension of two-dimensional gel electrophoresis, except that proteins are collected in solution following diffusion into sealed well chambers placed on top of the gel strip [114]. However, there is

concern with low recoveries of large proteins due to their low diffusion rate through the gel following separation and isoelectric precipitation application of this technique has been mostly performed for peptide separations [115-117].

1.7.3 Other Solution-based Techniques for Proteome Prefractionation

1.7.3.1 Chromatographic

As a protein separation tool, reversed phase liquid chromatography (RPLC) provides very high resolution and is a common part of protein fractionation strategies [118-120]. RPLC is typically applied as the second dimension of separation in many protocols, as a high resolution follow up to other separation methods, as well, it allows the removal of contaminants which may interfere with mass spectrometry.

Size exclusion chromatography (SEC) has been applied to the prefractionation of proteins for proteomics analysis [121-125]. An advantage of SEC is that separations can be carried out which retain the natural folding of the protein, preserving activity in the case of enzymes. However, SEC suffers from low resolving power, as separation is on a log scale requiring large differences in mass to achieve separation. Although, attempts to improve resolution through combinations of more columns have been performed [121].

Affinity chromatography, by its nature is not a technique which offers high resolution, but it can offer high selectivities in binding a target protein from a mixture, resulting in a two fraction system. Affinity chromatography is useful for either the purification of a particular protein or class of protein, or alternatively in proteomics analysis, for the depletion of high abundant proteins, such as the removal of albumins from blood serum [42, 126-128].

1.7.3.2 Electrophoretic

An alternative to chromatographic approaches for solution-based protein separations has been achieved through the use of electrophoresis, in particular capillary electrophoresis (CE) [129-132]. Although the resolving power of CE is very high, the loading capacity is low, which can limit the applicability of CE as a prefractionation tool. Peptide separations are more common in CE [130], and have also been coupled to mass spectrometry [133].

Size-based electrophoretic separations rely on the migration of protein through a porous gel in an applied electric field. Two commercially available devices are offered by Bio-Rad, the Prep Cell [134, 135] and the Whole Gel Eluter [136]. Recently, an alternative size-based electrophoretic method which allows the collection of protein in solution has been developed [137]. This device is termed gel-eluted liquid fractionation entrapment electrophoresis (GELFrEE). It is similar in design to the Prep Cell, but uses a much shorter gel tube to achieve protein separation, and the collection of fractions is discrete rather than continuous resulting in no increase in collection volume for later fractions. The design has also been multiplexed to achieve higher throughput separation [138].

1.8 Multidimensional Separations

The combination of orthogonal modes of separation results in the product increase in peak capacity from each separation mode. Many of the above mentioned solution-based protein separation methods have been coupled and are described below.

1.8.1 Charge/size

The combination of charge-based chromatographic and electrophoretic methods to further size-based separation have been performed, such as the case of IEC coupled to size exclusion chromatography [121], IEF in the form of the Bio-Rad Rotofor couple to the Prep Cell [139], or IEF coupled to GELFrEE [138]. These forms of multidimensional liquid phase separations become analogous to the 2D-PAGE proteome fractionation discussed in the previous chapter.

1.8.2 Charge/hydrophobicity

The coupling of charge separations to reversed phase chromatography is likely the most popular multidimensional liquid phase strategy in proteomics. This is due to the generally high loading capacity of ion exchange chromatography, and the high resolution of reversed phase, along with its ability to remove contaminating salts and buffers. Many combinations of IEC to reversed phase have been reported [81, 140, 141] as well as the coupling of chromatofocusing for high resolution separation of protein isoforms followed by non-porous reversed phase chromatography [109, 119, 142, 143].

1.8.3 Size/hydrophobicity

The chromatographic coupling of SEC to reversed phase [123], as well as electrophoretic coupling of the Prep Cell [144] to RPLC, has both been applied for the separation of proteins. However, due to the low orthogonality between size and hydrophobicity, multidimensional separations incorporating these separation platforms together are less common.

1.9 Quantitation Strategies for MS-based Proteomics

Ultimately, expression proteomics relies upon accurate quantitation methods for biomarker discovery. In 2D-PAGE, quantitation of proteins can be performed following visualization using staining procedures, allowing comparisons from one gel to the next. In mass spectrometry, differentiation between samples is performed by examining the relative signal intensity between mass-shifted peptides originating from the test and control samples. Label-free strategies are also being developed but these methods usually result in high standard deviations and low reproducibility.

It can be incorporated during cell growth, such as in SILAC (stable isotope labeling with amino acids in cell culture) which involves growing cells upon media with isotope enriched amino acids (e.g., lysine with enriched ^{13}C). Alternatively, isotopic or isobaric labeling is performed following protein extraction; the two most common strategies being ICAT (isotope-coded affinity tags), involving differential labeling of cysteines using heavy and light isotopic versions, and iTRAQ (isobaric tag for relative and absolute quantitation), which involves the labeling of free amines with an isobaric reagent. In iTRAQ, quantitation is performed by examining reporter ions in the fragmentation spectra. The use of isotope coded formaldehyde (CH_2O and CD_2O) for the methylation of peptide amino groups *via* reductive amination has also been reported [145-151]. Its incorporation into a quantitation strategy for proteome profiling is a central theme of this thesis, and as such will be explained in greater detail.

For the purpose of isotopic labeling, the addition of formaldehyde, or its deuterated counterpart, to a peptide mixture followed by reduction using sodium cyanoborohydride ($NaBH_3CN$) results in the replacement of the two hydrogen atoms of

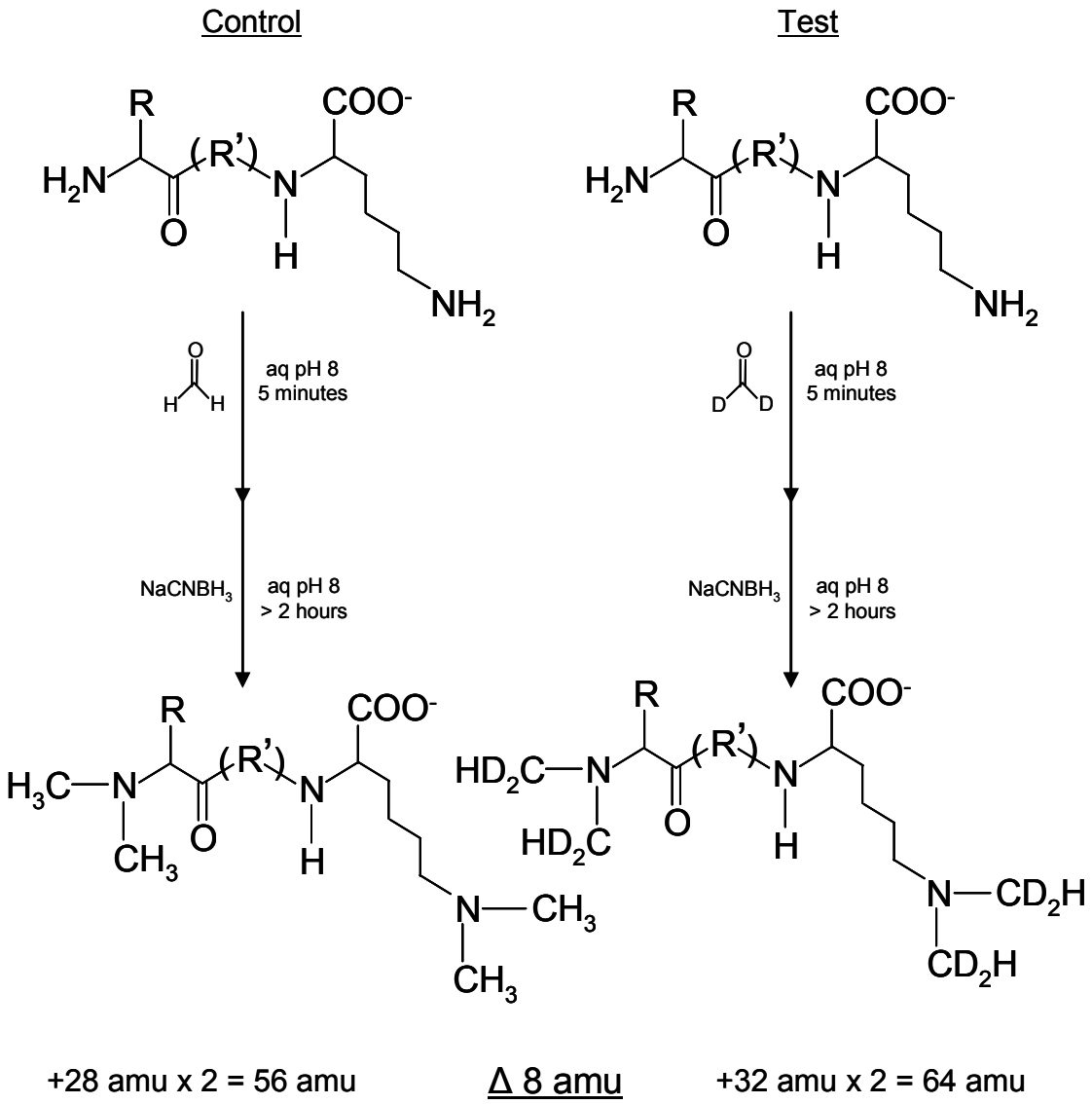

Figure 1.11 Dimethylation of lysine and N-terminus of a generic peptide. Formaldehyde, or deuterated formaldehyde, followed by sodium cyanoborohydride is used to generate a mass differential between peptides of a control and test sample. In the example above, a terminal lysine following tryptic digestion will have two available sites for reaction. Peptides which are products of cleavage following arginine will only have the N-terminus available for labeling. Each dimethylation generates a mass increase of +28 for the protonated formaldehyde and +32 for the deuterated formaldehyde, resulting in a mass differential of +4 amu between the peptides of the test and control samples.

the amine with a dimethyl group. The mechanism of this reaction is shown in Figure 1.11. The modification of R-NH_2 to R-$N(CH_3)_2$ or R-$N(CHD_2)_2$ results in a mass shift of either +28 for the protonated formaldehyde, or +32 for the deuterated species. This results in a mass differential of 4 mass units for each labeling site of the peptide. This is particularly convenient for tryptic digestions as many peptides generated using this enzyme will contain lysine residues, providing another labeling site in addition to the peptide N-terminus.

1.10 Biomarker Discovery

As previously mentioned, one of the practical goals of proteomics is the discovery of protein biomarkers which are indicative of particular human diseases, allowing for early diagnosis and improved treatment strategies. The road to biomarker discovery has two major components: 1) differential proteome profiling for potential biomarkers and 2) validation of these biomarkers over a large population of samples. A diagram illustrating these stages of analysis is shown in Figure 1.12. In the early stage of biomarker discovery the goal is to create a list of potential protein biomarker candidates. These candidates are obtained by performing a comprehensive proteome profiling of healthy and diseased proteomes from a small number of samples, searching for differences which may be indicative of the disease state. Once the list of potential biomarkers is obtained the late stage analysis subjects these candidate proteins to a more thorough and rigorous analysis over a wider population and potentially over a greater range of conditions such as disease stage, sex, age, etc. This stage of analysis uses targeted strategies which focus only on these candidate proteins, ignoring the remainder of the proteome. The use of MRM mass

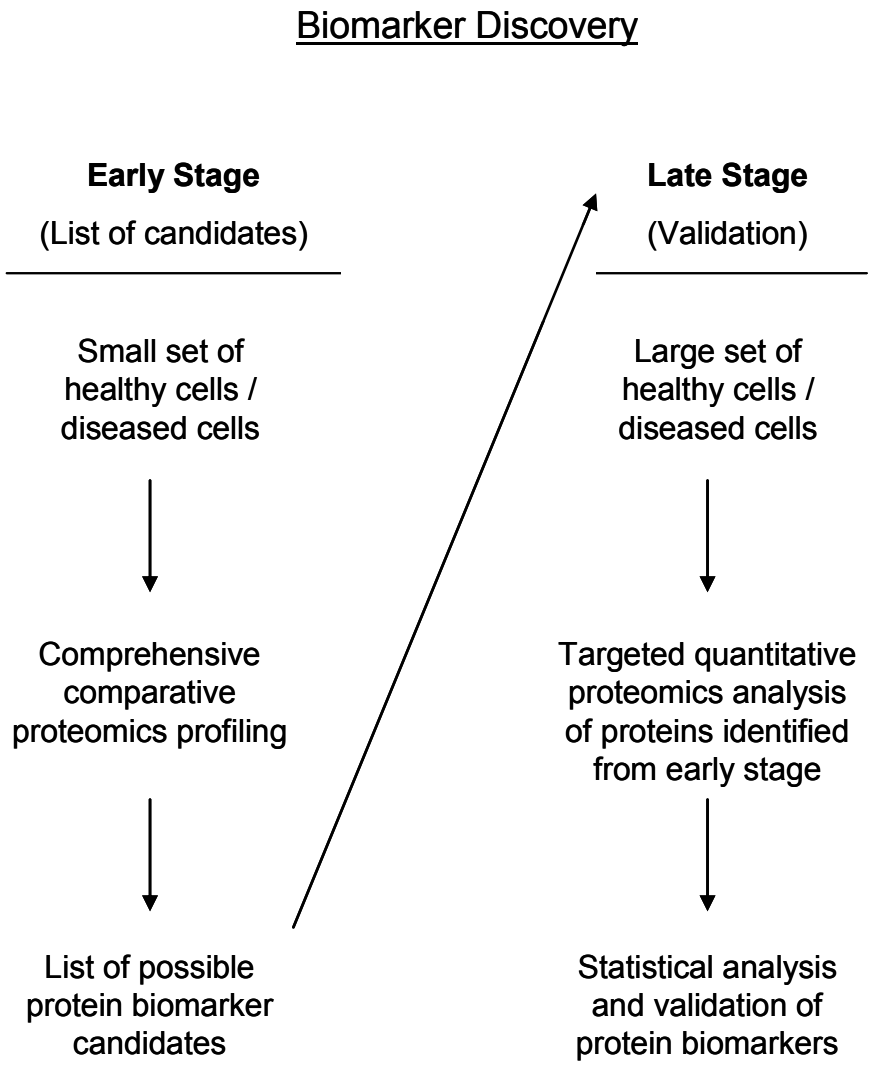

Figure 1.12 Diagram showing the stages for protein biomarker discovery. Early stage analysis involves the comprehensive comparative analysis of the healthy and diseased state of a relatively few samples to generate a list of potential biomarkers. These proteins are then subjected to targeted analysis over a large sample set in the late stage analysis. Validation of those proteins which constitute reliable biomarkers for a particular disease is obtained through statistical analysis of the quantitation data.

spectrometry for targeting particular peptide products from the target proteins is an area of interest for routinely quantifying potential biomarkers in a large number of complex samples [152-155]. However, the largest bottleneck is currently at the stage of generating lists of potential protein biomarkers. The two proteomics workflows which were described in Figure 1.1 are possible methods for profiling discovery of potential biomarkers. There are, however, significant drawbacks to these methods. While the gel-based 2D-PAGE approach is very good at generating easy to read differential maps of proteomes, the detection of the proteins largely relies on relatively insensitive staining methods. Also, 2D-PAGE followed by in-gel digestion and LC-MS/MS analysis of individual protein spots is labour-intensive and difficult to automate resulting in low throughput. In contrast, the bottom-up peptide sequencing approach relies on the more sensitive MS-based detection. However, it has the disadvantage of becoming overwhelmed by the large number of components generated from global enzymatic digestion. Additionally, because tandem MS peptide sequencing is performed on all peptides present in the sample, the large amount of data generated, as well as the presence of high abundant components present in the sample, can cause difficulties in detecting the components of interest.

The use of intact level proteome prefractionation has been mentioned as a way of combining the benefits of protein separations with the high-throughput analysis of peptide level bottom-up LC-MS/MS. However, as more separation modes are combined the throughput of analysis is reduced due to the increasing number of generated fractions which must be subsequently analyzed by later stages of separation. Also, the number of fractions which can result causes a bottleneck in bottom-up peptide sequencing as the

potentially hundreds of generated fractions from multidimensional protein separations would need to be analyzed.

Obviously there is a need to develop improved methodologies which can provide high-throughput proteome separations in solution while reducing the bottleneck in MS analysis for profiling of potential protein biomarkers. The development of such a platform is the subject of the research proposal and the work presented in this thesis.

1.11 Research Proposal

The research presented in this thesis represents efforts towards the development of a high-throughput proteomics workflow for differential proteome profiling using liquid phase separations and MS-based detection. This method aims to couple the best feature of 2D-PAGE, namely its ability to generate easy-to-read proteome maps, with the sensitivity of MS-based identification and quantitation. The work presented in this thesis aims to develop the tools necessary for the creation of a proteomics workflow which incorporates liquid phase proteome separation with MS-based profiling to enable the high-throughput analysis of proteome samples for biomarker discovery.

The tools evaluated and presented in the following 4 chapters cover distinct aspects of the proteomics workflow, beginning with intact protein separations, then moving on to the coupling of multidimensional and multiplexed separations with MALDI MS, the utilization of MALDI MS for proteomics profiling and ending with an investigation into improving protein digestion methods.

As a prefractionation tool, ion exchange chromatography is very popular as a first dimension of separation. However, there is little work in the literature which investigates

its effectiveness as a comprehensive prefractionation tool for proteomics applications at the intact protein level, although peptide level separations have been investigated thoroughly in the form of MudPIT. In chapter 2, ion exchange chromatography is evaluated as a first dimension of proteome prefractionation for use in a multidimensional liquid phase platform for comprehensive proteomics analysis. The fractionation ability, bias, and predictability of separation is determined, along with their implications for combining several stages of solution phase prefractionation methods.

In chapter 3, a design for a high-throughput parallel chromatographic separation system is presented. This system is intended to improve throughput of the multidimensional separation strategy by performing the second dimension of separation in parallel, rather than sequentially. Also presented in chapter 3 is a strategy for the coupling of the parallel chromatographic system to MALDI MS. A well plate device which attaches to the MALDI target allows for the collection of high volumes from the parallel chromatographic system. This allows the use of larger columns with greater loading capacity to be incorporated into the separation workflow. It also limits the number of sample manipulation steps by removing the need to transfer sample to the MALDI target following separations.

Although MALDI MS has been used effectively for the bottom-up analysis of peptides, it is not nearly as effective as LC-ESI-MS strategies in terms of quality or quantity of data. However, MALDI has advantages over ESI which makes it desirable to incorporate into proteomics workflow. These advantages include the static nature of samples, allowing for more freedom in data collection, as well as the generation of simpler MS spectra due to the formation of predominantly singly charged ions. In chapter

4, the evaluation and comparison of MALDI MS/MS and ESI MS/MS is performed in the context of a comprehensive peptide identification strategy to highlight the strengths and weakness of each ionization technique. Following this comparison, an alternative strategy is presented which uses MALDI as a technique for fast profiling of isotopically labelled peptides without the need for peptide sequencing. This strategy aims to provide differential analysis of proteome samples in a similar fashion which is available in the 2D-PAGE approaches, but without the limitations of using gel-based separations and with all the benefits of MS-based detection.

Lastly, chapter 5 explores the effects of organic solvents upon the enzymatic efficiency of trypsin for protein digestion. It has been reported that the use of 80% ACN for tryptic digestions results in improved sequence coverage of proteins following LC-MS/MS. The investigation of this for application into our workflow has resulted in an alternative explanation for the cause of improved data and reveals a novel strategy for proteome digestion to enhance protein identification and sequence coverage.

Chapter 2

Optimization and Evaluation of Ion Exchange Chromatography for Proteome Prefractionation

2.1 Introduction

Multidimensional separations incorporating both electrophoretic and/or chromatographic techniques are a common approach for the prefractionation of intact proteins [102]. The combination of orthogonal (or near orthogonal) separation technologies affords dramatic improvement in the total peak capacity of the platform [156, 157]. The effectiveness of each separation stage in a tandem platform is critical to the overall optimization of a proteomics workflow, both in terms of throughput and total number of protein identifications achieved [158-161]. The first dimension of separation in a multistep analytical workflow is most critical; poor fractionation during the first dimension, defined both in terms of low protein recovery or poor resolution, will have negative consequences on all downstream manipulations. This ultimately results in an impaired ability to characterize the proteome with mass spectrometry. Since peptide-level separation remains the most popular strategy for shotgun proteomics, most optimization studies have focused on this form of separation, rather than protein prefractionation, which is the focus of this current chapter.

The charge-based fractionation of proteins or peptides is a common choice as the first dimension of a multidimensional separation platform. In the classic two-dimensional gel electrophoresis experiment, for example, isoelectric focusing constitutes the first dimension, followed by size separation through SDS-PAGE [80]. The limitations of 2D-PAGE have prompted researchers to explore liquid-based alternatives for protein

separation [19], including electrophoretic (*e.g.* solution isoelectric focusing) and chromatographic (*e.g.* ion exchange) approaches [137, 126, 103 102, 120]. At the peptide level, charge-based separations followed by reversed-phase LC is a favoured platform for multidimensional proteome analysis [110]. Methods involving the pairing of strong cation exchange [110, 192], mixed bed platforms consisting of cation and anion exchange material [101], or alternatively, capillary isoelectric focusing [112] with RPLC illustrates the popularity of charge-based separation in the first dimension.

Undoubtedly, as a peptide separation tool, IEC has witnessed tremendous success in proteome workflows. However, there exists a strong motivation to perform proteome prefractionation at the 'intact protein' level (*i.e.* prior to enzymatic cleavage). Not only is intact separation critical for top-down proteome detection [144], but proteome prefractionation is an effective tool for improving the detection of low-abundance proteins [111]. A disadvantage of peptide-level separation is that peptides generated from high-abundance proteins will distribute across a majority of fractions. In contrast, at the intact level, these proteins have potential to be isolated in particular fractions, enabling the low-abundance 'hidden' proteome to be mined. Furthermore, differentiation of protein isoforms and post-translational modifications is possible, allowing a more detailed investigation of protein pathways, as shown for example through Lubman's use of chromatofocusing to differentiate phosphorylated proteins [119].

IEC has a long and familiar history as a purification tool for a wide range of applications, ranging from early isolation of hormones [103] and purification of viruses [104]; to highlight just a small portion of applications. However, its use as a comprehensive proteome prefractionation platform prior to MS has only recently been

explored. Here, the aim of IEC is no longer to isolate a single target protein, but to achieve high resolution and high recovery across the entire proteome mixture. To this end, variations on conventional salt gradient IEC have been presented and form the basis of effective solution-based proteome workflows. These strategies include Lubman's 2D platform, which pairs chromatofocusing with the Rotofor [110, 111] (a size-based electrophoretic separation platform), as well as the use of pH gradient IEC [109]. To date, these methods have yet to gain widespread adoption, perhaps owing to the reduced solubility of proteins at their isoelectric point, as well as the limited pH ranges of chromatofocusing buffers, reducing the comprehensive fractionation ability of these platforms. The addition of solubilizing agents such as glycine, taurine, or betaine can partly resolve recovery issues, but none of them are applicable over a wide pH range, as they would become charged passed a certain pH value and interfere with the chromatofocusing separation. The addition of detergents is possible, but they complicate the downstream workflow by interfering with reversed-phase separations and mass spectrometry analysis and therefore must be removed. While MS-compatible surfactants, such as RapiGest SF, PPS, and Invitrosol are useful for solubilization of protein samples prior to analysis, and have been shown to improve the number of peptide identifications following bottom-up peptide analysis [119], their use as an additive in an IEC buffer system would present concerns with associated cost. More recent developments of IEC for protein prefractionation include combinations of SAX and SCX with SDS PAGE [105] as well as application for protein depletion prior to further proteomics analysis [106]. Interestingly, despite the many examples which incorporate IEC as a

prefractionation tool into a proteome workflow, a critical examination of the effectiveness and applicability of IEC for proteomics has yet to be performed.

Comprehensive proteome characterization experiments are generally concerned with the detection of *all* protein components present in the mixture. An equally important concern is that the detection platform employed provides a maximal degree of experimental throughput. To succeed at both of these goals, one must critically evaluate all aspects of the analytical workflow. Here, focus is directed at the optimization of a charge-based prefractionation method, ion exchange chromatography. In this chapter, a critical evaluation of conventional salt gradient IEC as an effective proteome prefractionation tool is performed for both strong cation and strong anion exchange chromatography. Given the unknown nature of a complex proteome, in terms of number and type of proteins, optimization involved the use of well-characterized protein standards to facilitate the evaluation. In doing so, the recovery of individual proteins, as well as resolution and peak capacity of the separation is evaluated. Optimization of IEC maximizes both resolution and recovery over a range of proteins and examines the buffer pH, addition of organic modifiers, temperature, and slope of salt gradient. An emphasis is made on collecting the minimal number of fractions with respect to the true resolving power of IEC, thereby maximizing the throughput of the system.

2.2 Experimental

2.2.1 Materials

All protein standards, including trypsin (TPCK treated, cat. T8802), and

lyophilized *Saccharomyces cerevisiae* (cat. YSC2) were purchased from Sigma (Oakville, Canada). Milli-Q grade water was purified to 18.2 $M\Omega cm^{-1}$. HPLC grade solvents were purchased from Fisher Scientific (Ottawa, Canada). Reagents for gel electrophoresis were from Bio-Rad (Mississauga, Canada). All other chemicals were from Sigma and were used without further purification.

2.2.2 Sample Preparation for IEC

Lyophilized cells of *Saccharomyces cerevisiae* were suspended in 20 mM HEPES buffer pH 7.6, and proteins were extracted by three passes of the cells in a French press device (8000 psi). The extract was centrifuged, and the supernatant filtered by passing through a 0.45 μm cellulose acetate membrane. A final protein concentration of 10 mg/mL was obtained, as determined using the Bio-Rad protein assay. For chromatographic separation, a mixture of protein standards was prepared (250 μg/mL per protein) in the initial IEC solvent systems. The final concentration of the yeast samples, in their respective IEC buffers, just prior to separation was 2.5 mg/mL.

2.2.3 Protein Digestion

Protein fractions from IEC were individually precipitated overnight at -20^{o}C through addition of three volumes cold acetone. The resulting pellet was dissolved in 50 μL of 100 mM NH_4HCO_3. The sample was reduced with 2.5 μL of 200 mM DTT to break disulfide bonds which where then alkylated with the addition of 5.5 μL of a 200 mM iodoacetamide solution. Digestion of the fractions was performed overnight at 37^{o}C through addition of 2.5 μg trypsin per sample.

2.2.4 Liquid Chromatography

An Agilent 1100 quaternary HPLC system with a photodiode array detector set at

214 nm (Palo Alto, CA) was used. SCX was performed on a 1.0 × 100 mm ThermoFisher Biobasic SCX column (5 µm packing, 300 Å pore size) or a self-packed PolySULFOETHYL A column (5 µm packing, 300 Å pore size) with material obtained from The Nest Group, and SAX on a 1.0 × 100 mm Vydac SAX column (5 µm packing, 900 Å pore size), each at a flow rate of 50 µL/min. Weak cation exchange (WCX) was performed on a self-packed PolyCAT A column (1 mm × 100 mm, 5 µm packing, 300 Å pore size), also obtained from The Nest Group, operating under similar conditions to SCX. Optimal conditions for IEC, including solvent pH, percent acetonitrile, column temperature, and slope of salt gradient were determined as described in the results section of this chapter. For SCX, an acetate buffer (sodium acetate/acetic acid pH 4 - 7) was used, with the exception of pH 3 which used a 0.1% TFA solution, with 2 M NaCl added to comprise solvent B. SAX used a Tris-HCl buffer (pH 6 - 10) with 1 M NaCl added in solvent B. RPLC separations were performed on a 1.0 × 100 mm Vydac C_{18} column (5 µm packing, 300 Å pore size) using a linear gradient between solvent A (0.1% TFA/water) and B (0.1% TFA/ACN) starting at 10% B and increasing to 70% over 30 minutes.

2.2.5 Solution Isoelectric Focusing

A custom eight chamber sIEF device was used [162]. Yeast was prepared to a concentration of 2.5 mg/mL in 4 M urea, 50 mM DTT, 1% CHAPS, and 1% Biolyte 3/10 carrier ampholytes (Bio-Rad, Hercules CA), of which 3.2 mL was loaded into the device. The anode and cathode buffers consisted of 7×10^{-3} M H_3PO_4 (pH 2.5) and 20 mM lysine (pH 10.0) respectively. The device was operated at a constant power of 2 W until the current dropped below 1 mA. Fractions were collected and subject to RPLC, as described

in section 2.2.4.

2.2.6 LC-MS/MS and Database Searching

An Agilent 1200 LC nanopump and autosampler were coupled to a nanospray ionization source (2.5 kV) and MS data were collected using a LTQ linear ion trap mass spectrometer (ThermoFisher, San Jose, CA). Peptide separation was performed on a ThermoFisher 0.180 × 100 mm C_{18} capillary column (5 μm beads, 300 Å pore size) at a flow rate of 2 μL/min. Separation consisted of a linear gradient between solvent A (0.1% formic acid/water) and B (0.1% formic acid/ACN), starting with a hold at 2% B for 5 minutes then increasing to 42% B over 40 minutes, then to 80% B in 2 minutes. MS/MS spectra were obtained in data-dependant mode over the entire LC run by selecting the top 5 ions in the MS spectrum (350-2000 *m/z*) and fragmenting at 35% normalized collision energy. Precursor *m/z* values (-1.5, +2.5 amu) were placed on an exclusion list for 3 minutes. The eight fractions obtained from sIEF were each analyzed twice, while IEC fractions were analyzed once, providing a total of 16 LC-MS/MS runs for each of the charge-based separation platforms employed.

Data were searched with Bioworks 3.2 software (ThermoFisher) using the SEQUEST algorithm against the *S. cerevisiae* database. Database searching conditions were as follows: Precursor mass tolerance 1.5 amu; fragment ion tolerance 1 amu; up to 3 missed cleavages per peptide. Peptides were filtered as follows: Xcorr 1.9 (+1 charge ions), 2.2 (+2), 3.75 (+3); $\Delta CN \geq 0.1$; $Rsp \leq 4$; peptide probability $\leq 5\times10^{-4}$. Furthermore, two or more unique peptides, being assigned to the same protein, were required for a positive identification of the peptides. In cases where a unique peptide sequence could be assigned to multiple proteins, the peptide was assigned to the protein of highest molecular

weight. Using these strict criteria, no peptides or proteins were found in a yeast decoy database search, yielding a calculated false discovery rate of zero.

2.3 Results and Discussion

2.3.1 Optimization of IEC

For IEC to be properly evaluated as an effective tool for protein prefractionation, optimal separating conditions, which maximize both resolution and protein recovery, must first be determined. Optimization of the separation parameters for both strong cation and strong anion exchange was performed using a set of standard proteins, listed in Table 2.1, along with their molecular weight and pI. In salt gradient IEC, the two important variables influencing separation are the buffer pH of the running solvents and the profile of the salt elution gradient. The column temperature, as well as the addition of organic modifier (acetonitrile) to the solvent system, were also examined. As a means of quantifying resolution, the peak capacity of the separation was determined. An empirical value was obtained by dividing the total time of the elution window (last protein eluted subtracting first protein eluted) by the average peak width for all protein standards. For recovery, the protein yield of the individual protein standards was assessed based on the peak areas from the resulting chromatograms, and choosing operating conditions which maximize recovery for all proteins. A detailed assessment of protein yield was determined using the optimal separation conditions, as discussed in section 2.3.3.

Table 2.1 Protein standards used for optimization of IEC.

#	**Protein name**	UniProt Acession #	MW (Da)	pI
I	amyloglucosidase (*Aspergillus niger*)	P69328	65,789	4.17
II	bovine serum albumin	P02769	66,433	5.60
III	ovalbumin (chicken egg white)	P01012	42,750	5.19
IV	carbonic anhydrase (bovine)	P00921	28,982	6.40
V	α-casein (bovine)	P02662	22,974	4.91
VI	β-lactoglobulin (bovine)	P02754	18,281	4.83
VII	myoglobin (bovine)	P02192	16,946	6.97
VIII	cytochrome c (horse)	P00004	11,701	9.59
IX	lysozyme (chicken)	P00698	14,313	9.32
X	ubiquitin (bovine)	P62990	8,564	6.56
XI	Insulin (bovine)	P01317	5,739	5.39

Protein information obtained from ExPASy.org

Validation of IEC optimization was carried out through prefractionation of an extracted yeast proteome, with identification using bottom-up LC-MS/MS analysis. The optimized separation conditions determined for IEC are provided in Table 2.2.

2.3.1.1 Optimization of Buffer pH

The pH of the separation buffer solution determines the total charge of the protein, and thus is expected to be the most important parameter in IEC separations for intact proteins. The pH values examined for SCX ranged from 3 to 7, and for SAX ranged from 6 to 10. In contrast to protein purification strategies, where the pH can be tailored to maximize resolution and recovery for one particular protein, a proteome prefractionation strategy necessitates maximum performance for all proteins in the mixture.

The solvent pH had a major impact upon the retention and binding of proteins in the SAX and SCX separations. At extreme pH values tested for IEC, there were significant losses owing either to irreversible protein binding or to non-retention of proteins on the column. This is illustrated from the IEC chromatograms of the 11 protein mixture shown in Figure 2.1. Of interest from this figure is that at a low pH of 3, many proteins were fully retained on the SCX column and could not be eluted regardless of salt gradient employed. Interestingly, of the proteins observed in this chromatogram with generally well-behaved peaks (i.e., narrow and symmetrical peaks), namely lysozyme, insulin, and cytochrome c, one would not predict their retention based on protein pI. This low pH condition is often employed for SCX separation of peptides, and has been applied to protein-level SCX separation [106]. Clearly, this pH is not suitable for protein-level separation since most proteins bind too strongly to the column. Conversely, at pH 7,

Table 2.2. Optimal Conditions for SAX and SCX.

Technique	Solvent	Temperature	Separation Conditions
SCX	A) 20 mM NaAcetate pH 6 B) A with 2 M NaCl	24°C	50 mM/min
SAX	A) 20 mM Tris-HCl pH 8 B) A with 1 M NaCl	24°C	10 mM/min

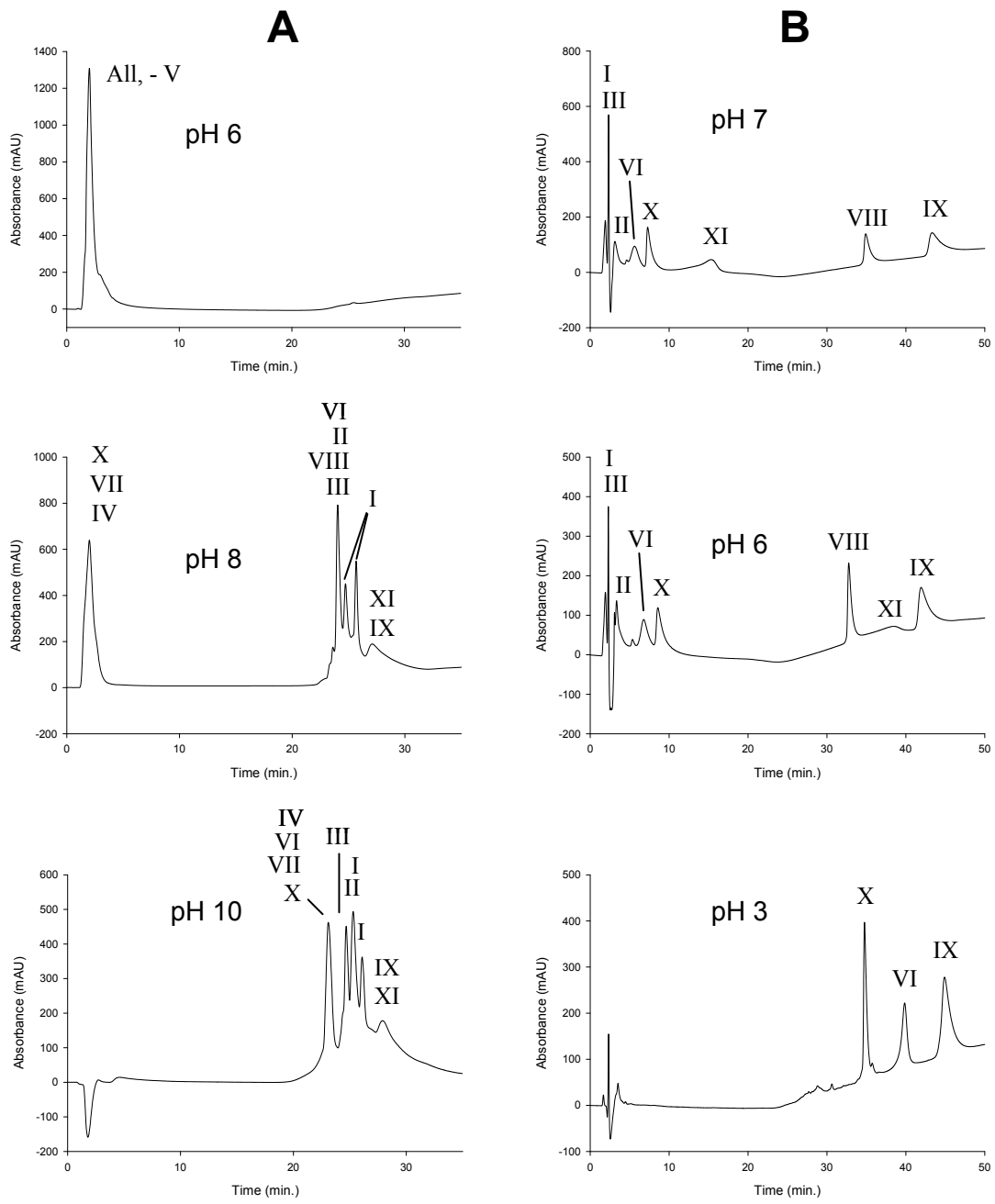

Figure 2.1 Chromatograms of the IEC separation of the 11 protein standard mixture. Separations using (A) SAX at pH 6, 8, and 10 and (B) SCX at pH 7, 6, and 3, with a 50 mM/min salt gradient at room temperature are shown. At low pH for SAX and high pH for SCX, the majority of the protein standards did not bind to the IEC column. With decreasing pH for SCX or increasing pH for SAX more proteins bound to the column and were observed. However, as the pH further increased or decreased for the respective separation, several protein standards showed decreased recovery as evidenced by the non-observance of a well defined elution peak in the chromatogram.

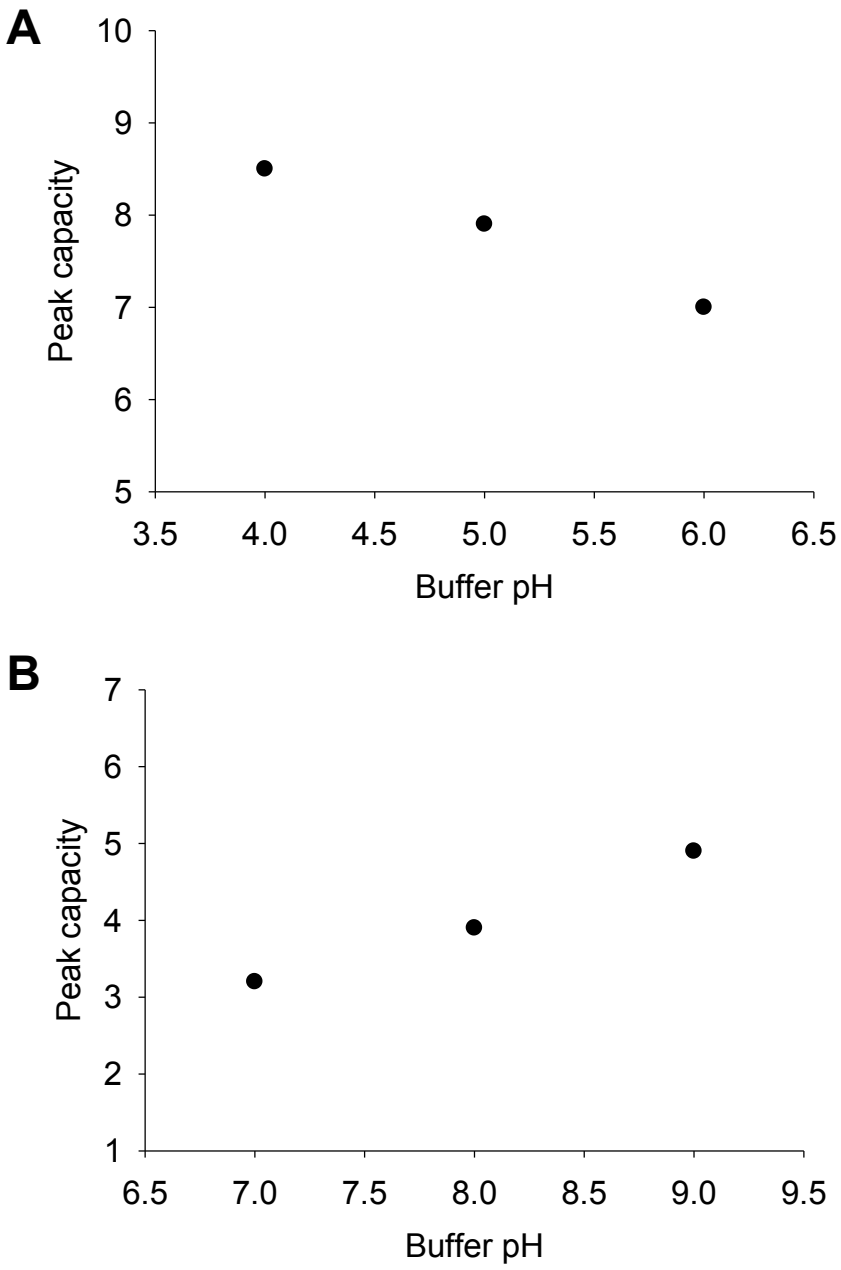

Figure 2.2 Measured peak capacities as a function of pH using the 11 protein mixture for (A) SAX and (B) SCX. At a low pH for SAX and high pH for SCX the peak capacities were at the lowest measured value. With increased pH for SAX and lowered pH for SCX the peak capacities improved. However, as some proteins were found to be retained by the columns and show reduced yield, these pH values were considered not applicable for comprehensive proteome prefractionation. Optimal pH values for SAX and SCX were chosen to be 8 and 6, respectively.

proteins which were previously retained, at pH 3, were now recovered following SCX; however, several proteins eluted in the injection peak. Similar results were obtained for SAX, with a pH range of 7 to 9 allowing for good recovery of the majority of the proteins tested. Alpha casein was a noted exception, which could not be observed in the majority of SCX and SAX separations.

From the observation of the separation of an 11-protein standard mixture, the pH range from 4 to 7 for SCX and 7 to 9 for SAX was considered reasonable for protein separation, while beyond these ranges essentially no separation is obtained. Peak capacities were thus determined over these pH values and are recorded in Figure 2.2. The measured peak capacities for SCX ranged from only 6 to 9 and for SAX from 3 to 5. Maximum peak capacities were observed at a pH of 9 for SAX and at pH 5 for SCX. However the recovery of proteins from the mixture also presents a major factor which must be considered in choosing an optimal pH value. Notable differences were seen in the recovery or retention of certain proteins, especially at the lowest and highest pH values tested. Although pH 9 provided the best measured peak capacity for SAX, some proteins, like β-lactoglobulin, were not observed in the chromatogram. For this reason pH 8 was chosen as the optimal value for the SAX separation. The optimal pH for SCX was chosen to be 6, based upon the ability to recover many of the protein standards from the column.

2.3.1.2 Optimization of Salt Gradient, Temperature, and Organic Modifier

Following optimization of the buffer pH, the elution profile of the salt gradient was varied with the goal of maintaining the highest peak capacity while keeping run time at a minimum. The change in peak capacity of the IEC separations for the 11 protein

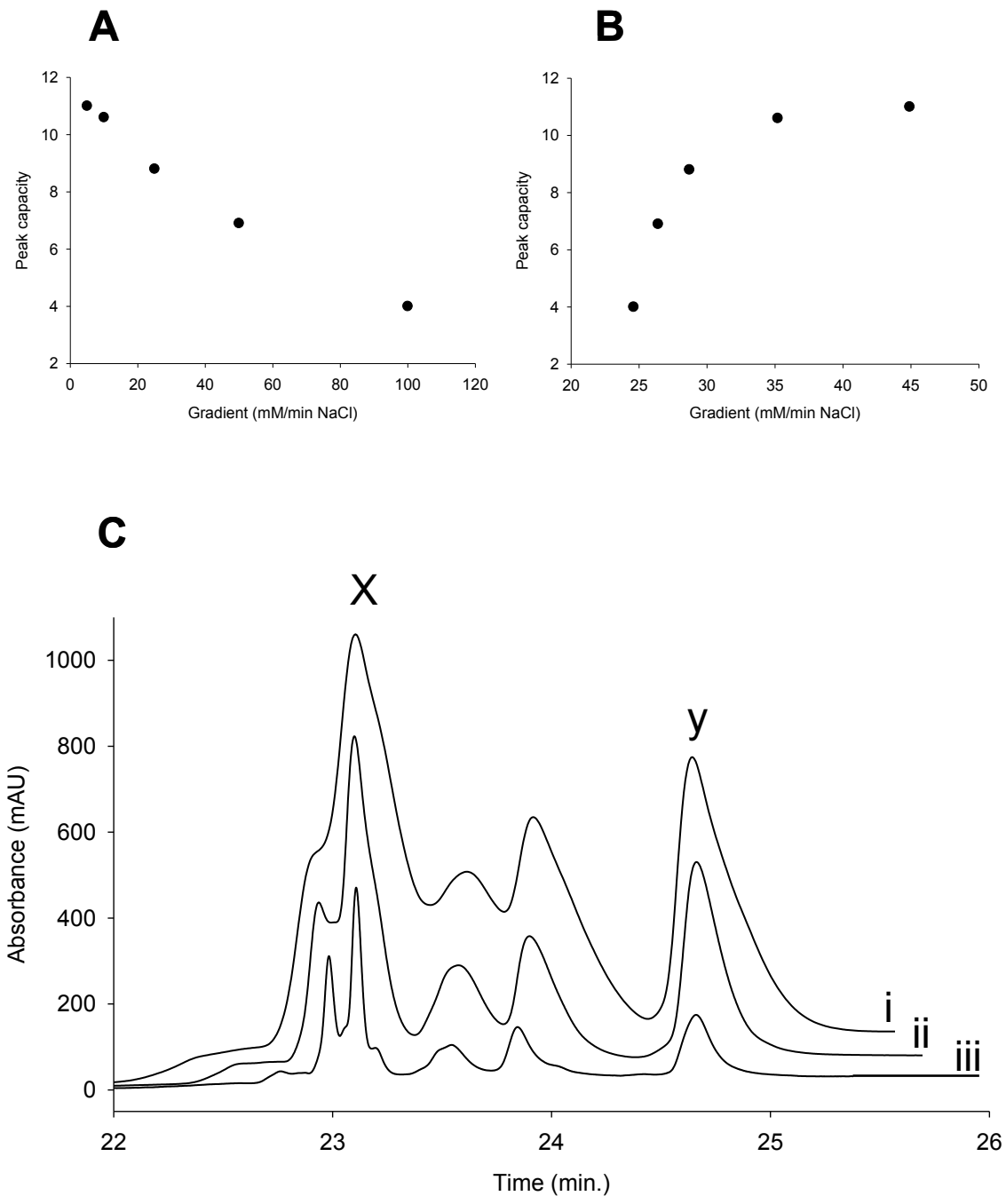

Figure 2.3 Optimization of the salt gradient slope for SCX and SAX separation using the 11 protein standard mixture. Peak capacity increased with reduced slope of the salt elution gradient, (A), but gave diminishing returns in comparison to the total separation time (B). (C) UV chromatograms obtained from the (i) 100, (ii) 50, and (iii) 10 mM/min salt gradients of the SAX separation. For ease of comparison, the 50 mM/min and 10 mM/min chromatograms were scaled along the X-axis to that of the 100 mM/min gradient by aligning the retention times of peaks x and y.

standards was evaluated over varying linear salt gradients ranging from 10 mM/min to 100 mM/min on both IEC columns. A gradient slope of a 10 mM/min for SAX and 50 mM/min for SCX was determined to provide an optimal balance of resolution and total run time. The peak capacity obtained for the SAX separation as a function of gradient slope (Figure 2.3A) and separation run time (Figure 2.3B), is shown to demonstrate the effects of gradient slope upon peak capacity. As the slope of the salt gradient was decreased, the time required for the separation increased and the peak capacity improved. However, a diminishing rate of return resulted in longer runs for only marginal improvements in separation, shown in Figure 2.3B. Chromatograms from the separation of the protein standards at the 100, 50, and 10 mM/min salt gradients using SAX are shown in Figure 2.3C. The x-axis of the 50 and 10 mM/min chromatograms were normalized to that of the 100 mM/min separation using the Agilent Chemstation software to facilitate the comparison. The improvement in resolution as the salt gradient was reduced is obvious. However, while improvements in peak resolution at shallower salt gradients were obtained, below a salt elution gradient slope of 10 mM/min using the SAX column, an increase in peak widths limited any gains in peak capacity.

Finally, both the temperature and the degree of organic modifier were explored as a possible means of improving resolution. It was found that increasing the temperature did not improve peak capacity, and at temperatures above 50^{o}C peak capacity was reduced (results not shown). Also, while it is noted that the addition of up to 25% acetonitrile is common for peptide separations using ion exchange chromatography [82], our observations showed no improvements with the addition of acetonitrile for protein prefractionation. We attribute this to the surface chemistry of the columns used in this

study, which is silica based rather than polymer based as is common in other IEC columns.

2.3.2 Evaluation of Optimized IEC for Protein Prefractionation

2.3.2.1 Detailed Evaluation of Protein Recovery from IEC

Protein recovery is a critical parameter when assessing the merits of an analytical manipulation. This parameter is more important than resolution, since the potential exists to add a second dimension of separation to improve resolution, while a low recovery of sample is not reversible. In the context of chromatography, one can expect some loss of protein; the very premise of the experiment is to allow proteins to bind to the column and then, hopefully, elute over the course of the gradient separation. An evaluation of protein recovery using individual protein standards allows for a detailed assessment of any bias associated with unbalanced recovery of some components relative to another.

The determination of percent recovery for some of the protein standards used for optimization was performed; the results of which are shown in Table 2.3. A large variation in yield between the individual proteins is noted, uncovering a recovery bias for protein components of the mixture. The SCX separations displayed a slightly greater degree of recovery variability (RSD of 68% for SAX and 77% for SCX), and both had approximately equal average recoveries (40% for SAX, 38% for SCX). Obviously, a total protein recovery would not have fully described the yield of protein from EIC, since with SCX it is noted that protein recovery for several proteins was below 25%, and α-casein was not detected following SCX or SAX separation. A low recovery of a particular protein likely indicates irreversible binding, or protein precipitation within the column. To further test the irreversible nature of protein binding, a weak cation exchange (WCX)

Table 2.3 Protein yields (%) under optimized conditions.

Protein	SCX	SAX
Insulin	24	22
Cytochrome c	73	66
Ubiquitin	57	65
α-casein phosphorylated	0	0
Carbonic Anhydrase	17	68
β-Lactoglobin	72	20
BSA	72	39
average	45	40
STD	31	27
%RSD	68	68

Table 2.4 Protein Yields from SCX and WCX.

	SCX[A]		WCX[B]	
Protein	% Recovery[C]	Loss before separation[D]	% Recovery	Loss before separation
Insulin	11	27	24	17
Cytochrome c	81	13	100	14
Ubiquitin	83	16	100	16
α-Casein	13	8	20	8
Carbonic Anhydrase	65	31	96	31
β-Lactoglobin	62	23	85	23
BSA	68	16	82	16
average yield (%)	55	19	72	18
%RSD	55	43	49	41

[A]SCX was on a polySULFOETHYL A SCX column, using conditions described in Table 1.
[B]WCX was on a polyCAT A WCX, using conditions described in Table 2.1.
[C]Data represents protein recovery relative to a protein standard prepared in aqueous 0.1% TFA
[D]Protein loss due to solubility calculated by subtracting recovery relative to a standard prepared *in 0.1% TFA vs* the recovery relative to a protein standard prepared in the initial HPLC solvent system

column was tested to examine if a different functional group had any impact on the protein yield. The results of this experiment are presented in Table 2.4. From the recovery data it is clear that using a different column type did have an effect on the protein yield, showing improved recoveries. However, a large variation in yield was still observed. It is possible that the loss of protein from IEC may be due to the initial pH of the buffering system, as described further in section 2.3.5. As also seen in Table 2.4, the buffering solution did have a serious negative effect on the protein yield for some of the protein standards, notably α-casein and insulin. Although the buffering solution could be adjusted to allow the recovery of a particular protein to improve, this would come at the cost of reduced yield for other proteins which may not behave so amiably under those conditions.

2.3.2.2 Correlation of Protein Elution and Protein Charge on IEC

One of the major incentives for performing proteome separation of whole proteins, rather than of the peptides generated following enzymatic digestion, is the ability to obtain information regarding intrinsic properties of the sample components (e.g., protein size, charge/pI, or hydrophobicity). In doing so, one can target the collection and/or analysis of specific types of proteins. Retention time information may also be used in conjunction with MS data to facilitate protein identification.

Correlation plots, shown in Figure 2.4, between protein retention time and protein pI, net charge, and total negative charge of the protein are given for the strong anion exchange separation. Protein charge was calculated by summing the charge of the basic and acidic amino acid residues at each pH tested using typical pKa values. From examination of these plots, it is clear that some correlation between the protein elution

order and the charge/pI of the protein could be made for IEC, though with a major deviation for cyctochrome c, as well as lysozyme (not shown in plot). Obvious deviations from predictable behaviour based on the net charge of the protein were observed, with counterintuitive results, such as the late retention of cytochrome c with a predicted net positive charge. However, the net charge does explain the non-retention of ubiquitin (pI 6.8), myoglobin (pI 7.2), and carbonic anhydrase (pI 6.4) (see Figure 2.1). These proteins were observed in the injection peak, but based on their pI one would predict that these proteins would at least partially retain on a positively charged support. However, the net charge of these proteins remains close to zero, which may explain their non-retention on the ion exchange column. In contrast, based on their pI, cytochrome c (pI 10.2) and lysozyme (pI 9.8) should not have been retained on the SAX column at pH 8, yet both were detected late in the salt gradient. Other attempts to find a correlation between a property of the protein and elution order, such as net negative charge (Figure 2.4C) did not reveal any useful pattern.

The difficultly in predicting the elution order of proteins on IEC is that the spatial distribution of charge on the protein, not just total charge, becomes important [163]. However, such information is rarely known about the protein under particular IEC separation conditions even if the amino acid sequence is known. This conclusion highlights a significant deficiency in ion exchange as a prefractionation tool for intact protein separation. Any useful information about the proteins becomes convoluted due to the ambiguous nature of the mechanism underplaying the separation of whole proteins in ion exchange chromatography.

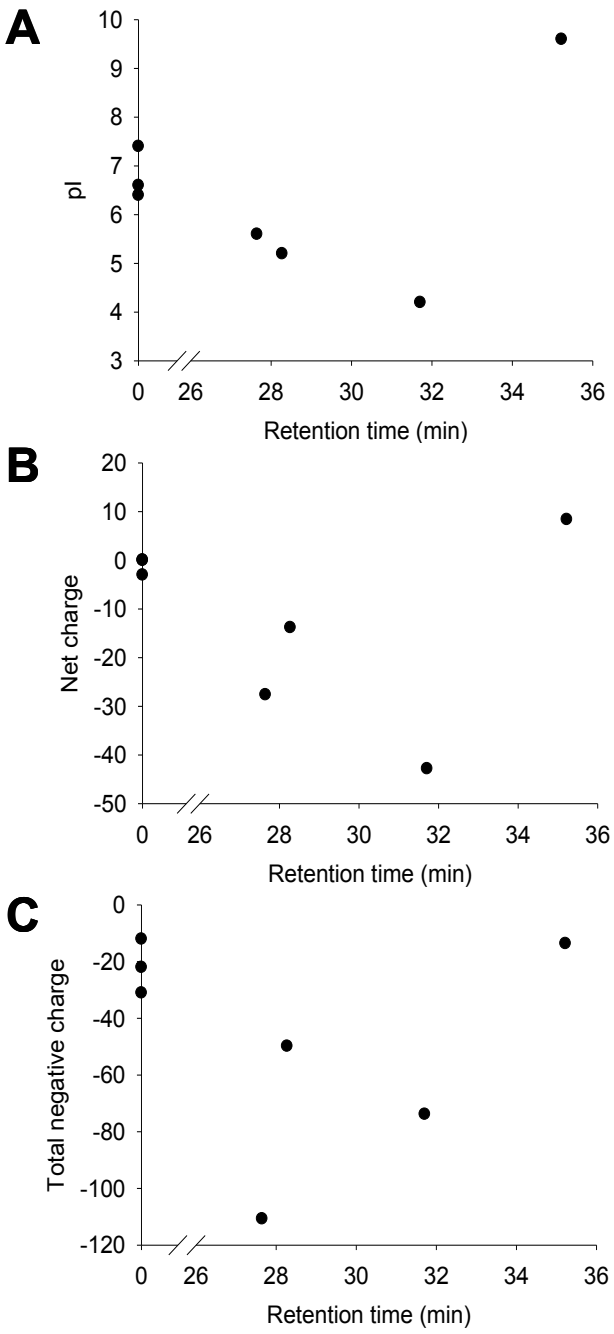

Figure 2.4 Plots showing protein retention time from SAX versus their (A) pI, (B) net charge, or (C) total negative charge. Some correlation was observed between protein retention time and pI, with the exception of cytochrome c. Plotting theoretical net charge of the protein versus retention time reveals that carbonic anhydrase, myoglobin, and ubiquitin, the three non-retained proteins, have a net charge close to zero, indicating they would bind poorly. Again, cytochrome c deviates significantly. Other properties, such as using total negative charge of the protein, did not reveal any discernable pattern to the elution order. These results indicate that the mechanism of protein/column interaction is complicated and that protein retention time cannot be used to provide any useful information about its properties (such as pI).

2.3.3 Bottom-up MS Analysis of an IEC Prefractionated Yeast Proteome

A bottom-up approach was employed to analyze the fractions collected from the SAX and SCX separations of a soluble yeast proteome extract to evaluate its effectiveness at analyzing a complex protein mixture. The SAX and SCX separation were divided into 16 fractions, each being subjected to single LC-MS/MS analyses. Figure 2.5A and 2.5B displays the number of proteins identified in each of the collected fractions from the two IEC separations, as well as the respective chromatograms of the yeast separation. Proteins were identified in each of the collected fractions, although their distribution across the fractions was varied. The chromatograms clearly show that a significant number of proteins showed limited binding to the column, being detected early in the run. This is consistent with the observations made using the protein standards during the pH optimization. From the results of the MS analysis, a Venn diagram was created to compare the proteins identified between the SAX and SCX separation, which is shown in Figure 2.5C. Overall, 173 proteins were identified from the SAX fractions, and 148 proteins were identified following prefractionation using the SCX platform, totalling 298 unique proteins from both systems. Of these, 107 were unique to SAX and only 36 were unique to SCX, with 155 proteins in common. Of the total number of proteins identified, 88% were detected from the SAX fractions. The greater number of proteins identified from the SAX separation is attributed to the reduced bias in protein recovery as demonstrated in the optimization and yield evaluation using the protein standards.

Replicate LC-MS/MS experiments were also conducted on the unfractionated yeast samples to determine the effectiveness of the proteome prefractionation strategy.

Given that 16 fractions were collected from IEC, and subsequently analyzed through 16 LC-MS/MS runs, the unfractionated sample was similarly analyzed through a total of 16 replicates. The results of this experiment are provided in Figure 2.6. Differences are noted between the SAX and SCX samples, which are likely due to variations in protein solubility under the respective buffering conditions. The increase in number of protein identification going from a single injection to multiple replicates of the same sample also illustrates the severe limitation of LC-MS/MS when presented with an extremely complicated sample. The act of performing replicate injections of the same sample provides opportunity for the mass spectrometer to sample peptides which were not selected for MS/MS in previous runs, elucidating additional protein identifications. This observation lends explanation to why the collection of a greater number of fractions from IEC than is reasonable (based on the peak capacity observed here) can result in an increase in the number of protein identifications. It is clearly shown that the separation power of IEC is rather limited and does not warrant collection of more than 10 to 20 fractions. With this realization, multidimensional separations can be simplified by limiting the number of fractions collected from the first dimension to a value more consistent with the peak capacity of the chromatographic technique, allowing the second dimension to have a greater throughput. The collected fractions can then be analyzed by replicate LC-MS runs to maximize the number of identified proteins.

2.3.4 2DLC of a BSA-spiked Yeast Mixture

From the results above, it would appear that proteome prefractionation does not achieve any appreciable benefit, beyond what can be gained through replicate MS analysis. It is noted that only a limited fraction of the entire yeast proteome was detected.

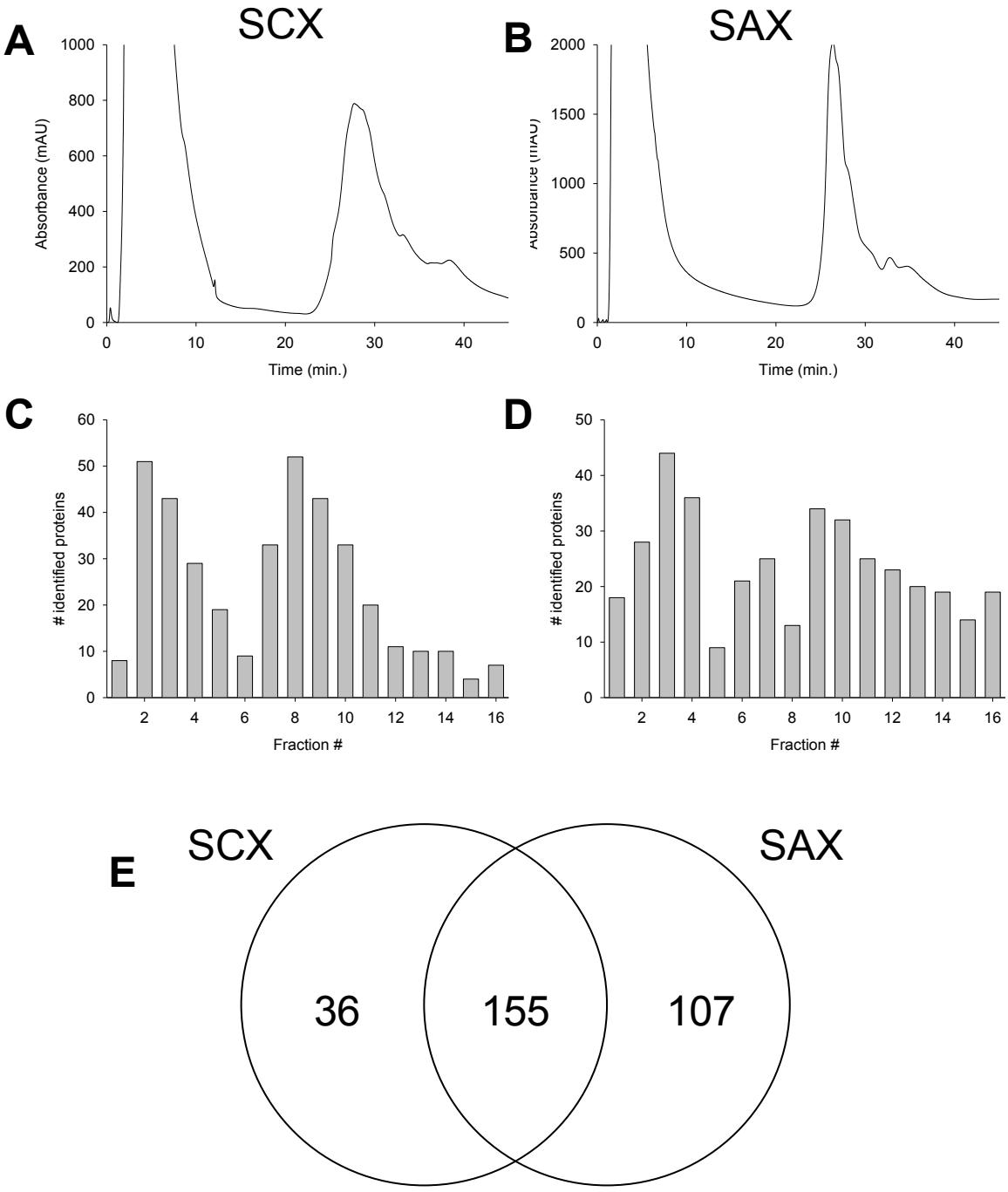

Figure 2.5 Chromatograms of the IEC separation of yeast proteins by (A) SCX and (B) SAX. The number of proteins identified in each of the 16 fractions obtained from (C) SCX and (D) SAX are shown. Venn diagram (E) comparing the total number of unique proteins identified from both SCX and SAX shows that a greater number of proteins were identified from the SAX separation, mostly attributed to its greater yields in comparison to SCX.

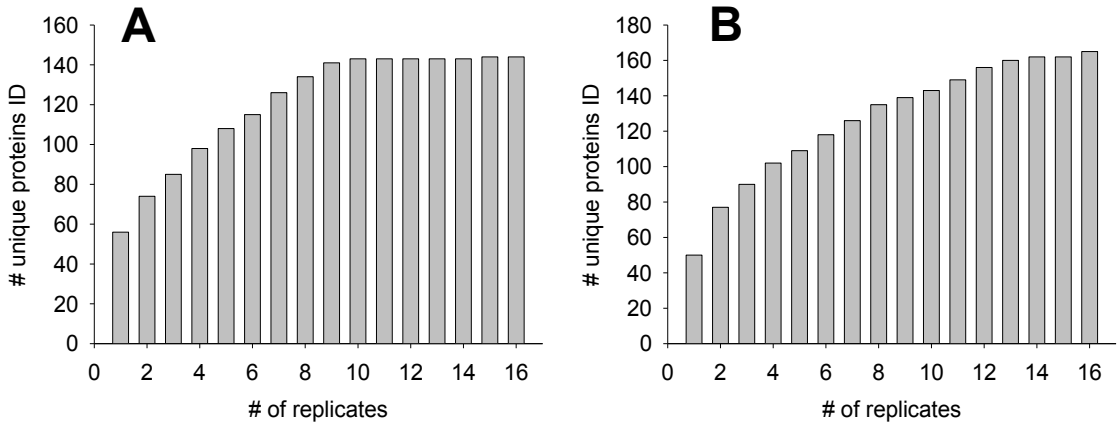

Figure 2.6 Number of total proteins identified from 16 replicate injections of the unfractionated yeast mixture in either (A) SCX buffer (20 mM NaAc pH 6) or (B) SAX buffer (20 mM Tris-HCl pH 8).

The combination of relatively low resolution with biased protein recovery results in a hampered detection of proteins following IEC separation. The greatest benefits of proteome prefractionation are realized through the depletion of high-abundance proteins. However, unlike the plasma proteome for example, yeast reveals a much more uniform proteome. To demonstrate the advantages of proteome prefractionation for depletion of high-abundance proteins, a yeast extract was spiked with bovine serum albumin to a level of 50% of total protein concentration by mass. This plasma-like proteome was then subjected to one and two-dimensional solution separation. Due to the inadequate fractionation ability of BSA on IEC, solution IEF was chosen for this experiment. The solution IEF separation yields 8 fractions, with BSA being primarily focused into a single fraction (fraction 5), shown in Figure 2.7, thus depleting the protein from other fractions. As a control, the number of proteins identified after prefractionation was compared to 8 replicate injections of the unfractionated mixture. The results from this experiment are shown in Table 2.5. Through prefractionation, the number of protein identifications was doubled, from 71 to 152 unique proteins from single injection of the 8 fractions obtained from the sIEF device.

Following sIEF, BSA focused primarily into fraction 5. As one would predict, LC-MS/MS analysis of this fraction was still able to identify other proteins other than BSA (65 from triplicate analysis of the fraction). In an attempt to improve detection of low abundance proteins in this fraction, the sample was subjected to a second dimension of separation intact proteins, in the form of RPLC, followed by tryptic digestion and analysis by LC-MS/MS. A total of 9 fractions from RPLC were collected, and as shown in Figure 2.7B, BSA was resolved into a single fraction (#8). This fraction was discarded.

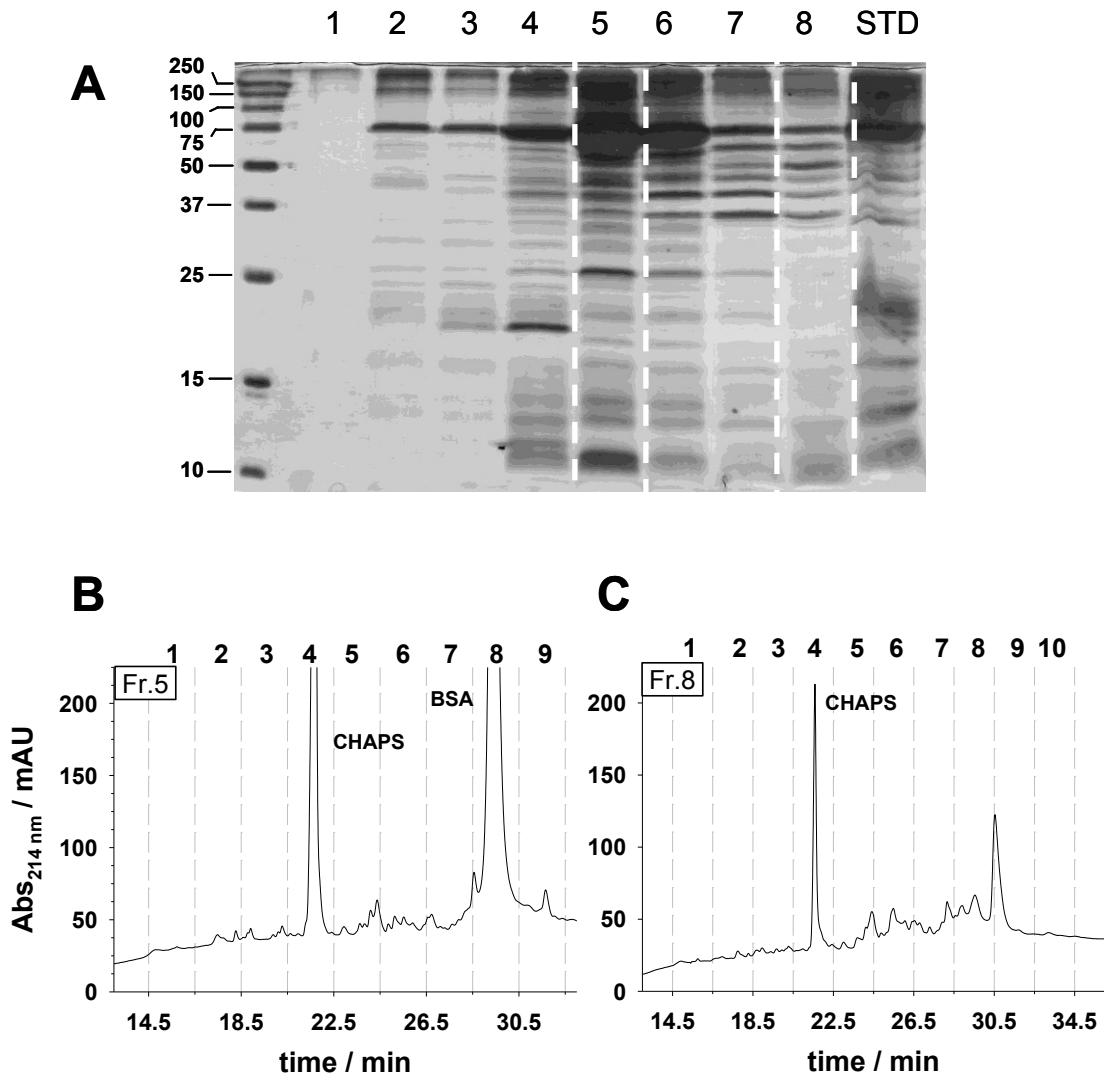

Figure 2.7 Prefractionation of a yeast proteome sample which had been spiked with BSA. Two-dimensional separation using (A) IEF followed by (B) RPLC separation of fraction 5, containing the focused BSA and (C) fraction 8 which contains no BSA.

Table 2.5 Peptides and proteins identified from 1D/2D prefractionation.

Sample	# unique peptides	# unique proteins
Unfractionated sample	374	71
IEF fractionated sample	910	152
IEF Fr5	359	65
IEF Fr5 BSA depletion via RPLC	58	14
IEF Fr5 with RPLC including BSA fraction	89	23
IEF Fr8 10 replicate runs	567	106
IEF Fr8 with 10 RPLC fractions	204	47

Again, to avoid the effects of replicate analysis, the remaining RPLC fractions were simply pooled and subjected to triplicate MS analysis. Interestingly, though BSA was further depleted in this sample, only 14 proteins were identified (significantly less than the 65 originally obtained from sIEF). A subsequent RPLC run which included the BSA-containing fraction was able to identify 23 proteins from triplicate MS analysis.

Similar results were obtained from 2D fractionation using sIEF fraction 8, a sample which contained only trace levels of BSA. From 10 RPLC fractions, each analysed individually, a total of 94 proteins were identified. However, replicate analysis (10 in total) of the fraction without subsequent separation yielded 164 unique proteins. In fact, each single run performed without a second dimension of separation detected more proteins than the 10 combined runs of RPLC fractionation (results not shown).

Remarkable run-to-run reproducibility in terms of number of identifications was obtained, (3% and 7% RSD for number of proteins and peptides respectively). From the list of proteins identified from each mixture (results not shown), it became clear that the high mass proteins were lost during reversed phase prefractionation. These results are consistent with those reported by others for packing material with relatively small pores (300 Å) [111, 158]. This problem may be alleviated by using non-porous beads. However, the RP column employed herein was marketed as a protein separation column and is of the type which is commonly used for proteomics applications.

2.4 Conclusions

From the optimization of IEC and its evaluation through protein yield and bottom-up LC-MS analysis, it is concluded that this chromatographic technique, as a protein

prefractionation platform, is not an ideal choice for proteomics workflows. In general, protein separation through IEC is extremely biased. While some proteins have high recoveries and display excellent peak shape, other proteins had broad elution profiles and/or very low recovery (e.g., phosphorylated α-casein). In light of the need to retain all proteins for a comprehensive proteome characterization, this study demonstrates that IEC would ultimately make detection of at least a portion of the proteome difficult, if not impossible.

As a more general consideration, any separation platform which does not maintain high recovery across the proteome would not serve well as a proteome fractionation system. While it appears that such concerns extend beyond IEC (e.g., loss of high mass proteins in RPLC), researchers should take care in choosing a prefractionation system. Methods must be employed which are unbiased towards particular components of the proteome to avoid loss of potentially important sample components, such as the discovery of biomarkers. In this light, it is not as critical to achieve a high total yield (averaging all components of the sample) as it is to obtain uniform protein recovery for all components. The solution IEF device, employed in this study, has previously been proven to provide a more uniform separation across all standard proteins [164], suggesting its use as an alternative to IEC as a charge-based solution phase separation platform.

As a final remark, it was also shown that replicate analysis of unfractionated mixtures can produce similar numbers of protein identifications compared to an unfractionated sample given equal LC-MS/MS analysis time. Frequently, many more fractions are collected from IEC than what was determined for peak capacity, with the

adage “the more the better”. While the collection of a greater number of fractions from a first dimension of separation would lead to more protein identifications, this would be, in large part, due to an increase in MS instrument time being applied to the sample. The collection of a greater number of fractions from a first dimension of separation would also lead to a greater bottleneck if a second dimension is applied to further fractionation of the sample.

Chapter 3

Design and Evaluation of a Multiplexed LC-MALDI MS Platform

3.1 Introduction

As discussed in the introductory chapter to this thesis, protein biomarker discovery and validation is a major driving force for proteomics research. It entails an evolving series of proteomics analysis experiments, first incorporating a comprehensive analysis (discovery phase) and moving towards a more strategic targeted analysis of putative protein biomarkers across a large population (validation phase). As an effective tool for high-throughput proteomics analysis, mass spectrometry (MS) can play a pivotal role in each stage of this pipeline. However, in the context of comprehensive analysis, such as during the initial stages of biomarker discovery, the detector limitations of MS necessitate sample prefractionation in order to overcome MS suppression effects [102]. With current technology, the comprehensive characterization of a proteome will typically incorporate liquid chromatography coupled to MS (LC-MS) for analysis of the many samples (tens to even hundreds) generated through proteome prefractionation. Each said fraction generally occupies 1 to 2 hours of LC-MS analysis time, resulting in days to possibly weeks in order to complete the analysis of a single proteome. Moreover, as described in chapter 1, the early stage of biomarker analysis, while not to the scale of the late stage, will require analysis of multiple proteomes in order to accurately profile and identify potential disease biomarkers. Considering the requirements for early stage biomarker identification, wherein the full proteome must be profiled for multiple samples, current proteomics analysis strategies do not possess adequate throughput to

justify comprehensive proteomics analysis for early-stage biomarker discovery. A high-throughput and robust method which enables screening of multiple proteome samples for potential biomarker detection is urgently required.

In terms of separation, multidimensional liquid chromatography is a favoured approach to protein prefractionation. However, considering the throughput of an experiment, multidimensional separations present a significant bottleneck. In the second dimension of separation, given that each fraction is traditionally analyzed in series, the time required to process all fractions generated from the first dimension of separation can be quite significant. This effect is compounded by the collection of an increasing number of fractions from the first dimension. For example, considering the collection of 10 fractions from a first dimension of LC (say ion exchange chromatography), this relatively small number of fractions will nonetheless represent some 10 to 20 hours for processing all samples through a second dimension of separation.

A potential solution to the problem of processing multiple fractions from a first dimension of separation is to carry out the second dimension in a parallel (multiplexed) format. For example, two-dimensional gel electrophoresis serves this very function in separating proteins across a spatial array. However, if liquid chromatographic approaches are to take advantage of a multiplexed separation design for the parallel separation of many fractions simultaneously, the coupling of numerous columns through an appropriate fluidic pathway is required.

Several designs for parallel chromatography have been reported in the literature, being capable of accommodating varying numbers of columns. The simplest designs focus on increasing the duty cycle of the LC-MS experiment. A staggered LC system

with two or more columns allows one column to be primed (solvent equilibration and sample loading) while performing chromatographic separation and MS analysis with the other column [165-167]. While up to a two-fold gain in improvement is afforded by these systems, they cannot be defined as true parallel chromatographic platforms as the chromatographic separations are still performed in a serial fashion.

True parallel separation platforms have been developed which permit the simultaneous delivery of a solvent gradient elution of proteins or peptides from multiple LC columns [168-170]. However, in all cases a significant concern lies with the system designs. The dedication of a number of switching valves, multiple HPLC and complex fluidic collections make for the automated operation of such systems difficult. More importantly, with these designs, it is difficult to ensure even flow rates between multiple columns, questioning column-to-column reproducibility. Also, the unintended sample cross-loading across multiple columns can become a concern with these systems. An improved system design permitting parallel LC column separation for a multidimensional platform is one of the subjects of the present chapter.

Following multidimensional fractionation, yet another bottleneck exists in the proteomics analysis workflow. While considerable gains in sample throughput can be met by performing separations in parallel, all said gains would be lost when considering the sequential analysis of fractions resulting from each separation with mass spectrometry. Baring the use of multiple mass spectrometers (a motivation to miniature MS systems, which is beyond the scope of this thesis), attempts to couple parallel chromatographic and electrophoretic separations to a single electrospray ionization interface have been made [171]. Perhaps the most well-known example of coupling multiple LC flow streams to

MS is through the multiplexed interface (MUX-technology^{TM}) available commercially through Waters/MicroMass. This interface allows multiple LC columns to be connected to a single ESI source, as shown in Figure 3.1, and has been applied to the analysis of small molecule combinatorial libraries [172, 173]. However, such systems only permit the eluent from a single column to be analyzed at any given moment. Given the complexity of proteome separations, combined with the short time scale in which peptides elute from a reversed phase column, such a platform is not suitable to proteomics analysis.

Alternative solutions to maintain throughput in coupling parallel separations to MS is to do away with electrospray ionization, switching instead to matrix-assisted laser desorption/ionization mass spectrometry. The advantage of MALDI over a parallel LC-ESI design is that the eluent from a parallel separation can be collected and stored on a MALDI target, allowing each fraction to be analyzed in an offline format. Thus, one does not sacrifice any information, as experienced when "sampling" multiple LC columns with ESI interfaces. Additionally, in terms of flexibility, given that the MALDI source is decoupled from the detection stage, an opportunity exists for customizing the analysis. For example, one may choose to perform either rapid or extended MS profiling of any given fraction. One concern with MALDI is the low sample volume which can be accommodated on the target plate, typically 1 to 2 microlitres per spot. Thus, considering the volume of fractions collected from analytical columns (on the order of hundreds of microlitres per fraction), a typical solution is to spot only a small portion of the collected sample. This degree of sample loss can result in poor sensitivity, thus the alternative method is to concentrate the sample through solvent evaporation followed by re-

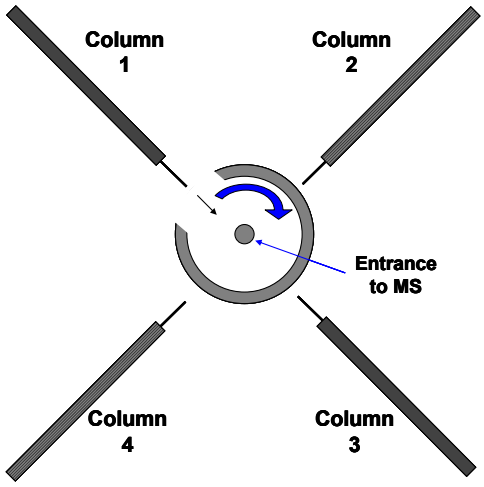

Figure 3.1 Schematic representation of the MUX-technology. Through it, 4 LC columns can be interfaced *via* ESI to a single mass spectrometer. However, the effluent from only one column can be analyzed at any given moment.

suspension and sample spotting. However, this approach is both labour intensive and carries an associated risk of sample loss.

The development of devices which allow direct interfacing of liquid chromatography to MALDI would significantly improve throughput and sensitivity of this approach. A host of devices have been described in the literature, primarily addressing the interface of reduced flow rate (low to sub microlitre per minute) LC or of capillary electrophoresis (CE) to MALDI [118, 174-182]. Some of these designs incorporate on-target digestion of proteins to further improve throughput, but all are generally limited in their sensitivity and sample loading capacity. Online LC-MALDI-MS interfaces have been developed which allow the direct analysis of LC eluent by MALDI MS similarly to ESI MS, but these systems require complex modification to existing MALDI sources which may not be easily accomplished by most proteomics facilities [183, 184, 184-188]. LC-MALDI interfaces which permit collection of larger quantities of material from higher flow rate separations have been described, such as through the use of heated interfaces [68, 189]. However, the removal of solvent immediately upon or during deposition onto the MALDI target provides an upper limit to solvent flow rate and also does not easily permit subsequent sample manipulation. The most promising appear to be on-target well plate devices which allow for the collection of higher volumes from LC separations as well as further down stream manipulations.

The use of detachable well plates which interface directly with the MALDI target allows a greater volume of sample to be collected and permits sample manipulation on collected fractions [190, 191]. In this instance, the MALDI target plate acts as the bottom of each individual fraction collection well. Following collection, target plates can be

placed into a Speedvac which removes the solvent and allows for the entire sample to be directly deposited onto the target spot. However, existing designs employing this strategy incorporates a protein/peptide capture phase material. This is seen through the use of polydimethylsiloxane (PDMS) material to construct the collection plate [190, 191], or by incorporating chromatographic resin into the collection wells [69]. The desired strategy is that proteins and peptides are retained on the PDMS/ chromatographic material, permitting additional washing/preconcentration steps to be performed on target. Such steps are not necessary in the present context, and also require additional steps to ensure non-retention of sample. The risk of sample loss is also high with such devices.

In this chapter, the design and evaluation of an integrated parallel chromatography to high volume LC-MALDI collection is described. This platform forms the basis of a high-throughput comprehensive proteome strategy based on MALDI MS, the development and application of which is described in chapter 4. The parallel chromatographic system consists of eight reversed phase columns and permits individual loading of samples onto a select column, thus permitting the online capture of up to eight fractions from a first dimension of protein or peptide separation. Evaluation of this system consists of determining sample cross-loading during injection, as well as flow rate equality and reproducibility between the columns of the chromatographic separation. The MALDI interfacing well plate accommodates up to 384 spots on a single target using an on-target well plate constructed of Teflon. The spot-to-spot cross contamination and effectiveness of concentrating large volume samples onto a MALDI spot is determined.

3.2 Experimental

3.2.1 Materials and Reagents

All standard proteins used throughout the experiments, including bovine trypsin (TPCK treated to reduce chymotryptic activity), were purchased from Sigma (Oakville, ON, Canada). Formic acid and trifluoroacetic acid (TFA) were also purchased from Sigma-Aldrich. Milli-Q grade water was purified to 18.2 MΩ cm^{-1}. The MALDI matrix, α-cyano-4-hydroxycinnamic acid (CHCA), used was obtained from ThermoFisher ProteoMass vMALDI Calibration Kit (MSCAL4). Solvents were of HPLC grade and were from Fisher Scientific (Ottawa, ON, Canada).

3.2.2 Preparation of Protein Standards and Matrix Solution

Protein standards were dissolved to a concentration of 2 mg/mL in water and kept at -25^{o}C before use. Protein stocks were diluted to 0.1 mg/mL in solvent A (see section 3.2.3) prior to reversed phase chromatography. A stock solution of MALDI matrix was prepared by dissolving 5 mg of CHCA into 0.5 mL of 84% ACN, 14% EtOH, and 2% water with 0.05% TFA. This stock solution was diluted using the same solvent composition to a concentration of 2 mg/mL prior to use.

3.2.3 Liquid Chromatography

Chromatographic separations were performed using a quaternary Agilent 1100 or 1200 HPLC system with degasser, autosampler, and diode array detector set to record absorbance values at 214 nm (Palo Alto, CA.). Reversed phase columns (C_4, 1 mm × 100 mm, 5 μm beads, 300 Å pore size) used in the parallel system were from Phenomenex (Torrance, CA). Separations were performed using a linear gradient from 95% solvent A (water, 0.1% TFA) to 40% B (acetonitrile, 0.1% TFA) over 80 minutes for peptide

separations, and 90% A to 70% B over 30 minutes for protein separations. Reversed phase cleanups of protein digests were performed using the same solvent and gradient conditions as described for the peptide chromatography except it was performed over a 5 minute time interval, with collection of eluent followed by removal of solvent using a SpeedVac.

3.2.4 Tryptic Digestion

A 50 µL sample of a 0.1 g/L BSA solution was reduced and alkylated with the addition of 2.5 µL of 200 mM dithiothreitol, being incubated for 20 minutes at 55^{o}C, followed by 5.5 µL of 200 mM iodoacetamide and incubation for 20 minutes in the dark. The digestion was started with the addition of 0.1 µg of trypsin (50:1 ratio protein to trypsin) and left overnight at 37^{o}C. The digestion was terminated with the addition of 5 µL of 10% TFA followed by reversed-phase cleanup to remove contaminates prior to use.

3.2.5 MALDI MS Analysis

MALDI-MS was performed using a vMALDI LTQ system (ThermoFisher, San Jose, CA). Reversed phase fractions collected from the parallel system onto the on-target well plate system were dried to completion in a SpeedVac. Next, the 2 mg/mL CHCA solution was combined with water in a 1.5:1 ratio and 2.5 µL was deposited onto the MALDI spot to resolubilize the dried sample which was then allowed to dry and co-crystallize on the MALDI target. MS data was collected using optimized laser power with AGC (automatic gain control) activated, and a MS signal threshold set to 5000 S/N, using 2 laser shots per spectrum, and averaging 3 to 5 MS scans per spectrum. Ion accumulation was set to 5×10^4 for MS scans over the mass range 800 – 4000 *m/z*. Data

were examined using the Thermo Xcalibur qualitative browser.

3.2.6 Design and Evaluation of the Parallel Chromatography Platform

A schematic of the parallel LC device is shown in Figure 3.2. The platform consists of two switching flow valves (V1 and V2) which control the direction of solvent flow from a single HPLC pump The first valve (V1), designated as the loading/running valve, is a 10 port, 2 position switching valve (Rheodyne, Oak Harbour, WA) which determines if flow from the HPLC pump is delivered to either V2 for column selection (loading mode, Figure 3.2A) or to a splitter (i) which directs flow to all reversed phase columns (ii) simultaneously (running mode, Figure 3.2B). V2 consists of an 11 port, 10 position switching valve, which allows for the selection of a particular column for loading while in loading mode. A restriction capillary (iii), consisting of a 20 μm ID, 10 cm long piece of silica capillary tubing, prevents a significant amount of flow to cross to the remaining columns during loading of a particular column.

The parallel system was evaluated to determine the equality in flow rate between columns in run mode as well as the effectiveness of the restriction capillaries to limit cross-loading. Flow rates from the system were measured at solvent compositions of 10% and 50% acetonitrile in water by weighing the collected eluent following 5 minutes at a pump flow rate of 200 μL/min (anticipating 25 μL/min per column). Sample cross-loading in the system was examined by examining the distribution of 2 and 10 μg of lysozyme following injection onto one of the RP columns. The reproducibility of separation between columns was examined by systematically injecting 1 μg of lysozyme onto each column of the parallel system and performing a gradient elution with that column connected to a UV detector. This was compared to the separation on each column

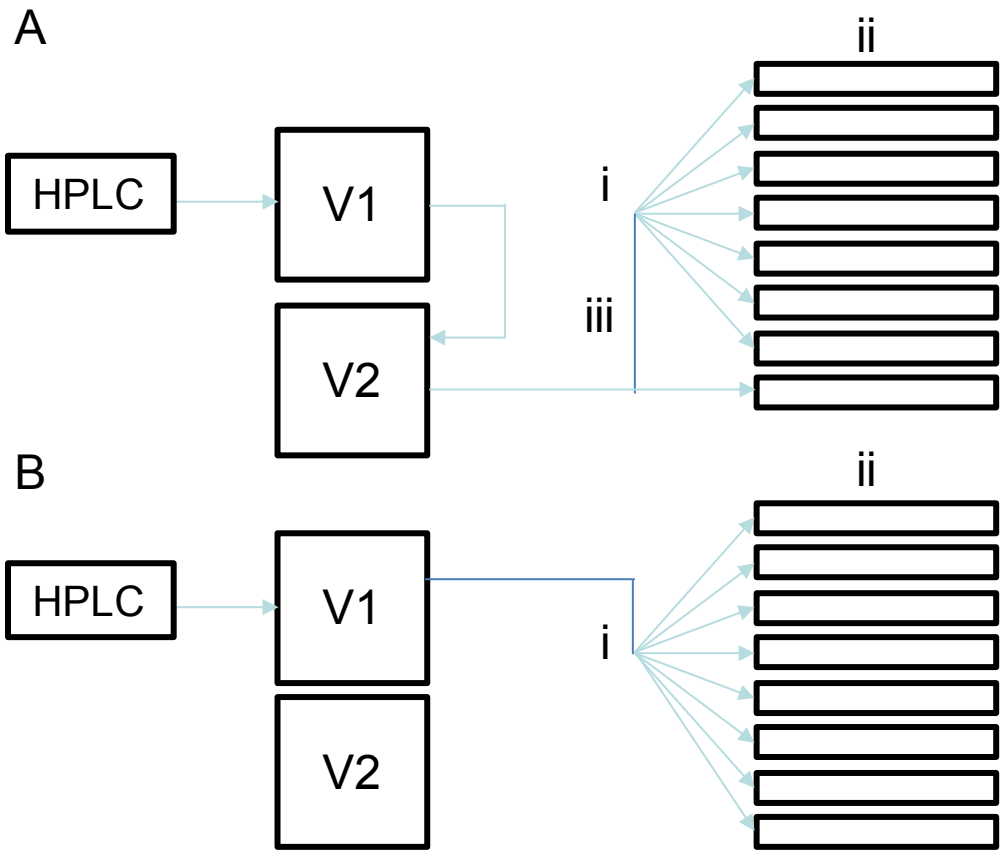

Figure 3.2 Diagram of the parallel chromatographic system showing load position (A) and run position (B). In load position, V2 allows for the majority of sample from the HPLC to be directed towards a particular column, in this case the last column shown in the figure above. A restrictor capillary (iii) prevents significant amounts of flow to the remaining columns (ii). In run position, V1 diverts all flow into a splitter (i) which evenly partitions flow to all columns connected to the system, allowing for multiplexed gradient elution.

independent of the parallel system.

3.2.7 Construction and Evaluation of the High-volume MALDI Well Plate

A schematic of the well plate assembled and disassembled is presented in Figure 3.3A and Figure 3.3B respectively. The on-target MALDI well plate is constructed of a Teflon block (i) of dimensions 1.1 × 8.2 × 12.3 cm with 1/8" diameter holes drilled through the block, a 1 mm thick PDMS layer 1 mm thick (ii), both corresponding in number and location to the spots present on a 384 well Thermo MALDI target (iii). The device is secured using four metal bars which cover columns 1, 8, 17, and 24 using four screws with washer and nut passing through rows A, F, K and P. As a result, 320 wells are available for collection of samples. Each of these wells has a dimension sufficient to allow collection of up to 100 μL of solution.

The MALDI well plate was examined to ensure no leakage was occurring between wells during collection or drying, by placing 15 pmol of BSA digest in 50 μL of 0.1% TFA water into a randomly chosen 3×3 array of the MALDI target, leaving a blank (50 μL of 0.1% TFA in water) in the central well. These nine sample on the plate were dried using a SpeedVac, after which the well plate was removed and matrix applied to the spots as described in section 3.2.5. Peptide yield following collection and drying of sample using the MALDI well plate was examined by placing 5 replicates of 1 pmol BSA digest prepared in 50 μL of 0.1% TFA water in the well plate and subjecting the sample to SpeedVac drying. The signal intensity of the most intense peptide was compared to that observed from the direct deposition of 1 pmol BSA digest (in 1 μL) onto the MALDI target.

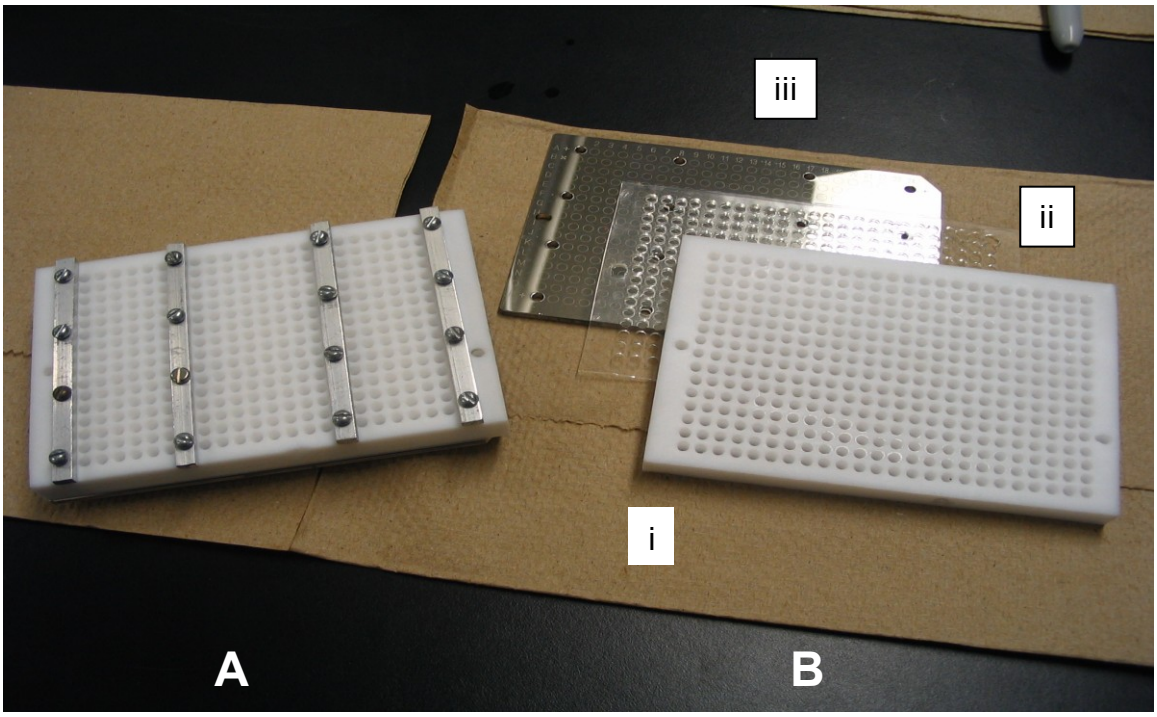

Figure 3.3 Picture of the MALDI on-target well plate device, assembled in A and disassembled in B. The device is constructed of a (i) Teflon block with 384 wells drilled through with a (ii) 1 mm layer of PSMS with holes corresponding to the size and location of the wells on the Teflon block, both of which correspond to the spots on a (iii) Thermo 384 MALDI target. The system is held together using 4 metal rods with screws which pass through the well block, PDMS layer, and MALDI target, using washers and nuts to tighten.

3.3 Results and Discussion

3.3.1 Evaluation of the Parallel Liquid Chromatography System

The parallel system incorporates a unique restriction capillary design, shown in Figure 3.2. The restriction capillaries serve two purposes. First, they limit the amount of cross-loading to other columns during loading. Second, it improves the equality in flow rates between the columns during multiplexed separations. These two features are demonstrated in the following section. The flow rate through the restriction capillary required to produce the same back pressure as the flow through the RPLC column during loading is much smaller, resulting in the majority of the flow from the HPLC to be directed onto the column. During parallel operation, the flow directed to the 8 RP columns is now forced to pass through the restriction capillaries. The combined back pressure of flow through the column, about 20 bar, plus the restriction capillary, about 120 bar, is much greater than through the column alone. This combined pressure results in reduced variability in pressure, and thus flow rate, between each column resulting in a more even partitioning of solvent across all columns. An evaluation of the parallel chromatographic system was performed to determine the equality of flow rate between the 8 chromatographic columns, the extent of sample cross-loading during injection, as well as the reproducibility of separations between columns.

3.3.1.1 Comparison of Flow Rates

The results of the determination of flow rate equality through the columns of the parallel system in run mode were determined for all 8 columns are presented in Table 3.1. From the normalized values, the variation in flow rate between the columns in a parallel mode of operation is seen to be extremely small with both solvent conditions. A relative

Table 3.1 Comparison of flow rates.

	Normalized flow rate	
Column	10% ACN	50% ACN
1	0.988	0.995
2	0.995	0.987
3	0.999	1.000
4	1.000	0.970
5	0.998	0.966
6	0.987	0.970
7	0.982	0.979
8	0.986	0.983
average	0.992	0.98
STD	0.007	0.01
%RSD	0.7	1.2

standard deviation in flow rate near 1% was seen with each solvent condition, though a slightly greater deviation in flow rate at 50% acetonitrile is attributed to a decrease in back pressure compared to that of the more aqueous solvent systems. This reduced back pressure lowers the effectiveness of the restriction capillaries. However, this variation is small and is not expected to cause significant variations in separation reproducibility between columns.

3.3.1.2 Evaluation of Separation Reproducibility and Conformity

To demonstrate the reproducibility and conformity in separation across the 8 reversed phase columns, the retention time of an arbitrarily chosen protein, in this instance lysozyme, using each column was measured while running in parallel mode, with results provided in Table 3.2. These results from parallel operation were compared to that obtained from each column when ran independent of the parallel system *(*i.e., "single column", in Table 3.2). Triplicate runs using the same column, column 1, in parallel mode and in direct format were also performed to assess variability of the retention time to that from the parallel system and between columns. The best reproducibility was seen through repeat direct injection using a single column. About a 0.5 second standard deviation in retention time was obtained from triplicate runs on column 1. Variation between the 8 reversed phase columns ran individually gave a standard deviation in retention time of about 3.5 seconds. This increase in standard deviation in retention time compared to that obtained from the use of the same column is to be expected given variations in column packing. Comparing the standard deviation obtained from running the columns individually in series to that obtained when performing the separations in parallel, chromatographic variability resulted in a deviation

Table 3.2 Comparison of reproducibility in retention time.

	Conventional LC		Parallel LC	
	Single column	Multiple columns	Single column	Multiple columns
Avg. t_r (min)	17.137	17.06	17.8	18.0
STD (min)	0.0082	0.064	0.11	0.12
STD (s)	0.49	3.8	6.6	7.5
%RSD	0.048	0.37	0.62	0.69

of about 7 seconds for column 1 run in triplicate as well as for all 8 columns However, as this parallel system is intended to couple to MALDI MS for data analysis, where the collection time of fractions may be upwards to several minutes, deviations in retention time of a few seconds are considered negligible.

3.3.1.3 Cross-loading During Injection onto the Parallel System

One concern with the design of this system, and other multiplexed chromatographic systems, is that a portion of the sample being injected can unintentionally make its way to other columns; in this case, through way of the flow splitter, shown in Figure 3.2. Prior to implementing the restriction capillaries, significant cross-loading was observed in the current system (results not shown). The results from the cross-loading experiments using the current design which incorporates restriction capillaries are provided in Table 3.3. As seen in this table, while an "open" flow path exists to all 8 columns, the necessity to flow through a restrictor channel to arrive at all columns, save the target column (column 4), favours loading onto this column only. Column 4 received approximately 97% of the injected sample, implying 97% of the flow is directed at this column. The remaining ~3% of the flow was distributed across the remaining 7 RPLC columns, though no column received more than 0.7% of the injected sample. There was a slight bias for sample to cross-load onto columns which were immediately adjacent to the intended column, columns 1, 2, and 3, based on their configuration in the 8 port splitter. Nonetheless, a majority of the sample is directed to the designated column in the parallel design. The loss of ~3% of the total sample to other columns represents a negligible loss in terms of comprehensive proteome analysis. Also, while the cross-loading of <1% of a given fraction onto a column may represent a source

Table 3.3 Percentage of lysozyme cross-loading.

Column	% protein on column
1	0.7
2	0.7
3	0.6
4	96.6 ± 0.6
5	0.4
6	0.3
7	0.3
8	0.4

of sample contamination, it is noted that the intended use of the parallel LC system is as a second dimension of separation in a multidimensional platform. Given the resolution of the first dimension, wherein a given analyte can easily distribute across multiple adjacent fractions, the contribution to sample carryover from the parallel LC platform can again be considered negligible.

These results, in combination with the flow rate determination experiments, confirm the dual purpose of the restriction capillary design. Not only does it maintain even flow rates among the columns, providing reproducible and similar separations across all columns, but it also limits the amount of cross-loading that is caused during an injection.

3.3.2 Testing of the MALDI Well Plate Target

The on-target MALDI well-plate was tested according to its ability to effectively contain sample in the collected well without leaking into adjacent chambers as well as the recovery of peptide following solvent drying. This is compared to conventional deposition approaches involving manual manipulation of samples.

3.3.2.1 Testing PDMS Seal

Teflon was chosen for the MALDI well plate to limit the loss of proteins and peptides during collection and drying of solvent. Since Teflon alone did not provide an adequate seal when pressed to the steel MALDI plate, it was necessary to include a thin gasket, composed of PDMS, to act as a seal between the Teflon and steel MALDI plate, and prevent leaks between adjacent wells. The results of this experiment are shown in Figure 3.4. The absence of BSA peptides detected in the central well which contained only solvent, while the surrounding adjacent wells contained 15 pmol of BSA peptides

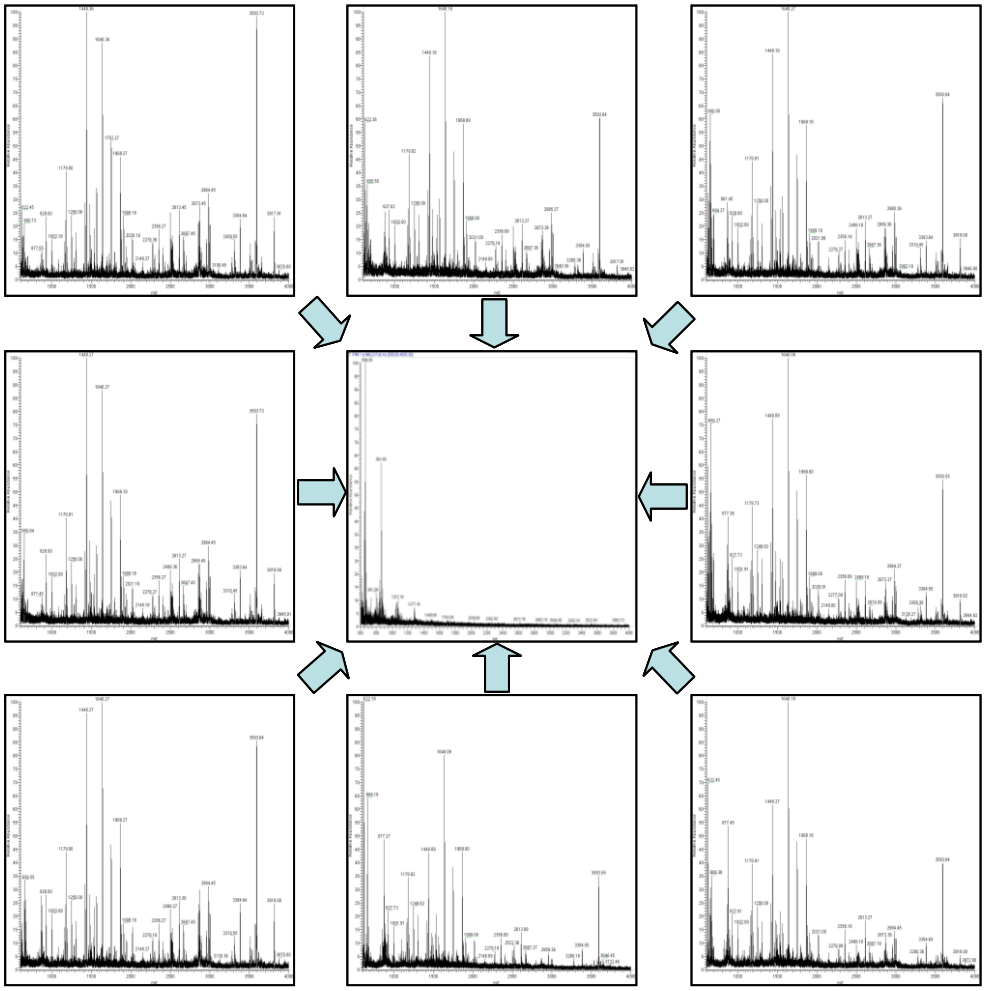

Figure 3.4 Testing of the PDMS layer to prevent leaks between adjacent wells. 15 pmol of digested BSA in 50 µL of solvent was placed into a 3x3 array of wells, leaving the central well empty (blank solvent added). Following drying, matrix addition, and MALDI MS analysis, no BSA peptides were detected in the central well (shown in center block of figure), demonstrating that the PDMS layer provides adequate sealing between the Teflon block and the MALDI plate.

each, demonstrates the lack of cross contamination and indicates that the PDMS layer and bracing system are effective methods to prevent sample from leaking into adjacent wells. It is noted that the sensitivity of the MALDI MS towards BSA peptides is sufficient to detect 5 fmol on target for this instrument (results not shown), which would represent a 0.5% carryover from any of the adjacent wells. Though an adequate seal was provided, the use of PDMS may have a negative impact upon yield, resulting from possible peptide adsorption onto its surface, as proteins are known to absorb onto PDMS. The impact upon signal intensity of a BSA digest using the well plate system compared to conventional direct deposition was investigated.

3.3.2.2 Comparison of Signal Intensities

Possible loss of sample associated with the PDMS gasket was also assessed by observing the signal intensities of BSA peptides obtained following use of the well plate device. Conventionally, higher volumes from LC separations than could be deposited directly onto the MALDI target would normally require either the spotting of only a small portion of the sample or first drying the sample to concentrate it before deposition. Using the well plate target, the drying down of sample can be performed directly on the MALDI target without the need for preconcentration through solvent evaporation or sacrificing a large percentage of the original fraction. A comparison of signal intensity was made by depositing 1 pmol of BSA peptides onto the MALDI target through three different deposition strategies: (1) transferring 50 μL of a 0.02 pmol/μL solution to the well plate device, (2) drying 50 μL of a 0.02 pmol/μL solutione in a vial and resolubilizing into 2.5 μL of matrix solution followed by direct target deposition, (3) direct deposition of 2.5 μL containing 1 pmol of BSA peptides from a concentrated stock without any prior drying

Table 3.4 Comparison of signal intensities.

	Signal intensity ($x10^4$)		
	Well plate	Vial	Direct deposition
Average (n = 6)	2.3 ± 0.5	2.2 ± 0.4	3.0 ± 0.8

steps. The results of this experiment are given in Table 3.4. As expected, the direct deposit method provided the highest average signal intensity from the three deposition methods. However, samples are rarely available as concentrated stock solutions. Comparing the average signal intensities of the samples deposited using the well plate device to the conventional (in vial) drying/ spotting strategy, similar peptide yields were obtained. From the signal intensities, the estimated yield of peptides onto the spot was about 77% for the well plate and 73% for the typical drying and solubilizing in a vial strategy. Overall, these results demonstrate that the well plate does not introduce any additional loss of sample than normally obtained during solvent evaporation steps. However, the well plate does provide a practical improvement in terms of sample preparation and throughput, as it eliminates the need to dry, solubilize, and individually deposit all fractions collected from a LC separation.

The need for high-throughput separations and MS analysis is important for characterizing the proteomes of large populations of individuals over time in order to identify potential disease biomarkers. The system presented in this chapter is designed to facilitate the fractionation of complex protein samples through the use of a parallel chromatographic platform coupled to MALDI MS through a high-volume on-target well plate attachment. This system permits reduced sample preparation time associated with direct solvent removal and sample spotting, which is normally required for LC-MALDI analysis. As a separation tool, the parallel LC system shows remarkably similar separations across all columns due to the conformity in flow rates generated through the use of restriction capillaries. The restrictor channel reduces the effects of variation in column back pressures due to packing differences between the columns. In comparison to

other parallel chromatographic designs [168, 169], this systems is much easier to automate through the use of programmable contact closures on a single HPLC pump. It also handles a greater number of columns while maintaining equal flow rates. Other designs rely on the equality of back pressures native to the columns employed to ensure even flow rates. However, this case is not guaranteed, especially over repeated column use where one or more columns may experience a clog.

The restriction capillaries used in this system also serve a second purpose by limiting the amount of cross-loading which occurs during injection of sample onto a particular column. Cross-loading is an inherent deficiency in an open column design (i.e., one which does not include valves to open/close the flow path to the various columns). However, the inclusion of restriction capillaries significantly reduced the cross-loading of sample onto the other columns (<1% per column). Considering that the intended use of this parallel LC platform is to enhance throughput in separation from a previous dimension of separation, this degree of cross loading is unnoticeable in relation to the imperfect prefractionation from the first dimension (see chapter 2).

The detachable well plate for on-target MALDI collection and high-throughput preparation has been shown to be effective and not prone to loss of sample during collection or drying or contamination into adjacent wells. The use of Teflon as the collection plate is an advantage over constructing the system completely from PDMS [191], which would necessitate additional workups and decrease the throughput of the methodology. Although it is necessary to have a thin layer of PDMS as a sealant between adjacent wells, it did not result in any noticeable loss of sample.

The use of a clamping system which covers portions of the plate wells was

necessary to provide adequate pressure for sealing. As each of the four bars covered one column of 16 spots, each plate looses 64 spots, leaving a total of 320 spots available for sample collection. Given that the parallel system uses 8 columns, this provides a total of 40 fractions per plate, per column. A total of two such on-target well-plate devices were constructed, enabling proper balance of the plates during solvent removal in the SpeedVac. With 8 columns, these two plates provide a total of 80 possible fractions per column. A typical reversed-phase separation of peptides for LC-ESI MS/MS analysis is about 2 hours in length, translating into the collection of 1.5 minute fractions. Thus, the present platform can meet the demands of a proteomics analysis experiment (described further in chapter 4). However, this number of fractions could be expanded or reduced as warranted, with limitations only restricted to the number of plates available.

3.4 Conclusions

This chapter presented the development and evaluation of a novel parallel LC-MALDI MS platform for the high-throughput analysis of proteome samples. The evaluation of the system shows it is effective at reproducible and consistent separations and reliable sample collection and high-throughput preparation for MALDI MS. The use of such a system will no doubt enhance the ability to analyze complex proteome samples faster than previously possible. However, the use of MALDI for large scale peptide identification has apparently lessened over the past few years, as identifications from LC-ESI MS/MS experiments provide bountiful and information-rich fragmentation data from CID of multiply charged peptides. Nonetheless MALDI MS may continue to play an important role in the search for protein biomarkers, provided that it is used appropriately.

Chapter 4 of this thesis focuses on the evaluation and comparison of MALDI MS to ESI MS for the analysis of complex peptide mixtures. From these results, an application of the LC-MALDI platform for high-throughput quantitative proteomics profiling is provided.

Chapter 4

LC-MALDI MS for High-throughput Proteome Profiling

4.1 Introduction

The boundaries of proteome characterization have continuously been expanded; through current proteomics technologies, the comprehensive analysis of a relatively simple proteome is now possible. Taking the yeast proteome as an example, the number of identifiable proteins in a single MS experiment increased from nearly 1,500 in 2001 (through a 2DLC platform) [83] to 4,399 quantifiable gene products (with a combination of 1D SDS-PAGE, peptide isoelectric focusing and bottom-up LC-MS) [56], a number which represents nearly the entire yeast proteome. Efforts are also being made to fully characterize more complex eukaryotic cells, perhaps the most successful example being the identification of 9,124 unique proteins from the *Drosophila melanogaster* proteome [55]. This number represents 63% of all predicted gene products. These impressive gains in protein detection and identification have partly come through improvements in MS technology, including higher mass accuracy, scan speed and sensitivity. A large part of comprehensive proteomics analysis is also owed to improved front-end sample preparation techniques. Most notably, incorporating proteome prefractionation improves the dynamic range of detection following bottom-up peptide analysis [192]. However, this increase in separation comes at the expense of increased analysis time. For example, Brunner's in depth characterization of the *Drosophila* proteome involved 1,700 LC MS/MS runs. Assuming zero downtime, this one experiment would easily consume some 20 weeks of continuous MS instrument time! Clearly, to enable the comprehensive

proteome screening of multiple samples, as is necessary for protein biomarker identification, a more high-throughput method of analysis must be employed.

Protein quantitation is essential to candidate protein biomarker identification. Quantitative workflows which employ two-dimensional gel electrophoresis, and variants such as differential gel electrophoresis (DIGE) [99, 193, 194], are capable of visualizing distinct proteins who's expression level has varied between two or more proteomes. Following this global visualization of the proteome, proteins of interest can be dissected and identified through MS. However, gel-based approaches are not sufficiently sensitive to visualize and enable identification of low abundance proteins, being the most likely candidates for protein biomarkers. Alternative workflows to quantitative proteomics analysis rely on bottom-up LC-MS/MS profiling of heavy/light tagged proteins (see chapter 1). However, as illustrated above, comprehensive profiling entails considerable fractionation and analysis time, as well as appropriate data mining to uncover candidate protein biomarkers. Clearly, flaws exist in each of these two workflows for biomarker discovery.

The premise of biomarker discovery is that only a small portion of a proteome is varying between two physiological states of the organism. By this reasoning, a shotgun approach to comprehensive proteome profiling is not only time consuming but inefficient in that the majority of the identified proteins do not display variable expression. A preferred alternative would be to conduct a protein identification experiment in a targeted fashion. A targeted proteomics experiment involves the detection of ions of a specific mass to charge ratio. This lends to greater sensitivity and speed of analysis. However, the determination of which proteins (peptides) should be targeted, and how these proteins can

be effectively visualized, is the focus of the current chapter.

The vMALDI LTQ linear ion trap mass spectrometer is a relatively recent addition to MS^n capable instruments with a MALDI source [195]. One of the key advantages of this instrument, over other MALDI TOF platforms, is the low sample consumption rate afforded by a low repetition laser and a sensitive ion trap analyzer which requires pulsed ion filling. This allows a single spot to be analyzed over an extended period without consuming the entire sample. Of course, a downfall to this slow acquisition laser is that the vMALDI LTQ is not ideal for comprehensive peptide identification through MS/MS. To date, the vMALDI LTQ has mainly been used as an imaging tool for profiling tissue samples [195-200]. However, we recognize a potential for the vMALDI LTQ in a protein biomarker discovery experiment. With a goal of sensitive analysis of a tagged (heavy/light) proteome mixture, MALDI MS could potentially serve a prominent role.

This chapter presents an evaluation and application of the Thermo vMALDI LTQ as a profiling tool for fast determination of peptides which are of interest for down stream analysis in a targeted quantitative workflow. First, a comparison of MALDI MS and MS/MS with ESI MS/MS on this instrument is performed, highlighting the strengths and limitations of each technique. From this, a comprehensive proteomics profiling workflow employing isotopically differentiated peptides is developed, incorporating the multiplexed LC-MALDI platform presented in chapter 3, which aims to improve the throughput of candidate protein biomarker discovery.

4.2 Experimental

4.2.1 Materials and Reagents

All standard proteins used throughout the experiments, including bovine trypsin (TPCK treated to reduce chymotryptic activity), were purchased from Sigma (Oakville, ON, Canada). *S. cerevisiae* (cat. YSC2), formic acid and trifluoroacetic acid (TFA) were also purchased from Sigma-Aldrich. Milli-Q grade water was purified to 18.2 MΩ cm^{-1}. The MALDI matrix, α-cyano-4-hydroxycinnamic acid (CHCA), was purchased from Sigma (cat. C2020), and was recrystallized from 60/20/20 ethanol/methanol/water prior to use. Solvents were of HPLC grade and were from Fisher Scientific (Ottawa, ON, Canada). Deuterated formaldehyde, originally obtained from Cambridge Isotope Laboratories (Andover, MA), was a gift from Ken Chisholm at the Institute for Marine Biosciences in Halifax.

4.2.2 Sample Preparation and Tryptic Digestion

Protein standards were dissolved to a concentration of 2 mg/mL in water and kept at -25^{o}C. Standards were diluted to 0.1 mg/mL in 100 mM Tris-HCl buffer prior to tryptic digestion.

Yeast proteins were extracted with three passes through a French press at 10,000 psi with 2% SDS in 20 mM Tris-HCl pH 8.0. Protein concentration was determined to be about 10 mg/mL by Bradford assay and 1 mg aliquots were frozen at –25^{o}C prior to use. Prior to digestion, SDS was removed using the chloroform/methanol precipitation method as described as described by Wessel and Flügge [201] with two additional wash steps with 400 µL of methanol to further reduce SDS concentration, centrifuging between each wash. This was followed by suspension in 500 µL of 100 mM triethylammonium

bicarbonate (TEAB) buffer at pH 8.0.

Prior to tryptic digestion, disulfide bonds were alkylated and reduced using 25 µL of 200 mM dithiothreitol, incubating for 20 minutes at 55^{o}C, followed by the addition of 5 µL of 200 mM iodoacetamide, with a 20 minute incubation in the dark, for every 5 µg of protein standard, using 10 times the volume for the digestion of the yeast proteome. Digestion was started with the addition of trypsin at a ratio of 1 to 100.

4.2.3 Differential Isotopic Labeling

Isotopic labelling of the resulting peptide digests was performed through the addition of 1.8 µL of a 20% protonated or deuterated formaldehyde solution for every 200 µg of sample, which subsequently derivatized any free amine present, namely the N-termini and lysines, followed by 5 minute incubation at room temperature. This was followed by the addition of 2.1 µL of a 6 M sodium cyanoborohydride solution and incubation for 2 hours to reduce the derivatized amines to form dimethyl functionalized groups with mass shifts of 28 or 32 amu for the light and heavy label respectively. Control and test samples were then combined in equal volumes and sample cleanup was achieved through reversed phase chromatography. The cleaned peptides from RPLC were lyophilized and kept at -25^{o}C until just prior to MS analysis.

4.2.4 Matrix Deposition

Following removal of the Teflon well plate, a 2.5 µL solution of 2 mg/mL CHCA, prepared in 50% ACN, 8% EtOH, and 42% water, was deposited on target. The solvent was allowed to evaporate at room temperature.

4.2.5 MALDI MS/MS and ESI-MS/MS

Mass spectrometry was performed using a ThermoFisher LTQ linear ion trap

mass spectrometer equipped with an ESI or vMALDI source. The MALDI source employs a 337 nm nitrogen laser operating at 20 Hz. All data were acquired in positive MS ion mode and CID of the precursor ions was accomplished using helium as the collision gas. Instrumental parameters common to all MALDI MS experiments were: automatic spectrum filter (ASF) and automatic gain control (AGC) both on; crystal positioning system (CPS) for MALDI was used to select the location for sample ablation; thresholds for MS and MS/MS signals were set to 5000 and 500 respectively. For both MALDI and ESI, a dynamic exclusion list was employed to remove ions selected for MS/MS analysis from further experiments. The ESI voltage was set to 2.5 kV and transfer capillary was set to 200^{o}C. The scan range for the MALDI MS experiments was set to 800 - 4000 *m/z*, and 400 - 2000 *m/z* for the ESI MS experiments. Max fill time for the ion trap was set to 100 ms, and automatic gain control was set to allow up to 30000 ions to enter the trap. For ESI, one MS scan was performed followed by 5 MS/MS. Parameters that were varied to optimize the MALDI MS/MS analysis were: number of microscans for both MS and MS/MS spectra, the data dependant acquisition procedure, and total analysis time per spot. An average of 5 independent measurements, from individually spotted samples, was used to obtain each measurement.

4.2.6 Data Analysis

Using the SEQUEST algorithm within the Bioworks software (Rev. 3.2), acquired MS/MS spectra were searched against a custom built database containing the 10 protein standards used in these experiments, including trypsin and keratin (a common contaminant). Yeast peptides were searched against the *S. cerevisiae* proteome (5/31/2005). Precursor mass tolerance was set to ±2, full tryptic digestion with 3 missed

cleavages, and dynamic modifications of carbamidomethylated Cys (+57.0215) and phosphorylated Ser, Thr, and Tyr (+79.9663). Xcorr of 1.9 (+1), 2.2 (+2), and 3.75 (+3), peptide probability < 1E-4 and ΔCN < 0.1 were set as filters, giving a false positive rate of < 1% when searched against a decoy database.

4.3 Results and Discussion

A general outline of the targeted quantitative proteome workflow is as follows:

(1) Individually digest, then respectively label a control (healthy), and test (diseased) proteome with formaldehyde (H/D)

(2) Combine control and test proteome samples, and fractionate through multidimensional/multiplexed (parallel) LC.

(3) Collect fractions and deposit on target with MALDI well plate collector, evaporate and prepare for MALDI MS analysis with matrix addition.

(4) Perform MS scans of the entire set of collected fractions.

(5) Data analysis identifies pairs of differentially expressed peptides, establishing targeted biomarker candidates

(6) Reanalysis of select spots, through MALDI MS/MS, establishes identity of candidate biomarkers.

The development of this quantitative MALDI MS profiling strategy incorporating the multiplexed LC-MALDI system (chapter 3), first involves a comparison of MALDI MS/MS and ESI MS/MS data dependent analysis strategies. This comparison established the effectiveness of MALDI MS/MS peptide sequencing to that of conventional ESI-MS/MS in terms of comprehensive peptide identification and, most importantly, speed of analysis. As tandem MS analysis of peptides is required for unambiguous identification,

its use in a quantitative profiling strategy would be extremely beneficial. However, as it will be shown, the use of MALDI MS/MS peptide sequencing for comprehensive, high-throughput profiling, is not an attainable goal. Fundamental limitations of this ionization method do not predispose MALDI MS for use in this fashion. However, the comparison also establishes the potential for MALDI MS to incorporate into a targeted proteomics analysis workflow, a strategy which would be difficult to perform through ESI.

4.3.1 Quality versus Quantity for Optimal MALDI MS/MS Profiling

The number of microscans, which is the number of spectra used to generate a final averaged spectrum, used for MS and MS/MS spectra acquisition has an effect on the number of MS spectra which can be collected in a given amount of time. A data dependent acquisition (DDA) amounts to an automated approach to selecting peaks of interest from an MS scan, and subjecting these peaks to MS/MS analysis. Optimization of this parameter essentially involves a balance between quantity and quality; more microscans will produce higher quality MS spectra, but this comes at the cost of collecting a lower total of spectra in a given amount of time. With DDA, a greater number of MS/MS scans per MS scan also results in acquiring more spectra for the purpose of protein identification. However, as the laser moves to different positions on a given MALDI spot, it is also possible that the spectrum will change and thus renewed profiling of the spot through MS may be beneficial. Optimization of both parameters is important to maximizing the number of peptide identifications in a time-limited experiment.

Optimization was performed using 5 replicates of 1 pmol or 100 fmol of a BSA digest on the MALDI target. For the microscan optimization, 1, 2, 3, and 5 scans were

evaluated for both the profiling MS and MS/MS scans. For the DDA optimization, from 1 to 10 MS/MS scans were examined per profiling MS scan. Finally, these parameters were also subjected to a time controlled experiment, acquiring data from a single spot for 15 s, 30 sec, 1 min, and 3 min (for microscan optimization) or 5 min (for DDA optimization). The results of the microscan optimization and DDA optimization are shown in Figures 4.1 and 4.2 respectively.

As seen in Figure 4.1, it was found that 1 microscan for the MS/MS experiment and 2 microscans for the MS profiling scan were sufficient to obtain maximal sequence coverage for both the 1 pmol (Figure 4.1A) and 100 fmol (Figure 4.1B) samples over all time points. Essentially, from these results, it would appear that maximizing "quantity", as opposed to increasing "quality" is the best strategy for data collection on the vMALDI LTQ. Averaging spectra limits the total number of scans produced in a given analysis time. This is essentially the same effect that is happening when the time for data collection is reduced. MALDI scan speed is constant, at approximately 1 second per scan, and so the best approach to maximizing the number of peptide identifications, represented as sequence coverage in this instance, is to ensure that every scan generates a unique spectrum.

The same general result was observed from the optimization of the data dependent acquisition strategy, which is shown in Figure 4.2. For this experiment, the number of MS/MS spectra which were collected for each profile scan performed was varied from 1 to 10. The premise for this is that generally the MALDI MS spectrum may vary over a given spot, due to possible "hot spots" in the sample from inhomogeneous crystallization. As such, performing several profile scans would be advantageous to ensure the sample is

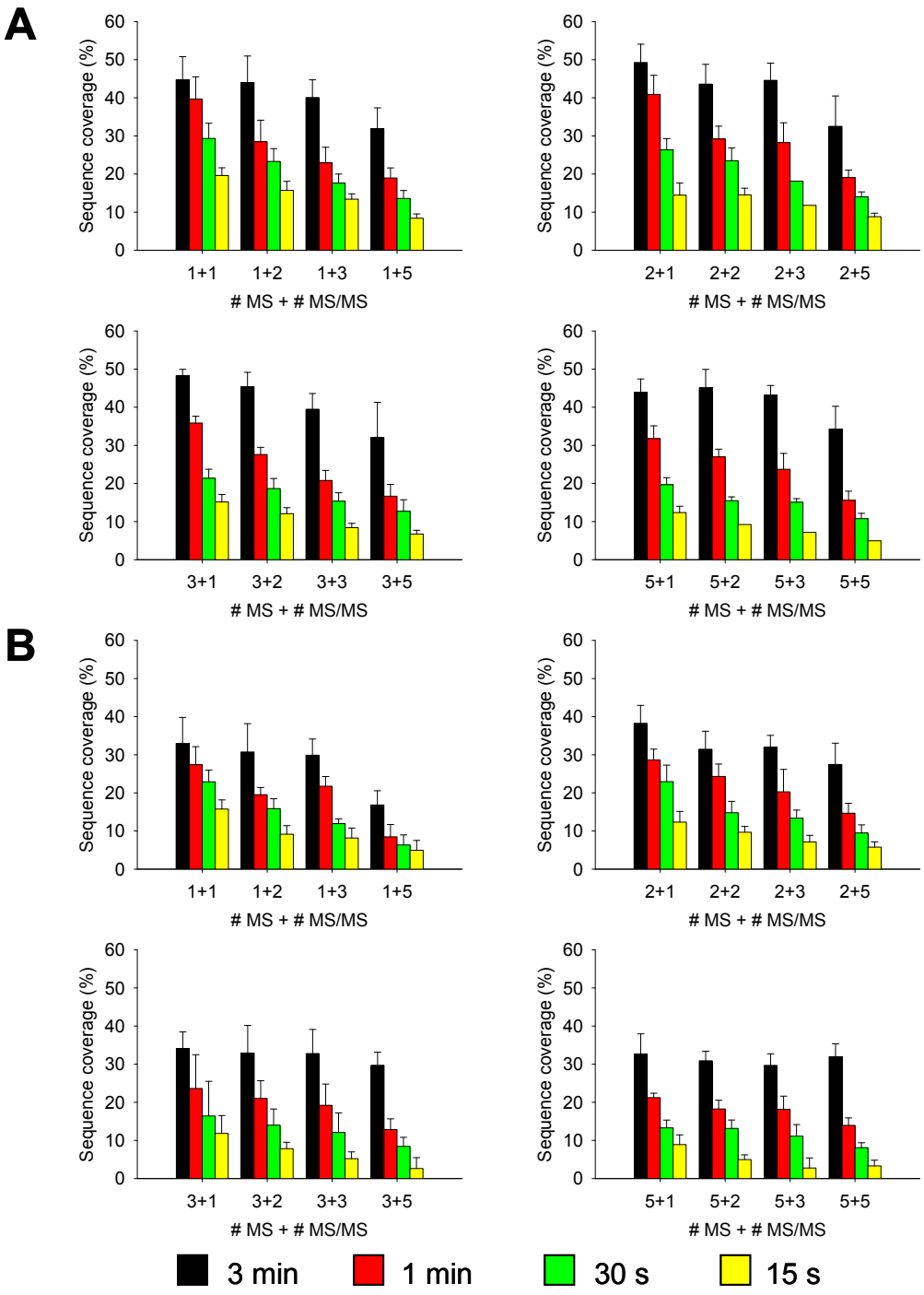

Figure 4.1 Plots of average (n = 5) BSA sequence coverage versus number of microscans. Data obtained through data dependent acquisition MALDI MS/MS analysis of (A) 1 pmol and (B) 100 fmol of sample using 1, 2, 3, or 5 microscans per MS and MS/MS spectra over varying amounts of time (see legend). Greatest sequence coverage was obtained from analysis that used no more than 1 microscan for collection of the MS/MS scan and 2 microscans for the MS scan for both sample amounts and data collection times.

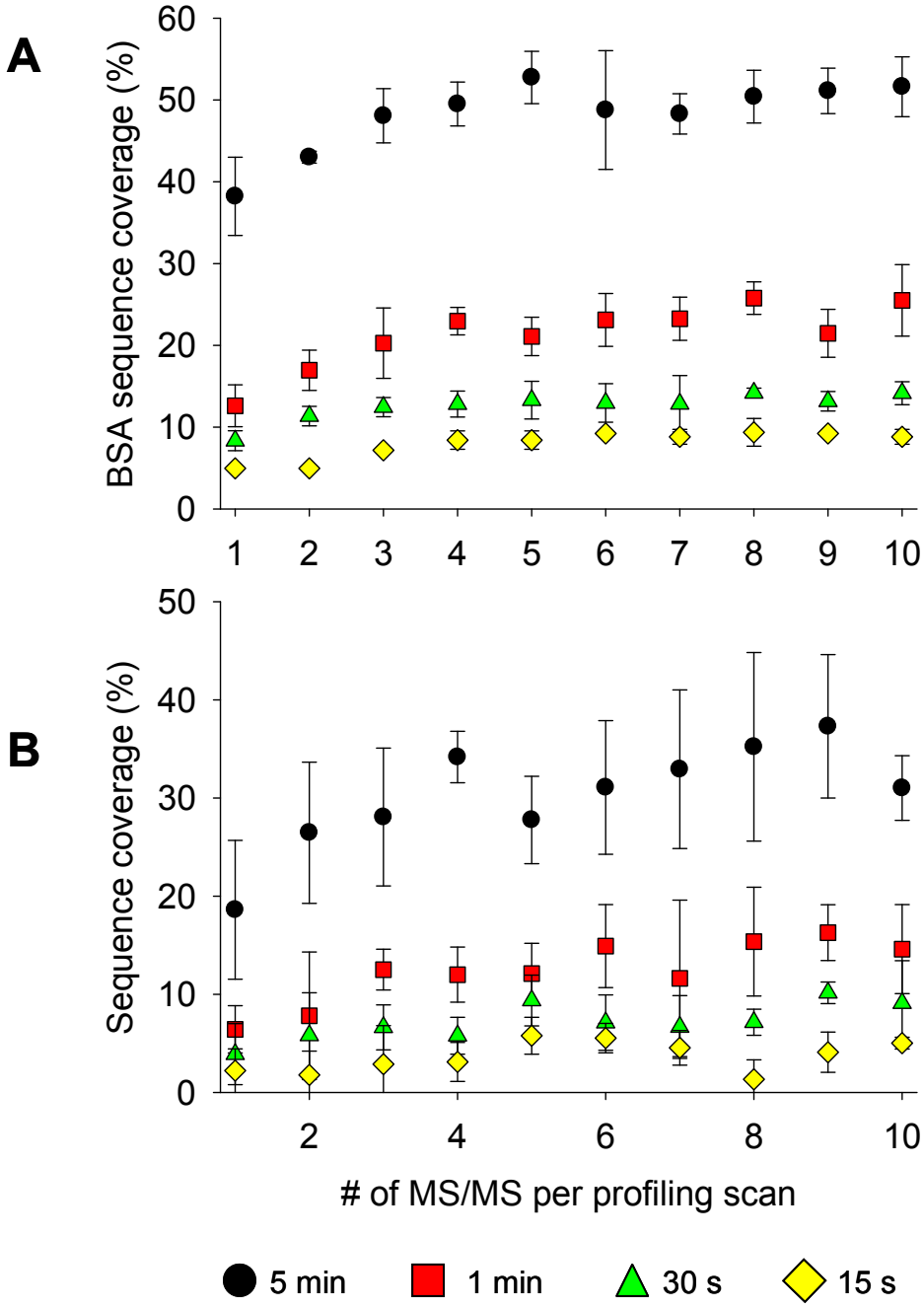

Figure 4.2 Plots of average (n = 5) BSA sequence coverage versus DDA profiling method. obtained from data dependent acquisition MALDI MS/MS analysis of (A) 1 pmol and (B) 100 fmol of sample. Numbers of MS/MS scans for every one MS profiling scan was varied allowing different amounts of time for analysis (see legend). Greatest sequence coverage was obtained when at least 4 or 5 data dependent MS/MS scans were performed for every one MS profiling scan.

thoroughly profiled. However, the time spent performing the profile scan sacrifices the number of MS/MS scans which could have been performed in its place. From Figure 4.2, a noticeable increase in identified peptides was obtained when increasing the number of MS/MS scans per MS profile scan from 1 to 4. Beyond this, no noticeable increase in sequence coverage was seen at all time points shown. Thus, performing 4 to 5 MS/MS scans for every MS scan appeared to be optimal for all time points.

4.3.2 Comparison of MALDI and ESI for Analysis of a Complex Mixture

Both of the experiments above show the sequence coverage obtained from a single protein, BSA, through varying collection parameters. While a single BSA digest will generate several peptide products, a real proteome sample, even with fractionation, could be much more complex. It is also of interest to compare the abilities of the vMALDI LTQ to handle MS/MS analysis a complex mixture relative to ESI-MS/MS. MALDI generally produces simpler spectra, due to the predominance of singly charged ions as well as reduced effects from the presence of contaminants, thus it may be that MALDI is a better candidate for the analysis of complex mixtures. A tryptic digest of a 10 protein mixture (listed in Table 4.1), results in the generation of hundreds of peptide products. From this, a mixture containing 1 pmol per protein was spotted for MALDI, while direct infusion of 1 pmol/μL per protein (2 μL/min) was chosen for ESI. A 20 minute analysis time was examined, enabling detection of a maximum number of peptides with each technique. The results of this comparison are shown in Table 4.1. From 20 minute analysis, it was found that the number of peptides and proteins identified by the MALDI MS/MS experiment from a single spot was greater than that obtained from the ESI experiment. A total of 33 peptides, corresponding to 9 of the 10 proteins,

Table 4.1 Proteins identified in 20 minutes using ESI and MALDI.

	Sequence coverage (%)	
Protein	MALDI	ESI
Amyloglucosidase	4.7	0.0
BSA	24.3	27.9
Ovalbumin	4.4	9.9
α-casein	26.6	0.0
Carbonic anhydrase	14.3	8.1
β-lactoglobulin	21.0	18.5
Myoglobin	21.5	57.5
Cytochrome c	25.0	67.3
Lysozyme	39.5	39.5
Ubiquitin	0.0	0.0
Total # peptides identified	33	27
Total # proteins identified	9	7

where identified from MALDI MS/MS analysis, while only 27 peptides corresponding to 7 proteins were identified from the ESI MS/MS analysis. This demonstrates that, given sufficient time, MALDI can handle considerably complex mixtures better than ESI. Even though the scan speed for ESI MS/MS is much faster than MALDI MS/MS (close to an order of magnitude greater spectral acquisition speed), MALDI still has potential to profile a complex mixture.

4.3.3 Evaluation of MALDI MS/MS for High-throughput Data Collection

While the favourable capabilities of MALDI MS/MS on the LTQ have been demonstrated, the practical applications of MALDI profiling in a high-throughput profiling strategy for biomarker discovery must be considered. Consideration to the time required for analysis is critical, and in the case of MALDI this translates to the time spent per spot for data collection. Table 4.1 illustrates that with "unlimited" time, MALDI can perform quite strongly. However, an optimal amount of analysis time per spot needs be chosen which balances both the amount of data collected, with the overall throughput of the experiment. Figure 4.3 plots the number of peptides, and corresponding proteins, identified over the 20 minute analysis time from the 10 protein mixture described in the preceding section. While several peptides/proteins were detected very early over the 20 minute data collection period, it is clear that, as the analysis progressed, a diminishing rate of return in terms of peptide identification is achieved. The curve reaches a near plateau at approximately 10 minutes of data collection, with marginal gains when the collection time was doubled to 20 minutes. The general conclusion from this figure is that MALDI MS/MS analysis of a complex mixture is possible, with potentially superior profiling relative to LC-ESI-MS/MS. However, the amount of time required to achieve

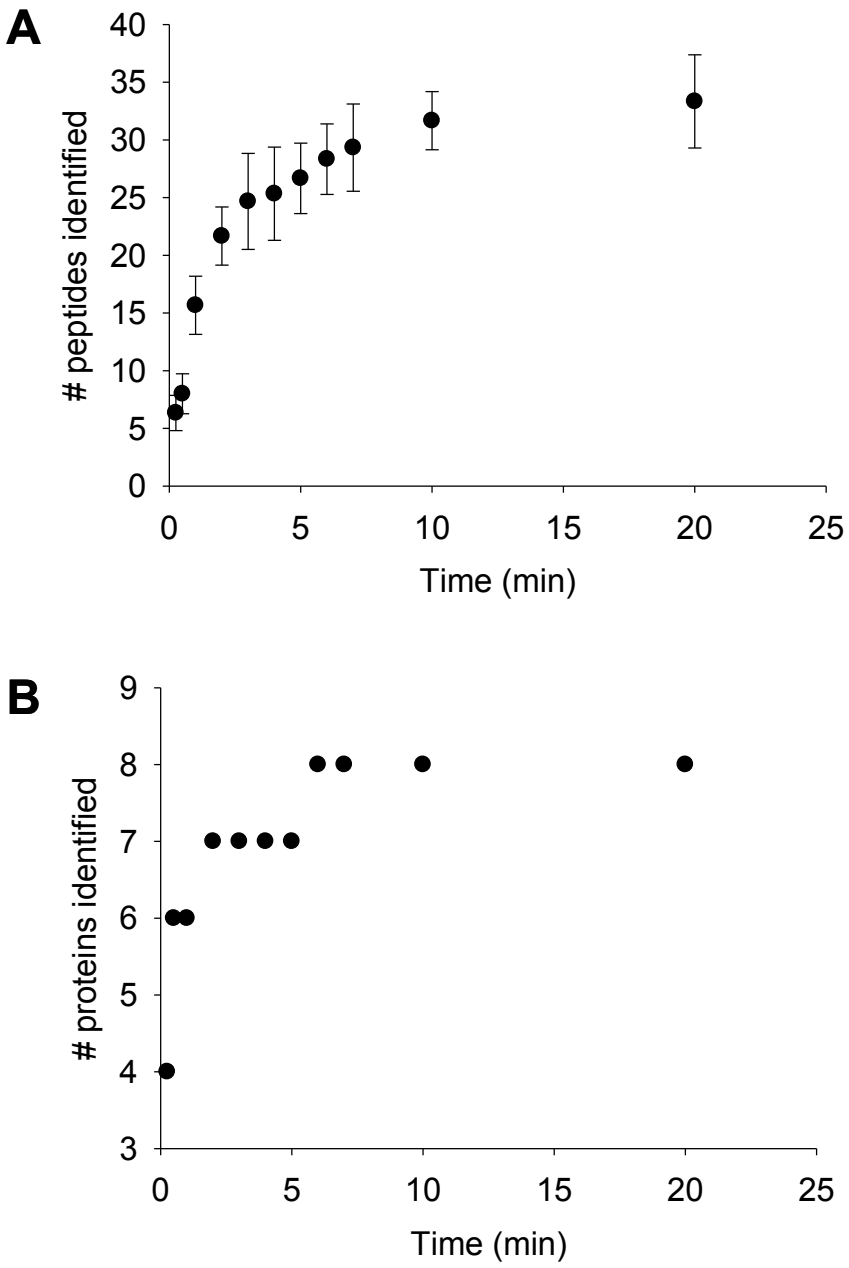

Figure 4.3 Plots showing the number of (A) peptides and (B) proteins identified over time from a 10 protein standard digest mixture using MALDI MS/MS in data dependent acquisition. Maximum protein and peptide identifications were reached by 10 minutes of data collection with only slight increases by the 20 minute time point.

high proteome coverage does not enable a high-throughput strategy. For example, considering a typical RPLC separation of a complex proteome digest, if 1.5 minute fractions are collected from the separation, this would result in approximately 80 MALDI spots. Optimal data acquisition (20 min/spot) from this relatively simple ~2 hour LC separation through DDA MALDI MS/MS would take a full day of MS analysis to complete. By contrast, similar analysis through online LC-ESI-MS/MS would naturally take as long as the separation. While the total number of peptides identified through online ESI-MS may not match that of an in depth MALDI MS/MS experiment, the time scale between these experiments still favours ESI. Experience with the LTQ in ESI mode dictates that approximately 500 yeast proteins can be identified per LC/MS run (results not shown).

For MALDI MS/MS to be an effective high-throughput method, the time scale of the analysis should be similar, if not shorter, than that of an LC-ESI-MS/MS experiment. This would require that 1 minute analysis time is allotted per spot from a typical separation with the collection of 1 minute fractions. From Figure 4.3, about 5 to 10 peptides can be identified from a spot in this time. Extrapolating from these numbers, and assuming that unique peptides are identified in each spot, an LC-MALDI experiment would identify roughly the same number of peptides to a typical LC-ESI-MS/MS run. However, given that LC-MALDI is much more difficult to automate relative to LC-ESI, the conventional experiment for protein identification is simply not practical.

The use of the vMALDI LTQ does permit MS/MS analysis, and thus can be used for targeted peptide (protein) identification. This is demonstrated through analysis of BSA peptides with MALDI and ESI MS in order to improve the confidence in

identification through targeted MS/MS analysis. The LC separation was repeated for replicate injections, subjecting the first run to online LC-ESI analysis, and the second run to LC-MALDI sample collection. While several BSA peptides were confidently identified (i.e., those who meet the filtering criteria) through LC-ESI-MS/MS, others did not meet the filtering criteria and could not be confidently identified. Targeted analysis of low confident peptides using MALDI MS/MS could improve their scores and BSA sequence coverage. For example, the BSA peptide KVPQVSTPTLVEVSR was targeted in the corresponding MALDI fraction determined from the LC-ESI-MS/MS run (results not shown). The Xcorr value of this peptide increased from below a threshold value of 1.37 to a confident value of 4.06. This demonstrates that, given sufficient analysis time, MALDI MS/MS can yield high confident protein identifications in a targeted strategy.

4.3.4 Quantitative MALDI MS Profiling of a BSA Digest

Beyond a targeted MALDI MS/MS analysis, MALDI MS profiling can analyze hundreds of fractions generated from an LC separation in a relatively short amount of time. In relation to early phase biomarker discovery, this approach is exactly what is required for high-throughput analysis. This approach, however, requires a method to differentiate between a test and control sample in order to determine which peptides may be selected for targeted analysis. One method is to use differential isotopic labeling of the test and control sample in order to uncover differences in relative abundances for respective peptides. The approach to protein labeling chosen incorporates dimethylation of free amines using formaldehyde or deuterated formaldehyde [145-151].

This labeling method was evaluated for use on the high-throughput LC-MALDI MS platform. A BSA digest was split in two, one being subject to light tagging and the

other to heavy tagging, with subsequent mixing of samples in a 1:1 ratio. A one dimensional LC separation was performed, injecting eight replicate samples onto the multiplexed RPLC columns (described in chapter 3). Fractions were collected using the LC-MALDI well plate device. Total analysis time for MALDI was 30 seconds per spot with averaging of 15 MS spectra. MS spectra, collected from two time points, on columns 1, 4, and 8 are presented in Figure 4.4. The MS spectra show that the consistency in elution time between the columns of the parallel system is very high. Also, the expected 1:1 ratio of BSA peptides (shown in Figure 4.4D) validates the isotopic labelling procedure for accurate quantitation analysis. A control test of this system for the purpose of targeted biomarker analysis is presented below.

4.3.5 High-throughput LC-MALDI MS Profiling for Biomarker Discovery

To test the MALDI workflow for high-throughput biomarker discovery, a yeast proteome was digested, split equally, differentially labelled, and combined at a 1 to 1 ratio; making the assumption that the relative abundances of the majority of proteins in differential proteomics analysis are at a ratio of 1 [202]. Isotopically labelled peptides generated from digestion of protein standards were spiked in to act as “biomarkers”. A list of the protein standards with the concentration they were present in the sample, the ratio at which they were spiked into the sample, as well as the number of peptides which were identified is given in Table 4.2. This mixture was subjected to strong cation exchange, splitting into 8 fractions, followed by multiplexed RPLC separation of each fraction. Eighty fractions per column were collected across two MALDI well plates (chapter 3). All spots (640 in total) were profiled using MALDI MS, averaging up to 15 microscans per spot, with an analysis time limited to no more than 45 seconds per spot

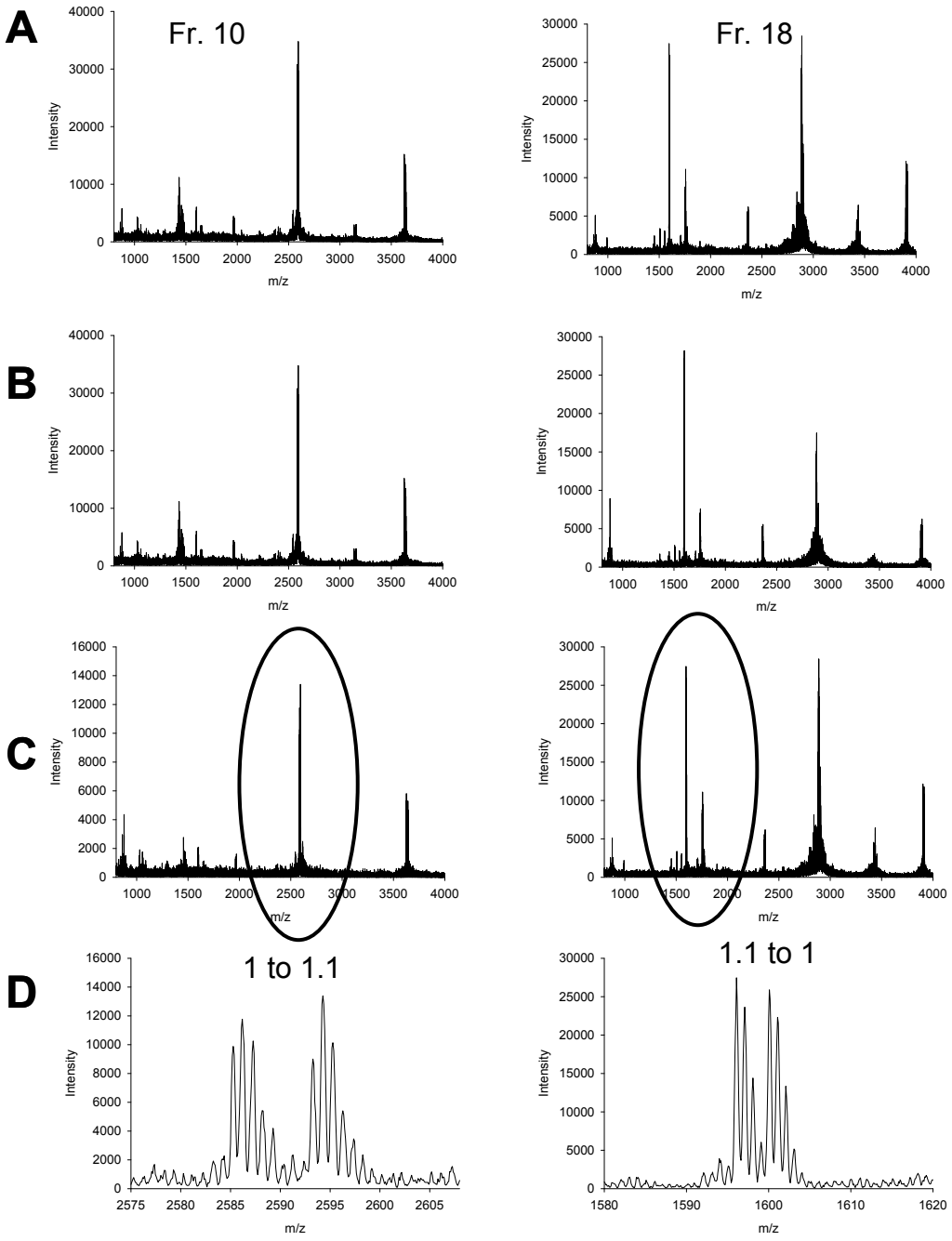

Figure 4.4 LC-MALDI MS analysis of differentially labeled BSA peptides. Scans shown are samples of fractions 10 and 15 from columns (A) one, (B), four, and (C) 8. The consistency in the elution between columns as well as the expected one to one ratio, clearly shown in D, shows that this approach may be useful for incorporation into a high-throughput quantitative profiling workflow.

with an average data acquisition is complete in 30 seconds per spot. Total analysis was complete in less than 5.5 hours. In comparison, ignoring replicate analysis, online LC-ESI-MS/MS analysis of 8 IEX fractions would take ~16 hours to complete employing a conventional 2 hour RPLC separation.

The majority of ion pairs observed in these MS spectra exhibit 1:1 ratios, and thus originate from the "background" yeast proteins. Peptide pairs exhibiting ratios other than 1:1 were also observed, and these peptides are of interest to a biomarker profiling experiment. For example, shown in Figure 4.5, the profiling of IEC fraction 1, RPLC fraction 28, revealed a peptide pair at masses 1278.0 and 1290 at a ratio close to 2:1. Indeed, when looking at the peptides produced from the protein standards, this pair matches to a peptide generated from the tryptic digestion and labelling of β-lactoglobulin, which was indeed spiked into the yeast sample at a ratio of 2 to 1 at concentration of 5% *w*/*w* of the total mass of yeast protein sample (Table 4.2). This strategy was capable of detecting the lowest protein spiked into the sample, ubiquitin, which was present at a concentration of 0.1% *w*/*w*, with one peptide being identified.

Some deviation in the relative abundance ratios was observed. This is attributed to the presence of background yeast peptides which overlap in *m*/*z* value to that of the peptides of the protein standards. For the purpose of early stage biomarker discovery, however, this would not be a major difficulty, as only peptide pairs with a relative abundance ratio deviating from a value of 1 are of interest. More accurate quantitation would be performed during the late stage targeted analysis. However, there is a possibility for false positive detection of potential biomarkers due to the interference of peptides with similar masses in the sample. Overlap in the MS spectrum would artificial

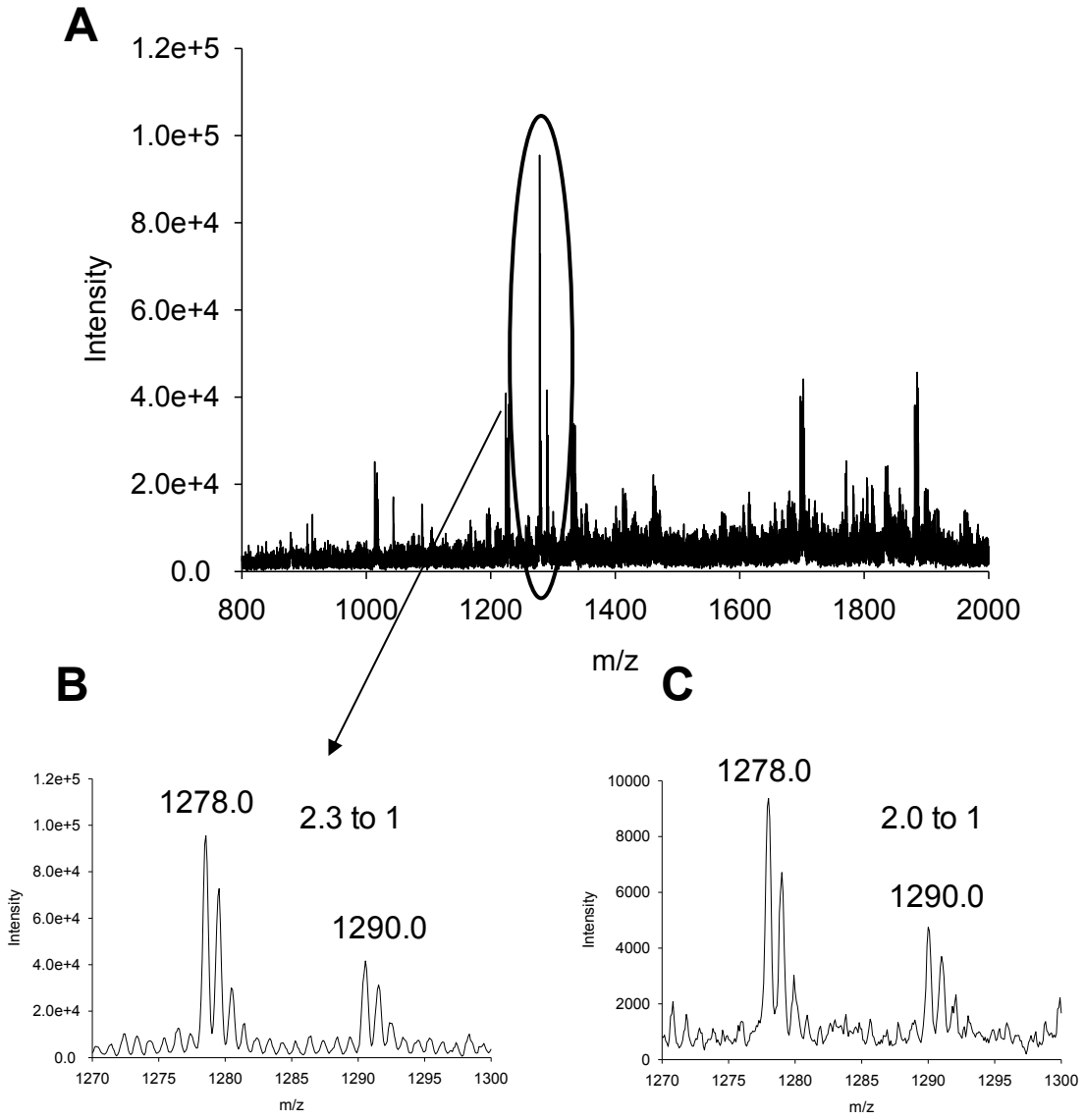

Figure 4.5 LC-MALDI MS profiling of a yeast peptide mixture containing spiked protein standards as model biomarkers. (A) MS profile scan of RPLC fraction 28 from IEC fraction 1. The majority of peptide ions observed were from yeast proteins and have relative ratios near 1 to 1 with their differentially tagged partner. One pair of masses, however, was observed with a ratio near 2:1. The masses and ratio of these peptides (B) were matched to those belonging to β-lactoglobulin A as observed in a control spectra (C).

Table 4.2 Detection of protein standard "biomarkers" in yeast.

Protein	Concentration (% *w*/*w*)	# peptides observed	Spiked ratio (H/D)*	Average observed ratio (H/D)*
β-lactoglobulin	5	5	2/1	2.0 ± 0.3
Carbonic anhydrase	1	4	1/3	0.4 ± 0.1
α-casein	0.2	1	50/1	N/A
Ubiquitin	0.1	1	5/1	N/A

H = dimethylated using light (protonated) formaldehyde

D = dimethylated using heavy (deuterated) formaldehyde

skew the relative abundance ratio from 1. While this may be a minor issue, it would lead to a loss in time during the validation stage as these false discoveries are examined later in targeted strategies. This highlights the need for effective separation methods at both the intact and peptide level.

4.4 Conclusions

In this chapter, the development of a high throughput MALDI MS proteome profiling strategy, incorporating multiplexed LC-MALDI separations is presented. Validation of the strategy followed a comparison of MALDI and ESI ion sources on the linear ion trap mass spectrometer, which highlighted how comprehensive MALDI MS/MS profiling is not a practical method for high-throughput proteome characterization. While MALDI MS/MS can compete with conventional LC-ESI-MS/MS data dependent analysis, generating large numbers of high confident peptide identifications, the time needed for comprehensive profiling is significantly greater than ESI. As an alternative strategy, targeted MS/MS analysis is possible, following complete MALDI MS profiling of multiple spots in a short period of time.

Chapter 5

Tryptic Digestions in Aqueous/organic Solvent Systems

5.1 Introduction

Sample preparation is a critical component of the proteomics workflow. In particular, bottom-up peptide analysis relies on effective methods for proteome digestion prior to MS analysis. Often incorporating enzymatic cleavage, this vital component of the analytical workflow aims to provide adequate quantities of suitable peptides for protein identification or quantitation. In terms of throughput, sample digestion can represent a bottleneck in the analysis and many protocols require overnight incubations to complete the digestion. In order to streamline and improve throughput of a proteomics workflow, including the LC-MALDI platform described in the previous two chapters, improved protein digestion protocols must be developed and evaluated.

The most popular protease used in proteomics analysis is trypsin. A variety of approaches to enhance tryptic digestion efficiency have been developed. These methods include the modification of solvent composition [203-206], the addition of solubilizing detergents [207], the use of microwave energy to enhance digestion [208, 209], the immobilization of trypsin or substrate to a bead or nanoparticle surface [210-212], as well as off-line and on-line microreactors [213-216]. Of interest are the reports of improved MS sequence coverage from tryptic digests in solution prepared with high concentrations of organic solvent [204, 205, 217-219]. Besides effecting the enzymatic activity, organic solvents also denature the protein sample, allowing for easy access of the protease to the protein. The simplicity of this procedure has significant benefits for performing routine

high-throughput analysis and would have potential to integrate into our LC-MALDI platform (chapter 3). In this case, fractions collected following intact RPLC separation could be immediately subject to in-well enzymatic cleavage, prior to concentrating the samples onto the MALDI collector plate. The incorporation of organic solvents has also been shown to have significant benefits for the solubilization and digestion of hydrophobic and globular proteins.

With a goal of incorporating the method of organic solvent-assisted tryptic digestion of proteins into the LC-MALDI workflow, a detailed investigation on the efficiency of tryptic digestion in high organic solvent was conducted. Our work confirmed the results in that the MS sequence coverage from tryptic digests in high organic solvent had increased.

Presented in this work is an examination of the effect of organic solvents on the proteolytic activity of trypsin and the nature of the resulting peptide products. Using a standard BAEE (N_{α}-benzoyl-L-arginine ethyl ester) assay, the level of tryptic activity was assessed over varying concentrations of acetonitrile at varying temperatures and pH. The proteolytic products were extensively characterized through a combination of liquid chromatography coupled to ultraviolet detection (LC-UV) as well as one-dimensional gel electrophoresis followed by silver staining, and qualitatively *via* MS-based peptide sequencing. A detailed examination and comparison of the particular peptide products generated under different digestion conditions was also provided through the use of isotopic labeling for relative quantitation of digestion products. Contrary to previous reports, these results reveal that the activity of trypsin at high concentrations of organic solvent is significantly impaired. An alternative explanation for improved sequence

coverage under apparently poor digestion conditions is therefore provided.

5.2 Experimental

5.2.1 Materials and Reagents

All standard proteins, including TPCK-treated trypsin and N_{α}-benzoyl-L-arginine ethyl ester (BAEE) were purchased from Sigma (Oakville, ON, Canada). *S. cerevisiae* (cat. YSC2), formic acid, trifluoroacetic acid (TFA), and trisodium phosphate were also purchased from Sigma. Tris HCl, iodoacetamide and dithiothreitol (DTT), along with gel preparation materials were purchased from Bio-Rad (Hercules, CA). Milli-Q grade water was purified to 18.2 $M\Omega cm^{-1}$. Solvents were of HPLC grade and were from Fisher Scientific (Ottawa, ON, Canada).

5.2.2 BAEE Assay

Trypsin activity was determined using a standard protocol [220, 221]. In brief, a 3 mL mixture of BAEE (0.086 g/L) was prepared in 67 mM Tris/HCl or 67 mM phosphate buffer at various pH, organic composition and temperature. To this mixture, 0.2 mL of trypsin (0.012 g/L) in 0.001 M HCl was added directly before measurement. Absorbance readings of the sample were taken at 253 nm over 5 minutes using an Agilent 8353 spectrophotometer (Palo Alto, CA) with a 1 cm pathlength quartz cuvette.

5.2.3 Tryptic Digestion

100 µL aliquots of protein sample were buffered to pH 8 with 100 mM Tris-HCl buffer. The breaking and reduction of disulfide bonds was performed by addition of 2 µL of 200 mM DTT and incubation at 55°C for 20 minutes, followed by 5 µL of 200 mM iodoacetamide and incubation at room temperature for a further 20 minutes in the dark.

Digestion was performed using a 50:1 protein to trypsin mass ratio at various temperatures and durations. Digestions were stopped by lowering the pH to approximately 3 with the addition of 5 μL of 10% formic acid or TFA.

5.2.4 Liquid Chromatography

Reversed-phase liquid chromatography was performed using an Agilent 1200 HPLC system (Palo Alto, CA) with autosampler and diode array detector. A 5 μL portion of digested sample was injected onto an Agilent Zorbax SB C_{18} column (150 × 0.5 mm i.d.) at a flow rate of 10 μL/min followed by gradient elution between solvent A (water, 2% acetonitrile, 0.1% TFA, *v*/*v*/*v*) and solvent B (100% acetonitrile, 0.1% TFA, v/v). The gradient consisted of a 5 minute hold at 2% B followed by an increase from 2 to 52% B over 50 minutes and then ramping to 80% B in 60 seconds to wash the column. The eluting peptides were monitored at wavelength of 214 nm.

5.2.5 Gel Electrophoresis

BSA samples digested with trypsin under varying conditions were processed by gel electrophoresis. An 18% resolving gel with a 4% stacking gel in a discontinuous buffer system was used [90]. Samples were heated at 95°C for 5 minutes in a reducing sample buffer, and loaded in 50 μg aliquots. Molecular weight standard markers (Bio-Rad, 10-250 kDa) were loaded onto each gel to provide a comparison point for samples. Gels were stained using the Bio-Rad Coomassie Blue staining method according to manufacturer's instructions.

5.2.6 Mass Spectrometry

LC-MS was performed using a ThermoFisher LTQ linear ion trap mass spectrometer equipped with an ESI source. The ESI voltage was set to 2.5 kV and

transfer capillary was set to 225°C. The scan range was set to 400-2000 *m*/*z* for all experiments. Max fill time for the ion trap was set to 100 ms, and automatic gain control was set to allow up to 30,000 ions to enter the trap. Data dependant acquisition was set to collect 1 MS scan followed by MS/MS of the top 5 ions. Dynamic exclusion was used to exclude previously selected ions from further MS/MS analysis for two minutes. Reversed phase LC was performed by gradient elution from 98% A (0.1% formic acid in water) to 50% B (0.1% formic acid in acetonitrile) over 50 minutes. A volume of 1 µL of each digest was injected onto a Phenomenex monolithic C_{18} column (150 × 0.1 mm) and separated at a flow rate of 2 µL/min.

5.2.7 MS Quantitation and Data Analysis

Relative quantitation was performed of BSA peptides generated from the different aqueous and organic conditions using protonated and deuterated formaldehyde (14.3 µL of a 20% solution per 200 µg of peptide, 5 minute incubation) to isotopically label free amine groups (N-terminus and lysines) with a +28 or +32 mass shift for each labeling site following reduction with 16.7 µL of a 6 M sodium cyanoborohydride solution per 200 µg of peptides for 2 hours. Isotopically labelled peptides were cleaned of residual reagents through reversed phase chromatography, dried to remove solvent, and subjected to LC-MS/MS analysis. Data was analyzed using customized quantitation software in MASCOT (Matrix Science, London, UK).

For quantitation of the BSA peptides, peak lists for MS/MS spectra were extracted using the analyst QS script MASCOT.dll after smoothing, deisotoping, and discarding all spectra with less than ten peaks. Peptide sequences were then assigned using MASCOT to search the Swiss-prot database, limiting to the class *Mammalia*, with

the following search parameters: Peptide tolerance ± 1.2 Da, MS/MS tolerance ± 0.8 Da, max 3 missed cleavages. Fixed modifications were: carbamindomethyl (C), $2CH_3$ (K, N-terminus), $2CD_2H$ (K, N-terminus). MASCOT searches were based on a significance threshold of $p<0.05$. Automated quantification was performed with a Visual Basic program written by staff of the National Research Council Institute for Marine Biology (NRC-IMB, Halifax, NS, CA). Briefly, the program parsed the MASCOT results file for peptide information and accessed the raw data *via* Analyst QS library files. Only peptides that matched to bovine serum albumin better than any other protein in the database (#1 rank) with MASCOT scores of 25 or greater were selected for quantification. Peptide quantitation results with a correlation value below 0.7 were discarded and the results and raw data of the remaining peptides were manually inspected to ensure correct assignment of peptide pair masses and removal of any low intensity/noisy spectra.

5.3 Results and Discussion

5.3.1 Sequence Coverage Following LC-MS/MS Analysis

The use of organic modifiers to improve the efficiency of tryptic digestions for proteomics applications is a common protocol used by several labs [203, 205, 206, 209, 218, 222, 223]. The ability to improve the sequence coverage of low quantities of proteins in very high (80%) concentrations of organic solvent following LC-MS/MS has been demonstrated in short digestion times (1 hour or less) [204, 205].

To confirm these results, a similar experiment was conducted by comparing the sequence coverage obtained *via* LC-MS/MS from the tryptic digestion of BSA at different concentrations of acetonitrile. The reaction temperature was chosen as an

additional variable, assessing sequence coverage at both the optimal temperature for trypsin at 37°C as well as cooler at room temperature (24°C); the results of this experiment are presented in Table 5.1. The results obtained confirm that improved protein sequence coverage can be obtained through tryptic digestion at high concentrations of organic solvent. Specifically, while it is noted that a drop in enzyme activity is experienced in a solution comprising 50% ACN, the sequence coverage increases to a maximum level when the ACN concentration is further increased to 80%. Interestingly, an even greater improvement in sequence coverage was obtained from the digest performed at room temperature compared to that at 37°C. This improvement was seen at all concentrations of acetonitrile, and has not been previously reported in the literature. This result is counter intuitive in that one would expect a 37°C digestion would provide optimal enzyme activity, and thus optimal sequence coverage.

The experiment was repeated using additional standard proteins, and comparing 1 hour digestions at 24°C and 37°C in 80% acetonitrile; these results are given in Table 5.2. Again, improved sequence coverage was obtained for each protein standard from digestion performed at room temperature, rather than at the typical 37°C, while in the presence of 80% ACN. To investigate the underlying cause for improved sequence coverage, the enzyme activity of trypsin at varying concentrations of acetonitrile in water was evaluated through a standard BAEE assay, investigating both the temperature of the reaction, as well as the solution pH, using the standard BAEE assay.

5.3.2 Tryptic Activity Measurements using the BAEE Assay

The BAEE assay involves monitoring the rate of hydrolysis of N-benzoyl-L-arginine ethyl ester by trypsin and observing the change in the ultraviolet absorbance of

Table 5.1 Percent sequence coverage from a 1 hour BSA digest.

% ACN	Temperature (°C)	
	24	37
0	27.7	21.2
10	36.4	33.5
50	22.4	11.6
80	57.4	23.6

Table 5.2 Percent sequence coverage of protein standards.

	Temperature (°C)	
Protein	24	37
BSA	55.4	33.7
Myoglobin	45.1	9.8
Cytochrome c	51.0	34.6
β-lactoglobulin A	34.6	28.4
Lysosyme	46.5	20.2

the solution over time; the reaction is shown in Figure 5.1.The digestion variables included organic solvent concentration (0 to 80%), temperature (4 to 37^{o}C), as well as pH (4 to 10). A 3D surface plot was constructed from the results of this experiment, displaying activity versus %ACN and pH at room temperature (Figure 5.2A) and activity versus %ACN and temperature at pH 8.0 (Figure 5.2B). Phosphate buffer was used as it is useful over the entire pH range being tested; however, it was not soluble at acetonitrile concentrations greater than 50%.

Assaying the enzymatic activity of trypsin in solutions of varying pH also served as a control to validate the BAEE assay. Trypsin has maximal activity at pH 8 in water with no acetonitrile and deviations from this pH should result in reduced enzymatic activity. As seen in Figure 5.2A, in a purely aqueous environment the enzymatic activity had a maximum value at pH 8, validating the results of this experiment. Even the presence of acetonitrile did not appear to have any impact on the optimal pH for tryptic activity. Enzymatic activity did improve with acetonitrile concentrations around 20 to 30% at pH 8 with a subsequent drop after 30%. Assays at a higher pH of 9 showed a decrease in the range of maximal activity to 20% ACN before the drop in activity while a lower pH value of 7 showed a shift to higher ACN concentrations with maximal activity observed between 30 and 40% ACN.

The examination of tryptic activity over varying temperature and acetonitrile concentration, as shown in Figure 5.2B, revealed that the enzymatic activity of trypsin in a purely aqueous environment improved with increasing temperature as expected. The addition of acetonitrile, however, had a profound effect. At 37^{o}C the activity of trypsin decreased quickly past 30% acetonitrile. At temperatures below 30^{o}C, the enzymatic

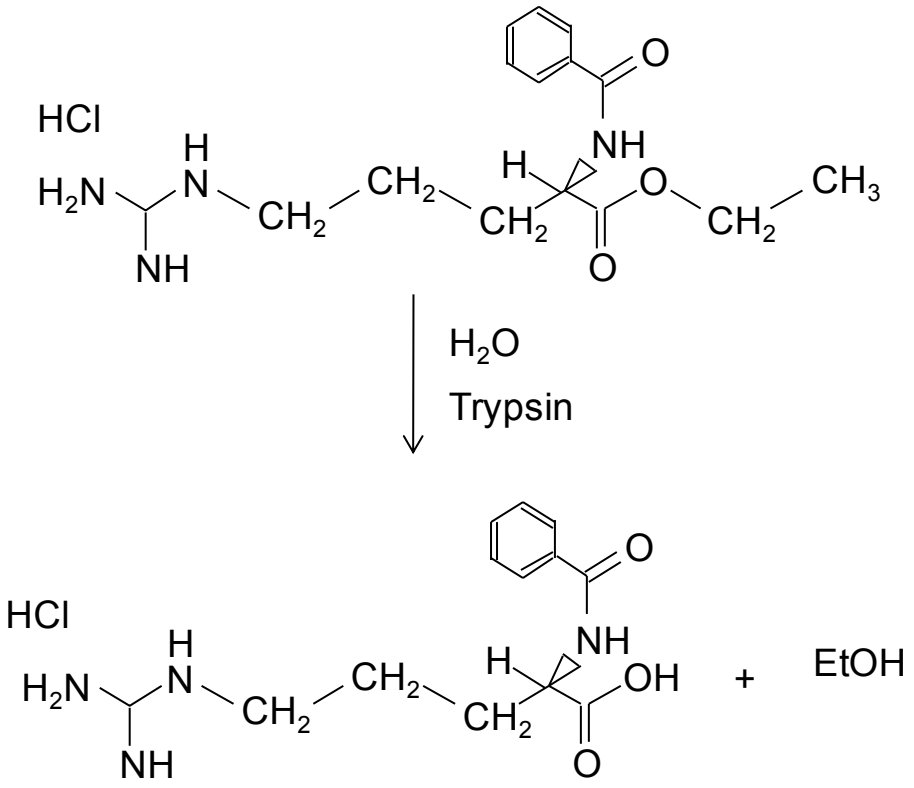

Figure 5.1 Chemical reaction of the BAEE assay. Nα-benzoyl-L-arginine ethyl ester is hydrolyzed by trypsin forming the product Nα-benzoyl-L_arginine and ethanol. The change in absorbance of the solution following addition of trypsin is monitored at a wavelength of 253 nm. The rate of change in absorbance is related to the enzymatic activity of trypsin in the solution.

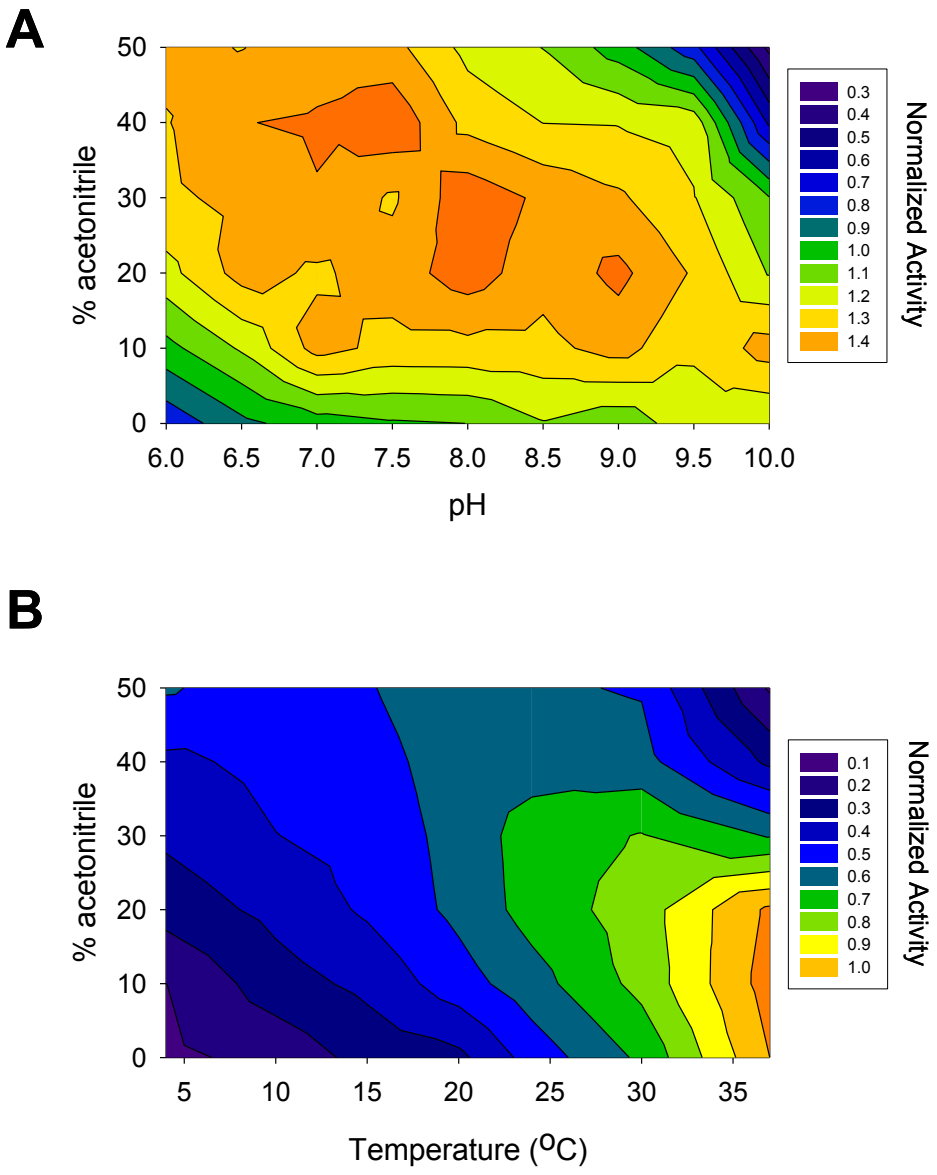

Figure 5.2 Enzymatic activity of trypsin under various conditions. Activity was normalized to that obtained in pure aqueous at pH 8.0, at different %ACN with varying (A) pH at 37°C and (B) temperature at pH 8. The addition of ACN at varying pH values did not appear to have a drastic effect upon the activity of trypsin, although activity did decrease sooner at the extreme pH values than it did at pH 8. Maximal tryptic activity was observed at an acetonitrile concentration between 20 – 30% and a temperature of 37°C. With increasing amounts of ACN at this temperature the activity quickly drops. However, the addition of acetonitrile at low temperatures appear to improve the activity.

activity of trypsin actually improved at 30% or greater concentrations of ACN in comparison to 37^{o}C, with activity decreasing gradually below 20^{o}C. This trend seems to correlate with improved sequence coverage of proteins digested in higher concentrations of organic solvent at room temperature rather than the typical 37^{o}C. However, it remains to be explained why it is that the sequence coverage of proteins improve following digestion in 80% ACN.

The enzymatic activity of trypsin was examined at higher concentrations of acetonitrile. Due to the insolubility of phosphate buffer in solutions composed of 50% ACN or greater, Tris-HCl buffer was chosen to permit assessment of enzyme activity at up to 80% acetonitrile, but at a constant pH of 8. Using Tris-HCl as the buffering agent, the tryptic activity was assessed at a constant pH of 8 and temperature of 24^{o}C over varying percentages of acetonitrile. The results of this experiment are shown in Figure 5.3 Examining this plot the greatest activity value for trypsin was obtained at an acetonitrile concentration of 20%. At acetonitrile concentrations greater than 20%, the activity continued to decrease. While 80% acetonitrile provided the greatest sequence coverage for all protein standards tested with LC-MS/MS, the activity of trypsin was in fact the lowest. The results from this experiment clearly contradict the explanation that observed improvements in sequence coverage of proteins at high concentration of acetonitrile are due to improved enzymatic activity of trypsin. The actual explanation for this observation must therefore reside in the types of peptides which are generated under differing digestion conditions. To test this hypothesis, an examination of the peptides generated from the tryptic digests under altering solvent compositions was performed.

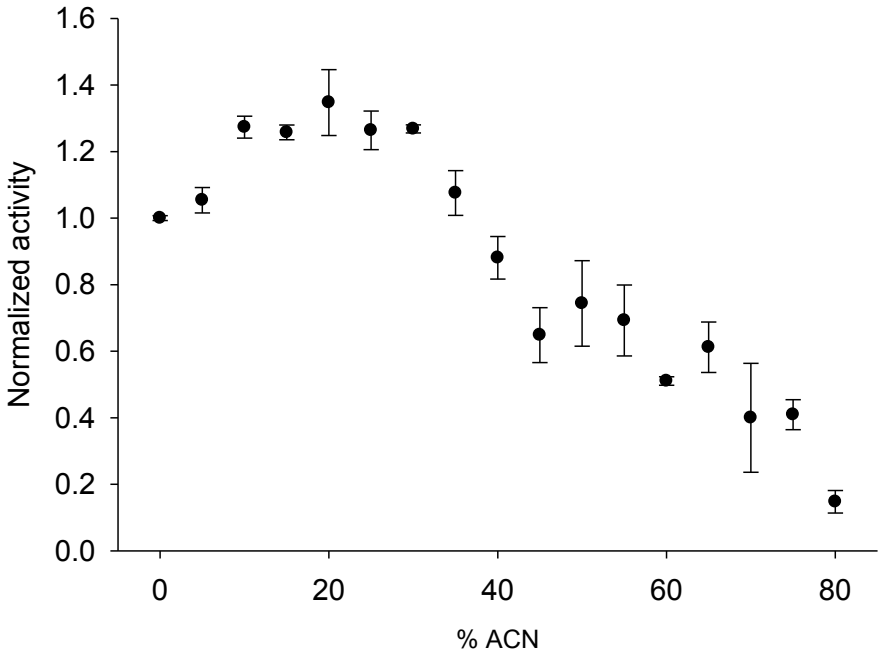

Figure 5.3 Tryptic activity from 0% to 80% ACN. Activity was normalized to 0% ACN, at 37°C, pH 8. As was already observed, and presented in Figure 5.2A, maximal activity was seen at an ACN concentration of 20%. Above this, a steady decrease in activity was observed with lowest activity at 80% ACN, with only 10% of the activity remaining as compared to pure aqueous.

5.3.3 Qualitative Examination using RPLC-UV and SDS-PAGE

One possible argument for the apparent disparity between the observed enzymatic activity of trypsin at a high concentration of acetonitrile and improved protein sequence coverage is that the BAEE assay, being based on a substrate mimic not an actual protein, does not accurately reflect the true activity of trypsin when it is acting on a protein under these conditions. To determine if this argument is valid, a qualitative analysis of the peptides produced from the tryptic digestion of BSA at various concentrations of acetonitrile and temperature was performed using RPLC with UV detection and SDS-PAGE. The RPLC chromatograms, as well SDS-PAGE visualizations, are presented in Figure 5.4. Examining these images it is clear that the extent of digestion at increasing concentrations of acetonitrile becomes progressively decreased. This is noted through observation of high levels of undigested intact BSA remaining in the digestions at 80% acetonitrile. From the gel image of the 80% ACN digests at either 24^{o}C or 37^{o}C, a significant number of larger protein bands are visible with apparent molecular weight above 10 kDa. Compared to digests performed in lower concentrations of acetonitrile (<10%), it is clear that digestion efficiency is severely compromised at higher concentrations of ACN. This observation is also confirmed in the UV chromatograms, which show significant amounts of intact BSA and decreased absorbance of the eluting peptides. The results from this experiment confirm the decreased enzymatic activity for trypsin at high concentrations of ACN, even when the enzyme is acting on a protein substrate. To determine why inefficient protein digestion, which leaves the majority of the protein undigested, and generating larger peptide fragments, results in improved sequence coverage by MS, a more detailed examination of the peptide fragments was

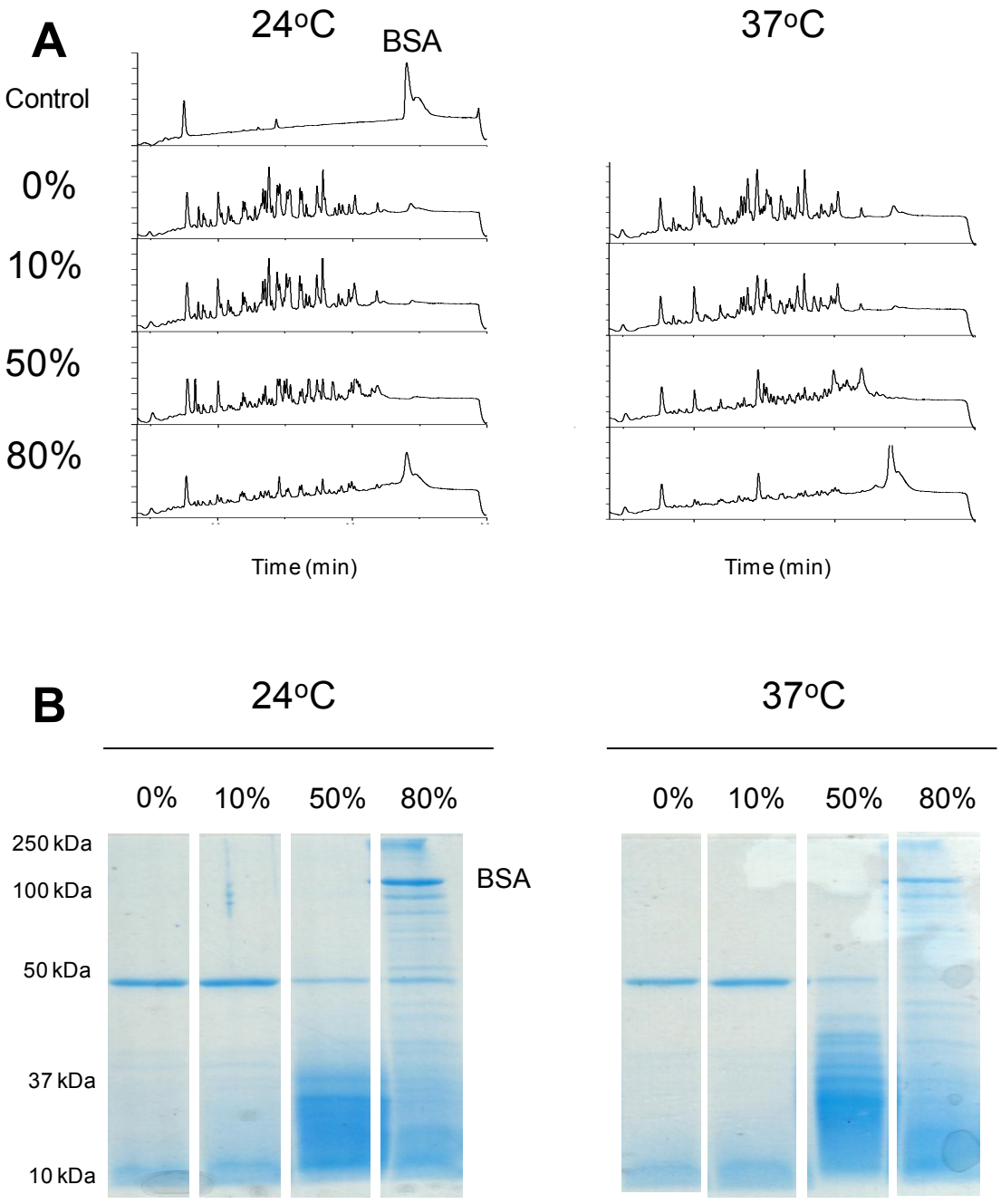

Figure 5.4 Results of the (A) RPLC-UV and (B) SDS-PAGE analysis of a BSA digest performed at 24°C and 37°C with 0, 10, 50, and 80 percent acetonitrile for 1 hour. Both of these techniques confirm that at higher percentages of acetonitrile the efficiency of the tryptic digest is very poor, which confirms the result from the BAEE assay, evident by the reduced number of peptide signals in the chromatograms, greater number of heavy peptides as seen in the SDS-PAGE images, as well as evidence of remaining BSA in both.

performed using LC-MS/MS.

5.3.4 Peptide Sequencing and Relative Quantitation

The contradicting results of poor digestion efficiency and improved sequence coverage may be resolved through a more detailed analysis of the peptides which were identified by the MS analysis. Shown in Table 5.3 are the list of peptides and the number of missed cleavage sites present of those that were identified in the aqueous digest at 37^{o}C along with those detected from the 80% ACN digest at room temperature using LC-MS/MS. Sequences with one or more missed cleavage site, which were not identified within peptides with no missed cleavages, are in red and underlined. The most important difference between these two lists is that the peptides identified from the organic digestion not only contain a greater number of peptides, but a greater number with missed cleavage sites. Interestingly, a number of peptides containing missed cleavages from the high organic solvent digestion were not identified as corresponding fully tryptic peptides in the purely aqueous digestion. For example, the peptide CCAADDKEACFAVEGPK, identified in the organic solvent digestion, contains one missed cleavage. A portion of its sequence, EACFAVEGPK, was detected alone in both the aqueous and organic digestions following LC-MS/MS. However, the remaining portion of the peptide, the sequence CCAADDK, was not detected as an individual peptide in either sample. From these results, it was deduced that improved sequence coverage for proteins digested by trypsin in the presence of high concentrations of acetonitrile may be due to the formation of a greater number of peptides with missed cleavage sites as compared to the same digestion performed in water. These peptides, with missed cleavage sites, may contain unique amino acid sequences from the parent protein that would otherwise go undetected

Table 5.3 Peptide sequence and number of missed cleavage sites.

Aqueous at 37°C		80% ACN at 24°C	
MC	Sequence	MC	Sequence
0	ALKAWSVAR	0	AEFVEVTK
0	KQTALVELLK	0	YLYEIAR
0	ECCDKPLLEK	0	DLGEEHFK
0	HLVDEPQNLIK	0	SHCIAEVEK
0	YICDNQDTISSK	0	ECCDKPLLEK
0	SLHTLFGDELCK	0	EACFAVEGPK
1	CASIQKFGER	0	LVNELTEFAK
1	LKECCDKPLLEK	0	SLHTLFGDELCK
1	LCVLHEKTPVSEK	0	HPEYAVSVLLR
1	VPQVSTPTLVEKVSR	0	YICDNQDTISSK
1	YICDNQDTISSKLK	0	TCVADESHAGCEK
1	CCAADDKEACFAVEGPK	0	ETYGDMADCCEK
2	LCVLHEKTPVSEKVTK	0	EYEATLEECCAK
		0	VPQVSTPTLVEVSR
		0	YNGVFQECCQAEDK
		0	RPCFSALTPDETYVPK
		0	DAIPENLPPLTADFAEDK
		1	KQTALVELLK
		1	LCVLHEKTPVSEK
		1	DDPHACYSTVFDKLK
		1	LAKEYEATLEECCAK
		1	CCAADDKEACFAVEGPK
		1	VHKECCHGDLLECADDR
		1	DAIPENLPPLTADFAEDKDVCK
		1	AFDEKLFTFHADICTLPDTEK
		2	LKECCDKPLLEK
		2	LSQKFPKAEFVEVTK
		2	KVASLRETYGDMADCCEK
		2	QEPERNECFLSHKDDSPDLPK
		2	YICDNQDTISSKLKECCDKPLLEK
		3	RMPCTEDYLSLILNRLCVLHEKTPVSEK

MC = number of missed cleavage sites

in MS analysis if digestion had gone to completion.

To test this hypothesis, a comparison of the types of peptides generated from the digestion of BSA in aqueous and 80% acetonitrile solutions was performed using relative quantitation through isotopic labeling. Each digestion was replicated five times and peptides were only considered for quantitation if detected in at least three of the five samples. Table 5.4 summarizes the results from this experiment. For the 5 minute digestion, a significantly greater concentration of peptides was obtained from the aqueous digestion relative to the organic digest. Even peptides with missed cleavages sites were present in greater amounts in comparison to the organic digestion. This is attributed to the greater yield from the aqueous digestion is such as short digestion period. Many of the peptides produced in the 5 minute digestion from the organic digestion were likely too large to be detected using LC-MS/MS on the linear ion trap. However, there were some exceptions, such as NECFLSHKDDSPDLPK, which was found in the 5 minute organic digestion at a greater than 5 fold concentration to that the aqueous digest. In contrast, it was found in the 1 hour digestion that average the number of peptides with zero missed cleavage sites was double to that found in the organic digest while the number of peptides with one missed cleavage site was found to be double in the organic, and 5 times the number of peptides with 2 missed cleavage sites was found in the organic digestion in comparison to the aqueous digestion. Still, peptides with zero missed cleavage sites were found at a greater concentration in the aqueous compared to the organic, though the concentration of these peptides in the organic digest was greater than at a rapid 5 min digest. This would seem to rule out the use of a 5 minute organic digestion as an optimal method for generating suitable peptide products for LC-MS/MS; 1 hour organic digestion

Table 5.4 Relative quantitation of BSA peptides.

# missed Cleavages	5 min. A/O ± STD	1 hour A/O ± STD	Overnight A/O ± STD
0	9 ± 4	2.1 ± 0.7	1.3 ± 0.4
1	10 ± 6	0.5 ± 0.2	0.7 ± 0.2
2	16 ± 12	0.16 ± 0.06	0.7 ± 0.6
All peptides	11 ± 6	1.3 ± 0.5	1.1 ± 0.4

A = Aqueous 37°C
O = 80% ACN 24°C

generated more suitable peptide products. Although several peptides unique to the organic digestion were found, the overall yield for many of the other peptides was very low compared to a standard aqueous digestion.

The results from the overnight digestion show that the ratio of peptides with zero, one, and two missed cleavage sites had shifted closer to 1 between the organic and aqueous digestion, indicating that the sample composition had become more similar in comparison to the 5 minute and 1 hour digestions. With a purpose of maximizing the generation and detection of complementary peptide sequences, as well as maximizing yield, this would indicate that a 1 hour digestion in 80% ACN would be the most useful for enhancing protein sequence coverage compared to the 5 minute or the overnight digestion.

This is further highlighted from the sequence coverage obtained of the 5 minute, 1 hour, and overnight digestion in organic shown in Table 5.5. Average sequence coverage from the organic digestion of 43% was obtained from the 5 minute digestion, 63% from the 1 hour, and 58% from the overnight, in comparison to the aqueous digestion which had a nearly consistent 60% sequence coverage across all three time points. Combining the peptides from all replicate samples, the multiconsensous (aggregate data) results show that the organic digestion contains a greater number of unique BSA peptides compared to the aqueous digestion, with the 1 hour providing the greatest sequence coverage of 76% compared to the greatest sequence coverage obtained from the aqueous digestion performed overnight with 63%.

The use of a 1 hour organic digestion seems to be useful for providing unique peptides which are not observed in a typical overnight aqueous digestion. Assuming that

Table 5.5 Sequence coverage of BSA from five replicate digestions.

	Aqueous 37°C		80% ACN 24°C	
Time	Sequence coverage (%)	Multi[A]	Sequence coverage (%)	Multi[A]
5 min	60 ± 4	71	43 ± 10	71
1 hour	59 ± 3	65	63 ± 7	76
overnight	62 ± 6	63	58 ± 13	72

[A]multiconsensus results: aggregate % sequence coverage using all peptides identified from all replicates

these results are not unique to the BSA protein, one may extrapolate the method to enhance proteome coverage (i.e., digestion and analysis of complex protein mixtures). The application of a 1 hour organic digestion, combined with overnight aqueous digestion was conducted for analysis of a yeast proteome.

5.3.5 Application for the Improved Coverage of a Yeast Proteome

From the quantitation data, a 1 hour room temperature digestion in 80% acetonitrile is optimal for producing suitable peptide fragments which lead to greater protein sequence coverage. However, many complementary peptides can be generated in high yield from a purely aqueous digests. Thus, a combination of these two different protocols should provide increased sequence coverage and higher identifications for proteins from a complex proteome mixture.

A yeast proteome extract was split into two equal portions and subjected to digestion using two protocols: aqueous at 37^{o}C overnight and 80% ACN at room temperature for 1 hour. The resulting peptide mixtures were independently separated by cation exchange chromatography into 9 fractions which were then subject to duplicate analysis using LC-MS/MS. A Venn diagram comparing the identified proteins and corresponding peptides between the organic and aqueous digests and their respective replicate runs are presented in Figure 5.5. The use of replicate analysis of a complex proteome mixture results in improved proteome coverage for both the organic and aqueous digested samples. The number of proteins identified from the aqueous digestion increased from 692 to 751, nearly an increase of 10%. For the proteins identified from replicate analysis of the organic digest, an increase from 536 to 616, about a 15% increase, was observed. In terms of peptides, an approximate increase of 20% of the total

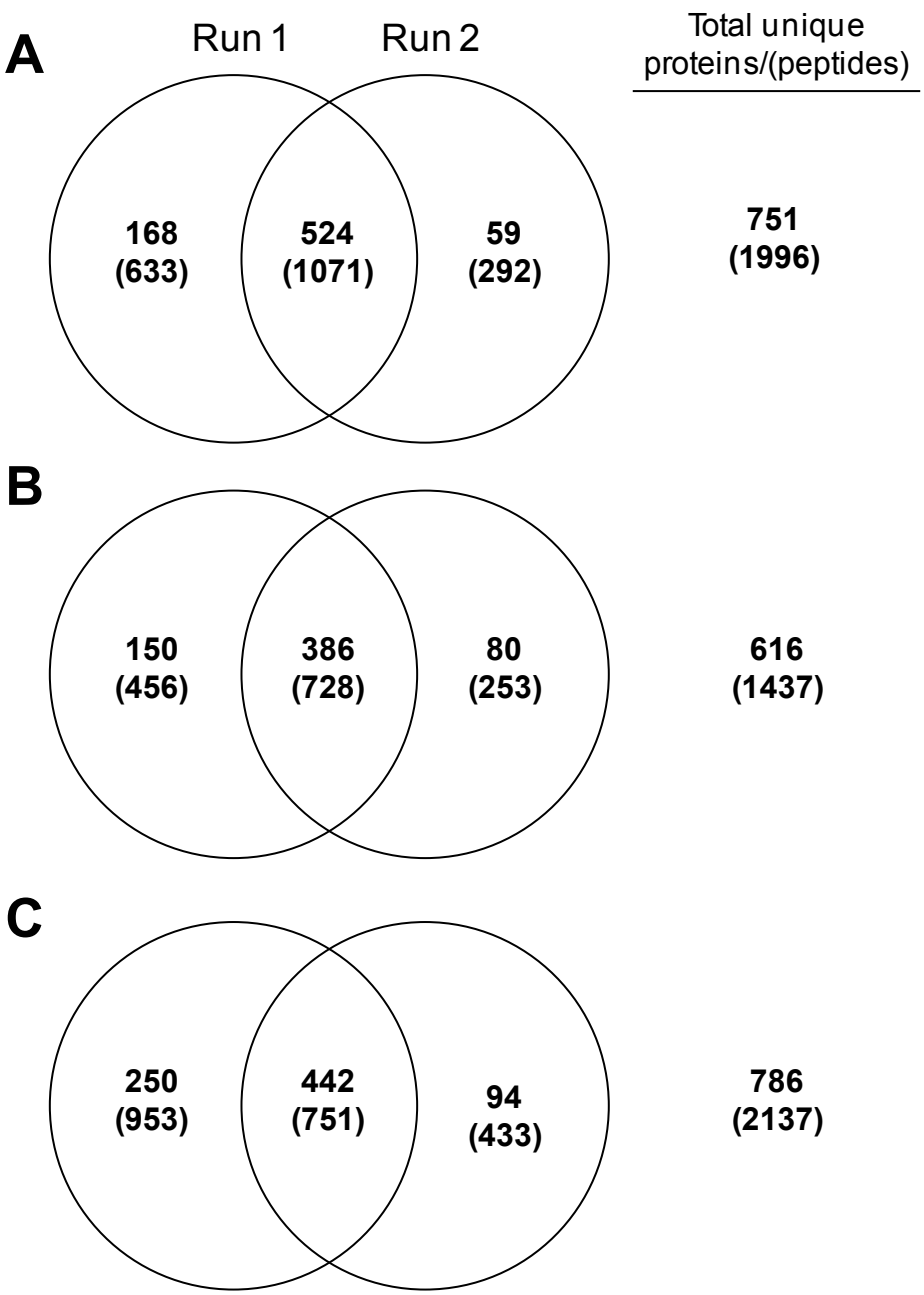

Figure 5.5 Venn diagrams of the yeast proteins and peptides identified from the aqueous and organic digestions following SCX and LC-MS/MS analysis. Comparisons are made between (A) replicate analyses of the aqueous digest, (B) replicate runs of the organic digest, and (C) combined organic and aqueous digest. Replicate analysis of the peptide fractions from SCX of the organic and aqueous yeast proteome digests improved the total number of unique proteins and peptides identified. As independent protocols, the aqueous digest is superior to that of the organic digestion for this complex sample. However, as shown in C, combining the two digestion protocols resulted in the identification of a greater number of proteins than possible from either of the protocols alone.

number of identified peptides was observed. This would indicate a slightly reduced bias in the types of proteins identified from the organic digestion. The greater number of peptide identifications obtained from aqueous compared to organic digestion is attributed to the non-complete digestion which occurs at 80% ACN, which was observed in the RPLC-UV and SDS-PAGE experiments. This would affect the total peptide loading onto the IEC and LC-MS/MS which was not taken into account. This may be overcome by further enriching the peptide mixture, as through the use of a low molecular weight cutoff filter to eliminate larger peptide fragments, and undigested (intact) proteins. This would ensure appropriate loading of low molecular weight peptides (<3 kDa) which are suitable for LC-MS/MS analysis.

Despite the reduced peptide yield from the organic digest, it is shown in Figure 5.5C that a greater number of unique proteins and peptides can be obtained if the second LC-MS/MS run involves a combination of the two digestion protocols. In this instance, an additional 35 unique proteins, and 141 unique peptides, were identified when results from one run from each of the aqueous and organic digestions is combined, compared to a replicate analysis of a single digestion protocol. This is attributed to greater differences in the peptide composition generated between an aqueous and organic digestion.

This was examined by determining the percentage of peptides identified from the aqueous and organic digestions which contain 0, 1, 2, or 3 missed cleavage sites. These percentages are given in Table 5.6. The percentage of peptides which contain no missed cleavage site was found to be greater in the aqueous digestion. The percentages of peptides identified which did have at least one missed cleavage were found to be greater from the yeast sample digested in 80% acetonitrile. This supports the observations made

Table 5.6 Percentage of yeast peptides with missed cleavage sites.

# missed cleavage sites	% of identified peptides	
	Aqueous (Overnight 37°C)	Organic (1 h 24°C)
0	69.0	63.2
1	26.9	30.4
2	3.8	6.0
3	0.3	0.4

previously which determined that the digestion efficiency of trypsin in this solvent composition is lower than in purely aqueous solvent. Although the digestion efficiency is low, it does provide a greater number of uniquely identified peptides when the results were pooled with that from an aqueous digestion than just from replicate injections of the sample samples.

5.4 Conclusions

Presented in this work is an investigation into the underlying cause of enhanced sequence coverage following bottom-up LC-MS/MS analysis for tryptic digestion of protein in the presence of 80% acetonitrile. While improved coverage was confirmed with a set of protein standards, a BAEE assay to determine tryptic activity at various temperatures, pH, and organic composition showed that enzymatic activity was hampered in high concentrations of organic solvent. Further investigation of BSA digestion products showed consistently incomplete digestion products, including peptides which contained missed cleavage sites and residual undigested protein. However, these peptides with missed cleavage sites permit an opportunity to detect a greater portion of the standard proteins when analyzed by mass spectrometry, thus leading to increased sequence coverage. The application of this to a complex yeast proteome showed that a combination of aqueous and organic solvent digestion techniques followed by two-dimensional liquid chromatography coupled to bottom-up MS/MS analysis improved the overall number of unique peptides identified. It is noted, however, that although digestions in high concentrations of an organic solvent does lead to higher sequence coverage due to peptides with missed cleavages, a portion of protein in the sample likely

remains undigested and results in reduced sample loading for further downstream analysis. However, this may be avoided if the workflow takes this into consideration and a greater volume of the digested mixture is loaded for LC-MS/MS analysis to counteract this issue.

Chapter 6

Conclusions

6.1 Thesis Summary

This thesis describes work towards the development of a high-throughput proteomics workflow for early stage protein biomarker discovery. In Chapter 1, the goals and difficulties of current MS-based proteomics analysis workflows wre discussed. It presented and highlighted the disadvantages inherent to two common proteomics analysis workflows: comprehensive bottom-up peptide analysis, such as through multidimensional protein identification technology (MudPIT) approach, and two-dimensional gel electrophoresis (2D-PAGE) approach. The essential disadvantage of MudPIT style workflows are that comprehensive analysis generates considerable amounts of data, but potentially misses the components which are of actual of interest. While prefractionation can improve this, it comes at the expense of lower throughput. In contrast, the 2D-PAGE approach, while possibly quicker in the sense that the proteins of interest are targeted from differential staining, does not take advantage of the sensitivity of MS-based detection. The need for high-throughput analysis which combines the best features of these workflows is required. The development of this workflow includes the need to evaluate and develop new effective techniques to comprise the proteomics toolbox.

In chapter 2, the evaluation of ion exchange chromatography (IEC) was conducted as a comprehensive proteome prefractionation tool prior to bottom-up peptide identification. It was found that IEC has inherent bias towards components of the sample which make it unsuited for proteome prefractionation, illustrating that “more” separation

is not necessarily “better” separation. Additionally, it was demonstrated that prefractionation of samples containing high concentrations of contaminating protein greatly improves MS-based peptide identification. Also, it was shown that the choice of separation method can have a detrimental impact upon the composition of your sample, as illustrated by the loss of high mass proteins through porous reversed chromatography. As a general conclusion, all separation platforms used in a comprehensive proteomics analysis workflow should be critically evaluated in terms of separation and protein recovery. The use of well characterized protein standards, as opposed to unknown protein mixtures, permits this critically important evaluation.

Chapter 3 presented the development of an 8-column multiplexed reversed phase liquid chromatography (RPLC) separation system which can couple to matrix-assisted laser desorption/ionization (MALDI) mass spectrometry through a high-volume well plate adaptor. This multiplexed RPLC system, through the use of restriction capillaries, displayed excellent equality in flow rates between all columns of the system, as well as minimal cross-loading onto the remaining columns during sample loading. The MALDI well plate adaptor, constructed of Teflon and sealed with a layer of polydimethylsiloxane (PDMS), allowed the collection of up to 100 µL onto a MALDI target with no observed leaking between wells and the retention of 75% of signal intensity compared to direct deposition. This system greatly improves the throughput of separations compared to conventional serial separation of fractions generated from a first dimension of fractionation. The platform also permits automated LC-MALDI, by greatly reducing the number of manual steps involved, highlighting MALDI MS as a viable alternative to LC-ESI-MS.

Employing the multiplex LC-MALDI system developed in chapter 3, chapter 4 presented an application of MALDI MS on the vMALDI LTQ as a high-throughput analysis strategy for fast quantitative proteome profiling. Although it was shown that MALDI MS/MS analysis using this instrument could generate comparable, or perhaps even greater, amounts of peptide identifications compared to an ESI MS/MS approach, the time required to achieve these results was simply too great to be part of a high-throughput analysis strategy. Instead, MALDI MS was used as a profiling tool alone. Following differential isotopic labeling of yeast and the addition of protein standards as test "biomarkers", this strategy was able to identify low amounts of candidate peptides, displaying differential levels of expression, amongst a complex background of undifferentiated peptides. A two-dimensional liquid chromatographic separation approach, takes advantage of the multiplexed LC-MALDI system to further improve the throughput of the workflow. This approach couples the advantages of liquid phase separations and MS sensitivity, while retaining the core principle of targeted analysis through differential mapping, inherent to 2D-PAGE. It is expected, with further development, that this system will become a valuable addition to the proteomics toolbox for preliminary discovery of candidate protein biomarkers.

Finally, in chapter 5, the investigation of the effects which high concentrations of acetonitrile (ACN) have upon the proteolytic activity of trypsin was explored. The use of organic solvents to enhance MS-based protein sequence coverage is well known. However, explanations for why these solvent systems enhance protein sequence coverage following MS analysis were conflicting, given that a standard trypsin activity assays and qualitative examination of the peptide products from a BSA digest under these conditions

showed that the digestion quality was very poor. A quantitative examination using differential isotopic labeling revealed that the presence of a high concentration of ACN caused the formation of a greater number of peptides with missed cleavage sites. These larger peptide fragments produced under these conditions often contained unique amino acid sequences which were not detected from digests conducted in lower concentrations of ACN. This knowledge leads to the development of a new approach to sample preparation prior to bottom-up MS analysis. The strategy takes advantage of the greater differences in sample composition following digestion in aqueous and 80% ACN which was showed to improve the total number of peptide and protein identifications from a complex yeast sample.

6.2 Future Work

The main purpose of this thesis was to develop a high-throughput proteomics profiling workflow for the early stage discovery of potential protein biomarkers. While the application of real samples using this system was beyond the scope of this thesis, future work using this system may provide accurate and timely results on the proteomics state of various samples. Not only can this system be used for the direct comparison of proteomes under various disease states, its use for the time-resolved analysis of proteomes in a time dimension is also possible. In this thesis, the manual analysis of data collected from the LC-MALDI MS platform was carried out, with full knowledge of the peptides which were of interest. In contrast, the analysis of real samples will require the development of computer software to identify potential peaks of interest following data collection. Work in this area is currently underway.

In terms of applications of this system, a potential collaboration with research in biology has presented interest in the analysis of brine shrimp development and reproduction based on a time-resolved analysis. Due to the large number of samples and time points that would be required for analysis, this high-throughput system is ideal for discovering the most important protein features. Also, due to its high-throughput, the analysis of the change in a proteome make up over time could also be matched to DNA transcription information.

While the work presented here focused on a peptide level separation scheme, ultimately the use of protein prefractionation techniques will be incorporated into the system. The multiplexed LC system is ideal for increasing the throughput of separation following an initial stage of protein prefractionation or, alternatively, as a high-throughput peptide cleanup system. As IEC was shown to be an inadequate tool for comprehensive proteome prefractionation, other tools developed in this lab, such as a fast and easy to use solution isoelectric focusing (sIEF) device, as well as a gel-based molecular weight separation device which allows for collection of samples in solution named GELFrEE (gel eluted liquid fractionation entrapment electrophoresis), will be incorporated into the high-throughput analysis system. The combination of sIEF with GELFrEE is a solution phase equivalent to 2D-PAGE. The use of this strategy prior to the multiplexed LC-MALDI MS analysis would delegate the RPLC separation to a high-throughput cleanup method, with optional peptide level separations using reversed phase during collection onto the MALDI well plate. However, this strategy would require the development of a protein level isotopic labeling method. A strategy for this has recently been developed in this lab, and could be applied to further the work presented in this

thesis.

Additional work for improving this system will lay with sample preparation. The dynamic range in concentration of the proteome makes the analysis of low abundant proteins of interest difficult to analyze. Increasing the loading of the sample does little to help as it also increases the presence of contaminating components. Without knowledge of what it is which may be of interest, purification and enrichment through affinity methods are not feasible. Currently, the use of combinatorial affinity columns which bind all components of a proteome, resulting in the compression of the dynamic range, may be the best approach to resolving this issue. This would also allow for the refined control over the amount of sample which is applied to the sample plate for MALDI MS profiling, as the ratio of sample to matrix has a large effect upon sensitivity. Protein level isotopic labeling prior to this preparation would also retain the relative quantitation information of the sample, allowing this high-throughput profiling method to be applied.

6.3 Conclusions

The proteome is very complex, and tools which allow for the comprehensive analysis of a proteome are still urgently needed. The clinical application of proteomics has enormous potential, and will be realized as new tools become available to improve proteome characterization. Presented in this thesis is work towards the development of an alternative proteomics workflow which focuses on high-throughput protein biomarker discovery. While work remains to tackle issues of dynamic range and sample preparation strategies, it is anticipated that, in the near future, the analysis of complete proteomes for routine biomarker discovery will be possible, facilitate through technologies presented

here. This will allow for the creation of improved methods for disease diagnosis as well as personalized treatment plans to improve human health.

References

[1] Vlahou, A., Fountoulakis, A., Proteomic approaches in the search for disease biomarkers. J. Chromatogr. B 2005, *814*, 11-19.

[2] LaBaer, J., So, you want to look for biomarkers - (Introduction to the special biornarkers issue). J. Proteome Res. 2005, *4*, 1053-1059.

[3] Rifai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat. Biotechnol. 2006, *24*, 971-983.

[4] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. Nature 2003, *422*, 198-207.

[5] Mann, M., Hendrickson, R. C., Pandey, A., Analysis of proteins and proteomes by mass spectrometry. Annu. Rev. Biochem. 2001, *70*, 437-473.

[6] Righetti, P. G., Castagna, A., Herbert, B., Prefractionation techniques in proteome analysis. Anal. Chem. 2001, *73*, 320A-326A.

[7] Lescuyer, P., Hochstrasser, D., Rabilloud, T., How shall we use the proteomics toolbox for biomarker discovery? J. Proteome Res. 2007, *6*, 3371-3376.

[8] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H.O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. Q. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J. H., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, C., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng,Z. M., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W. M., Gong, F. C., Gu, Z. P., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z. X., Ketchum, K. A., Lai, Z. W., Lei, Y. D., Li, Z. Y., Li, J. Y., Liang, Y., Lin, X. Y., Lu, F., Merkulov, G. V., Milshina, N.,

Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B. X., Sun, J. T., Wang, Z. Y., Wang, A. H., Wang, X., Wang, J., Wei, M. H., Wides, R., Xiao, C. L., Yan, C. H., Yao, A., Ye, J., Zhan, M., Zhang, W. Q., Zhang, H. Y., Zhao, Q., Zheng, L. S., Zhong, F., Zhong, W. Y., Zhu, S. P. C., Zhao, S. Y., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H. J., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H. Y., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky,A.; Lewis, M., Liu, X. J., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M. Y., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X. H, The sequence of the human genome. Science 2001, *291*, 1304-1351.

[9] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs,R. A., Muzny,D. M., Scherer,S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H. M., Yu, J., Wang, J., Huang, G. Y., Gu, J., Hood,,L., Rowen,,L., Madan, A., Qin, S. Z., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H. Q., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P.,

Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G. R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W. H., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J. R., Slater, G., Smit, A. F. A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., Initial sequencing and analysis of the human genome. Nature 2001, *409*, 860-921.

[10] Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L. J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S. G., Nielsen, C. B., Butler, J., Endrizzi, M., Qui,D. Y., Ianakiev, P., Pedersen, D. B., Nelson, M. A., Werner-Washburne, M., Selitrennikoff, C. P., Kinsey, J. A., Braun, E. L., Zelter, A., Schulte, U., Kothe, G. O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stabge-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R. L., Perkins, D. D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W. X. Pratt, R. J., Osmani, S. A., DeSouza, C. P. C., Glass, L., Orbach, M. J., Berglund, J. A., Voelker, R., Yarden, O., Plamann, M., Seller, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D. O., Alex, L. A., Mannhaupt, G., Ebbole, D. J., Freitag, M., Paulsen, I., Sachs, M. S., Lander, E. S., Nusbaum, C., Birren, B., The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 2003, *422*, 859-868.

[11] Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M. C., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z. W., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de

Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chatuverdi, K., Christophides, G. K., Chrystal, M. A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z. P., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z. X., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J. J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H. G., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J. T., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z. J., Unger, M. F., Walenz, B., Wang, A. H., Wang, J., Wang, M., Wang, X. L., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E. M., Zhang, H. Y., Zhao, Q., Zhao, S. Y., Zhu, S. P. C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C. W., Louis, C., Kalush, F., Mural, R. J., Myers, E. W., Adams, M. D., Smith, H. O., Broder, S., Gardner, M. J., Fraser, C. M., Birney, E., Bork, P., Brey, P. T., Venter, J. C., Weissenbach, J., Kafatos, F. C., Collins, F. H., Hoffman, S. L., The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 2002, *298*, 129-149.

[12] Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green ,E. D., Gregory,

S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., ,Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Larty, E. O., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Strange-Thomann, N. S., Sugnet,C ., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Vidal, A. U., Vinson, J. P., von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C., Lander, E. S., Initial sequencing and comparative analysis of the mouse genome. Nature 2002, *420*, 520-562.

[13] Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble,A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S.,

McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, L., Jimenez,,J., Sanchez,,M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerrutti, L., Lowe, T., McCombie, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Ussery, D., Barrell, B. G., Nurse, P., The genome sequence of *Schizosaccharomyces pombe*. Nature 2002, *415*, 871-880.

[14] Anderson, N. L., Anderson, N. G., The human plasma proteome - History, character, and diagnostic prospects. Mol. Cell. Proteomics 2002, *1*, 845-867.

[15] Abbott, A., A post-genomic challenge: learning to read patterns of protein synthesis. Nature 1999, *402*, 715-720.

[16] Graveley, B. R., Alternative splicing: increasing diversity in the proteomic world. Trends in Genet. 2001, *17*, 100-107.

[17] Mann, M., Jensen, O. N., Proteomic analysis of post-translational modifications. Nat. Biotechnol. 2003, *21*, 255-261.

[18] Jensen, O. N., Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr. Opin. Chem. Biol. 2004, *8*, 33-41.

[19] Santoni, V., Molloy, M., Rabilloud, T., Membrane proteins and proteomics: Un amour impossible? Electrophoresis 2000, *21*, 1054-1070.

[20] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan,

J. X., Gooley, A. A., Hughes, G., HumpherySmith, I., Williams, K. L., Hochstrasser, D. F., From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Bio-Technology 1996, *14*, 61-65.

[21] James, P., Protein identification in the post-genome era: the rapid rise of proteomics. Q. Rev. Biophys. 1997, *30*, 279-331.

[22] Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., Eisenberg, D., Detecting protein function and protein-protein interactions from genome sequences. Science 1999, *285*, 751-753.

[23] Schwikowski, B., Uetz, P., Fields, S., A network of protein-protein interactions in yeast. Nat. Biotechnol. 2000, *18*, 1257-1261.

[24] von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., Bork, P., Comparative assessment of large-scale data sets of protein-protein interactions. Nature 2002, *417*, 399-403.

[25] Chou, K. C., Progress in protein structural class prediction and its impact to bioinformatics and proteomics. Curr. Protein Peptide Sci. 2005, *6*, 423-436.

[26] Galvao-Botton, L. M. P., Katsuyama, A. M., Guzzo, C. R., Almeida, F. C. L., Farah, C. S., Valente, A. P, High-throughput screening of structural proteomics targets using NMR. FEBS Lett. 2003, *552*, 207-213.

[27] Liu, H. L., Hsu, J. P., Recent developments in structural proteomics for protein structure determination. Proteomics 2005, *5*, 2056-2068.

[28] Rehm, T., Huber, R., Holak, T. A., Application of NMR in structural proteomics: Screening for proteins amenable to structural analysis. Structure 2002, *10*, 1613-1618.

[29] Shin, J., Lee, W., Lee, W., Structural proteomics by NMR spectroscopy. Expert Rev. Proteomics 2008, *5*, 589-601.

[30] Szyperski, T., Atreya, H., du Penhoat, C., Eletsky, A., Liu, G., Parish, D., Shao, Y., Sheng, Y., Sukumaran, D., Xu, D., Protocol for NMR-based structural proteomics. Mol. Cell. Proteomics 2005, *4*, S14-S14.

[31] Vinarov, D. A., Lytle, B. L., Peterson, F. C., Tyler, E. M., Volkman, B. F., Markley, J. L., Cell-free protein production and labeling protocol for NMR-based

structural proteomics. Nat. Methods 2004, *1*, 149-153.

[32] Yee, A. A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X. H., Skarina, T., Evdokimova, E., Liu, C. S., Semesi, A., Guido, V., Edwards, A. M., Arrowsmith, C. H., NMR and x-ray crystallography, complementary tools in structural proteomics of small proteins. J. Am. Chem. Soc. 2005, *127*, 16512-16517.

[33] Yokoyama, S., Protein expression systems for structural genomics and proteomics. Curr. Opin. Chem. Biol. 2003, *7*, 39-43.

[34] Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A., Robinson, C. V., Protein complexes in the gas phase: Technology for structural genomics and proteomics. Chem. Rev. 2007, *107*, 3544-3567.

[35] Downard, K. M., Ions of the interactome: The role of MS in the study of protein interactions in proteomics and structural biology. Proteomics 2006, *6*, 5374-5384.

[36] Guan, J. Q., Chance, M. R., Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. Trends Biochem. Sci. 2005, *30*, 583-592.

[37] Naylor, S., Kumar, R., Emerging role of mass spectrometry in structural and functional proteomics. Adv. Protein Chem. 2003, *65*, 217-248.

[38] Sihlbom, C., Davidsson, P., Nilsson, C. L., Prefractionation of cerebrospinal fluid to enhance glycoprotein concentration prior to structural determination with FT-ICR mass spectrometry. J. Proteome Res. 2005, *4*, 2294-2301.

[39] Meng, F. Y., Forbes, A. J., Miller, L. M., Kelleher, N. L., Detection and localization of protein modifications by high resolution tandem mass spectrometry. Mass Spectrom. Rev. 2005, *24*, 126-134.

[40] Temporini, C., Callerli, E., Massolini, G., Caccialanza, G., Integrated analytical strategies for the study of phosphorylation and glycosylation in proteins. Mass Spectrom. Rev. 2008, *27*, 207-236.

[41] Dai, J., Jin, W. H., Sheng, Q. H., Shieh, C. H., Wu, J. R., Zeng, R, Protein phosphorylation and expression profiling by Yin-Yang multidimensional liquid chromatography (Yin-Yang MDLC) mass spectrometry. J. Proteome Res. 2007, *6*, 250-262.

[42] Corthals, G. L., Aebersold, R., Goodlett, D. R., Identification of phosphorylation sites using microimmobilized metal affinity chromatography. Methods Enzymol. 2005, *405*, 66-81.

[43] Zhang, B., VerBerkmoes, N. C., Langston, M. A., Uberbacher, E., Hettich, R. L., Samatova, N. F., Detecting differential and correlated protein expression in label-free shotgun proteomics. J. Proteome Res. 2006, *5*, 2909-2918.

[44] Poon, H. F., Castegna, A., Farr, S. A., Thongboonkerd, V., Lynn, B. C., Banks, W. A., Morley, J. E., Klein, J. B., Butterfield, D. A., Quantitative proteomics analysis of specific protein expression and oxidative modification in aged senescence-accelerated-prone 8 mice brain. Neuroscience 2004, *126*, 915-926.

[45] Mazzanti, R., Solazzo, M., Fantappie, O., Elfering, S., Pantaleo, P., Bechi, P., Cianchi, F., Ettl, A., Giulivi, C., Differential expression proteomics of human colon cancer. Am. J. Physiol-gastr. L. 2006, *290*, G1329-G1338.

[46] Lesley, S. A., High-throughput proteomics: Protein expression and purification in the postgenomic world. Protein Expr. Purif. 2001, *22*, 159-164.

[47] Kennedy, S., The role of proteomics in toxicology: identification of biomarkers of toxicity by protein expression analysis. Biomarkers 2002, *7*, 269-290.

[48] Joo, W. A., Kang, M. J., Son, W. K., Lee, H. J., Lee, D. Y., Lee, E., Kim, C. W., Monitoring protein expression by proteomics: Human plasma exposed to benzene. Proteomics 2003, *3*, 2402-2411.

[49] Hanash, S. M., Madoz-Gurpide, J., Misek, D. E., Identification of novel targets for cancer therapy using expression proteomics. Leukemia 2002, *16*, 478-485.

[50] Freeman, W. M., Hemby, S. E., Proteomics for protein expression profiling in neuroscience. Neurochem. Res. 2004, *29*, 1065-1081.

[51] Foster, L. J., Zeemann, P. A., Li, C., Mann, M., Jensen, O. N., Kassem, M., Differential expression profiling of membrane proteins by quantitative proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation. Stem Cells 2005, *23*, 1367-1377.

[52] Evans, C. A., Tonge, R., Blinco, D., Pierce, A., Shaw, J., Lu, Y. N., Hamzah, H. G., Gray, A., Downes, C. P., Gaskell, S. J., Spooncer, E., Whetton, A. D., Comparative proteomics of primitive hematopoietic cell populations reveals differences in expression of proteins regulating motility. Blood 2004, *103*, 3751-3759.

[53] Bradley, B. P., Shrader, E. A., Kimmel, D. G., Meiller, J. C., Protein expression signatures: an application of proteomics. Mar. Environ. Res. 2002, *54*, 373-377.

[54] Alexander-Kaufman, K., James, G., Sheedy, D., Harper, C., Matsumoto, I., Differential protein expression in the prefrontal white matter of human alcoholics: a proteomics study. Mol. Psychiatry 2006, *11*, 56-65.

[55] Brunner, E., Ahrens, C. H., Mohanty, S., Baetschmann, H., Loevenich, S., Potthast, F., Deutsch, E. W., Panse, C., de Lichtenberg, U., Rinner, O., Lee, H., Pedrioli, P. G. A., Malmstrom, J., Koehler, K., Schrimpf, S., Krijgsveld, J., Kregenow, F., Heck, A. J. R., Hafen, E., Schlapbach, R., Aebersold, R., A high-quality catalog of the Drosophila melanogaster proteome. Nat. Biotechnol. 2007, *25*, 576-583.

[56] de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Froehlich, F., Walther, T. C., Mann, M., Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 2008, *455*, 1251-U60.

[57] Quilliam, M. A., Ross, N. W., Analysis of diarrhetic shellfish poisoning toxins and metabolites in plankton and shellfish by ion-spray liquid chromatography mass spectrometry. ACS Sym. Ser. 1996, *619*, 351-364.

[58] Rejtar, T., Hu, P., Juhasz, P., Campbell, J. M., Vestal, M. L., Preisler, J., Karger, B. L., Off-line coupling of high-resolution capillary electrophoresis to MALDI-TOF and TOF/TOF MS. J. Proteome Res. 2002, *1*, 171-179.

[59] Marshall, A. G., Hendrickson, C. L., Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrom. Rev. 1998, *17*, 1-35.

[60] Comisaro.Mb, Marshall, A. G., Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. Chemical Physics Letters 1974, *25*, 282-283.

[61] Makarov, A., Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. Anal. Chem. 2000, *72*, 1156-1162.

[62] Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M., Cooks, R. G., The Orbitrap: a new mass spectrometer. J. Mass Spectrom. 2005, *40*, 430-443.

[63] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. Science 1989, *246*, 64-71.

[64] Karas, M., Hillenkamp, F., Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. Anal. Chem. 1988, *60*, 2299-2301.

[65] Karas, M., Bachmann, D., Hillenkamp, F., Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules. Anal. Chem. 1985, *57*, 2935-2939.

[66] Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of flight Mass Spectrometry. Rapid Commun. Mass Spectrom. 1988, *2*, 151-153.

[67] Zhang, N., Doucette, A., Li, L., Two-layer sample preparation method for MALDI mass spectrometric analysis of protein and peptide samples containing sodium dodecyl sulfate. Anal. Chem. 2001, *73*, 2968-2975.

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

[69] Hattan, S. J., Vestal, M. L., Novel Three-Dimensional MALDI Plate for Interfacing High-Capacity LC Separations with MALDI-TOF. Anal. Chem. 2008, *80*, 9115-9123.

[70] Chen, H. S., Rejtar, T., Andreev, V., Moskovets, E., Karger, B. L., High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. Anal. Chem. 2005, *77*, 2323-2331.

[71] Roepstorff, P., Fohlman, J., Proposal for a Common Nomenclature for Sequence Ions in Mass-Spectra of Peptides. Biomed. Mass Spectrom. 1984, *11*, 601-601.

[72] Eng, J. K., Mccormack, A. L., Yates, J. R., An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. J. Am. Soc. Mass Spectrom. 1994, *5*, 976-989.

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

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

[75] Bogdanov, B., Smith, R. D., Proteomics by FTICR mass spectrometry: Top down and bottom up. Mass Spectrom. Rev. 2005, *24*, 168-200.

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

[77] Han, X., Jin, M., Breuker, K., McLafferty, F. W., Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. Science 2006, *314*, 109-112.

[78] Lee, J. E., Kellie, J. F., Tran, J. C., Tipton, J. D., Catherman, A. D., Thomas, H. M., Ahlf,Dorothy, R., Durbin, K. R., Vellaichamy, A., Ntai, I., Marshall, A. G., Kelleher, N. L., A Robust Two-Dimensional Separation for Top-Down Tandem Mass Spectrometry of the Low-Mass Proteome. J. Am. Soc. Mass Spectrom. 2009, *20*, 2183-2191.

[79] Vellaichamy, A., Tran, J. C., Catherman, A. D., Lee, J. E. *et al.*, Size-Sorting Combined with Improved Nanocapillary Liquid Chromatography-Mass Spectrometry for Identification of Intact Proteins up to 80 kDa. Anal. Chem. 2010, *82*, 1234-1244.

[80] O'Farrell, P. H., High Resolution Two-dimensional Electrophoresis of Proteins. J. Biol. Chem. 1975, *250*, 4007.

[81] Opiteck, G. J., Lewis, K. C., Jorgenson, J. W., Anderegg, R. J., Comprehensive on-line LC/LC/MS of proteins. Anal. Chem. 1997, *69*, 1518-1524.

[82] Wolters, D. A., Washburn, M. P., Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 2001, *73*, 5683-5690.

[83] Washburn, M. P., Wolters, D., Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 2001, *19*, 242-247.

[84] Ofarrell, P. H., High-Resolution 2-Dimensional Electrophoresis of Proteins. J. Biol. Chem. 1975, *250*, 4007-4021.

[85] Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W., The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 2000, *21*, 1037-1053.

[86] Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R., Postel, W., Isoelectric-Focusing in Immobilized Ph Gradients - Principle, Methodology and some Applications. J. Biochem. Biophys. Methods 1982, *6*, 317-339.

[87] Gorg, A., Postel, W., Gunther, S., The Current State of Two-Dimensional Electrophoresis with Immobilized pH Gradients. Electrophoresis 1988, *9*, 531-546.

[88] Svensson, H., Isoelectric Fractionation, Analysis, and Characterization of Ampholytes in Natural Ph Gradients .1. Differential Equation of Solute Concentrations at a Steady State and its Solution for Simple Cases. Acta Chem. Scand. 1961, *15*, 325-341.

[89] Vesterbe.O, Synthesis and Isoelectric Fractionation of Carrier Ampholytes. Acta Chem. Scand. 1969, *23*, 2653-2653.

[90] Laemmli, U. K., Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. Nature 1970, *227*, 680-685.

[91] Shapiro, A. L., Vinuela, E., Maizel, J. V., Molecular Weight Estimation of Polypeptide Chains by Electrophoresis in SDS-Polyacrylamide Gels. Biochem. Biophys. Res. Commun. 1967, *28*, 815-820.

[92] Ornstein, L., Disc Electrophoresis .I. Background and Theory. Ann. N. Y. Acad. Sci. 1964, *121*, 321-349.

[93] Davis, B. J., Disc Electrophoresis .2. Method and Application to Human Serum Proteins. Ann. N. Y. Acad. Sci. 1964, *121*, 404-427.

[94] Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., Improved Staining of Proteins in Polyacrylamide Gels Including Isoelectric-Focusing Gels with Clear Background at Nanogram Sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis 1988, *9*, 255-262.

[95] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Anal. Chem. 1996, *68*, 850-858.

[96] Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., Patton, W. F, A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. Electrophoresis 2000, *21*, 3673-3683.

[97] Berggren, K., Chernokalskaya, E., Steinberg, T. H., Kemper, C., Lopez, M. F., Diwu, Z., Haugland, R. P., Patton, W. F., Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis 2000, *21*, 2509-2521.

[98] Patton, W. F., Detection technologies in proteome analysis. J. Chromatogr. B 2002, *771*, 3-31.

[99] Unlu, M., Morgan, M. E., Minden, J. S., Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. Electrophoresis 1997, *18*, 2071-2077.

[100] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C., Identifying Proteins from 2-Dimensional Gels by Molecular Mass Searching of Peptide-Fragments in Protein-Sequence Databases. Proc. Natl. Acad. Sci. U. S. A. 1993, *90*, 5011-5015.

[101] Motoyama, A., Xu, T., Ruse, C. I., Wohlschlegel, J. A., Yates, J. R., Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Anal. Chem. 2007, *79*, 3623-3634.

[102] Righetti, P. G., Castagna, A., Antonioli, P., Boschetti, E., Prefractionation

techniques in proteome analysis: The mining tools of the third millennium. Electrophoresis 2005, *26*, 297-319.

[103] Heideman, M. L., Purification of Bovine Thyrotropic Hormone by Ion Exchange Chromatography. Endocrinology 1953, *53*, 640-652.

[104] Brument, N., Morenweiser, R., Blouin, V., Toublanc, E., Raimbaud, I., Cherel, Y., Folliot, S., Gaden, F., Boulanger, P., Kroner-Lux, G., Moullier, P., Rolling, F., Salvetti, A., A versatile and scalable two-step ion-exchange chromatography process for the purification of recombinant adeno-associated virus serotypes-2 and-5. Molecular Therapy 2002, *6*, 678-686.

[105] Ottens, A. K., Kobeissy, F. H., Wolper, R. A., Haskins, W. E., Hayes, R. L., Denslow, N. D., Wang, K. K. W., A multidimensional differential proteomic platform using dual-phase ion-exchange chromatography-polyacrylamide gel electrophoresis. Anal. Chem. 2005, *77*, 4836-4845.

[106] Gao, M. X., Zhang, J., Deng, C. H., Yang, P. Y., Zhang, X. M., Novel strategy of high-abundance protein depletion using multidimensional liquid chromatography. J. Proteome Res. 2006, *5*, 2853-2860.

[107] Sluyterman, L. A. A., Elgersma, O., Chromatofocusing - Isoelectric-Focusing on Ion-Exchange Columns .1. General Principles. J. Chromatogr. 1978, *150*, 17-30.

[108] Sluyterman, L. A. A., Wijdenes, J., Chromatofocusing - Isoelectric-Focusing on Ion-Exchange Columns .2. Experimental-Verification. J. Chromatogr. 1978, *150*, 31-44.

[109] Shan, L., Anderson, D. J., Gradient chromatofocusing. Versatile pH gradient separation of proteins in ion-exchange HPLC: Characterization studies. Anal. Chem. 2002, *74*, 5641-5649.

[110] Chong, B. E., Yan, F., Lubman, D. M., Miller, F. R., Chromatofocusing nonporous reversed-phase high-performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry of proteins from human breast cancer whole cell lysates: a novel two-dimensional liquid chromatography. Rapid Commun. Mass Spectrom. 2001, *15*, 291-296.

[111] Buchanan, N. S., Hamler, R. L., Leopold, P. E., Miller, F. R., Lubman, D. M., Mass mapping of cancer cell lysates using two-dimensional liquid separations, electrospray-time of flight-mass spectrometry, and automated data processing. Electrophoresis 2005, *26*, 248-256.

[112] Wang, W. J., Guo, T., Song, T., Lee, C. S., Balgley, B. M., Comprehensive yeast proteome analysis using a capillary isoelectric focusing-based multidimensional separation platform coupled with ESI-MS. Proteomics 2007, *7*, 1178-1187.

[113] Chen, J. Z., Balgley, B. M., DeVoe, D. L., Lee, C. S., Capillary isoelectric focusing-based multidimensional concentration. Anal. Chem. 2003, *75*, 3145-3152.

[114] Wenz, C., Hoerth, P., Wulff, T., Ruefer, A., Preckel, T., Demonstrating reproducibility of OFFGEL isoelectric focusing of proteins. Mol. Cell. Proteomics 2006, *5*, 859.

[115] Lam, H., Josserand, J., Lion, N., Girault, H. H., Modeling the isoelectric focusing of peptides in an OFFGEL multicompartment. cell. J. Proteome Res. 2007, *6*, 1666-1676.

[116] Ernoult, E., Bourreau, A., Gamelin, E., Guette, C., A Proteomic Approach for Plasma Biomarker Discovery with iTRAQ Labelling and OFFGEL Fractionation. J. Biomed. Biotech. 2009, *2010*, 303-311.

[117] Chenau, J., Michelland, S., Sidibe, J., Seve, M., Peptides Offgel electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with iTRAQ labeling. Bull. Cancer 2008, *95*, S19-S19.

[118] Zheng, S., Yoo, C., Delmotte, N., Miller, F. R., Huber, C. G., Lubman, D. M., Monolithic column HPLC separation of intact proteins analyzed by LC-MALDI using on-plate digestion: an approach to integrate protein separation and identification. Anal. Chem. 2006, *78*, 5198-5204.

[119] Yoo, C., Patwa, T. H., Kreunin, P., Miller, F. R., Huber, C. G., Nesvizhskii, A. I., Lubman, D. M., Comprehensive analysis of proteins of pH fractionated samples using monolithic LC/MS. J. Mass Spectrom. 2007, *42*, 312-334.

[120] Lubman, D. M., Kachman, M. T., Wang, H. X., Gong, S. Y., Yan, F., Hamler, R. L., O'Neil, K. A., Zhu, K., Buchanan, N. S., Barder, T. J., Two-dimensional liquid separations-mass mapping of proteins from human cancer cell lysates. J. Chromatogr. B 2002, *782*, 183-196.

[121] Lecchi, P., Gupte, A. R., Perez, R. E., Stockert, L. V., Abramson, F. P., Size-exclusion chromatography in multidimensional separation schemes for proteome analysis. J. Biochem. Biophys. Methods 2003, *56*, 141-152.

[122] Zhang, J., Xu, X., Gao, M., Yang, P., Zhang, X., Comparison of 2-D LC and 3-D LC with post- and pre-tryptic-digestion SEC fractionation for proteome analysis of normal human liver tissue. Proteomics 2007, *7*, 500-512.

[123] Simpson, D. C., Ahn, S., Pasa-Tolic, L., Bogdanov, B., Mottaz, H. M., Vilkov, A. N., Anderson, G. A., Lipton, M. S., Smith, R. D., Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling. Electrophoresis 2006, *27*, 2722-2733.

[124] Lubke, M., LeQuere, J. L., Barron, D., Prefractionation of aroma extracts from fat-containing food by high-performance size-exclusion chromatography. J. Chromatogr. A 1996, *729*, 371-379.

[125] Choi, M. H., Wishnok, J. S., Tannenbaum, S. R., Capillary size-exclusion chromatography as a gel-free strategy in plasma proteomics. Mol. Cell. Toxicology 2005, *1*, 87-91.

[126] Gong, Y., Li, X., Yang, B., Ying, W. T., Li, D., Zhang, Y. J., Dai, S. J., Cai, Y., Wang, J. L., He, F. C., Qian, X. H., Different immunoaffinity fractionation strategies to characterize the human plasma proteome. J. Proteome Res. 2006, *5*, 1379-1387.

[127] Seok, H. J., Hong, M. Y., Kim, Y. J., Han, M. K., Lee, D., Lee, J. H., Yoo, J. S., Kim, H. S., Mass spectrometric analysis of affinity-captured proteins on a dendrimer-based immunosensing surface: investigation of on-chip proteolytic digestion. Anal. Biochem. 2005, *337*, 294-307.

[128] Gygi, S. P., Rist, B., Griffin, T. J., Eng, J., Aebersold, R., Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. J. Proteome Res. 2002, *1*, 47-54.

[129] Bushey, M. M., Jorgenson, J. W., Capillary Electrophoresis of Proteins in Buffers Containing High-Concentrations of Zwitterionic Salts. J. Chromatogr. 1989, *480*, 301-310.

[130] Wiktorowicz, J. E., Colburn, J. C., Separation of Cationic Proteins Via Charge Reversal in Capillary Electrophoresis. Electrophoresis 1990, *11*, 769-773.

[131] Zhao, Z. X., Malik, A., Lee, M. L., Solute Adsorption on Polymer-Coated Fused-Silica Capillary Electrophoresis Columns using Selected Protein and Peptide Standards. Anal. Chem. 1993, *65*, 2747-2752.

[132] Keough, T., Takigiku, R., Lacey, M. P., Purdon, M., Matrix-Assisted Laser Desorption Mass-Spectrometry of Proteins Isolated by Capillary Zone Electrophoresis. Anal. Chem. 1992, *64*, 1594-1600.

[133] Zhang, H., Stoeckli, M., Andren, P. E., Caprioli, R. M., Combining solid-phase preconcentration, capillary electrophoresis and off-line matrix-assisted laser desorption/ionization mass spectrometry: Intracerebral metabolic processing of peptide E in vivo. J. Mass Spectrom. 1999, *34*, 377-383.

[134] Lewis, U. J., Clark, M. O., Preparative Methods for Disk Electrophoresis with Special Reference to Isolation of Pituitary Hormones. Anal. Biochem. 1963, *6*, 303-315.

[135] Zerefos, P. G., Vougas, K., Dimitraki, P., Kossida, S., Petrolekas, A., Stravodimos, K., Giannopoulos, A., Fountoulakis, M., Vlahou, A., Characterization of the human urine proteome by preparative electrophoresis in combination with 2-DE. Proteomics 2006, *6*, 4346-4355.

[136] Andersen, P., Heron, I., Simultaneous Electroelution of Whole Sds-Polyacrylamide Gels for the Direct Cellular Analysis of Complex Protein Mixtures. J. Immunol. Methods 1993, *161*, 29-39.

[137] Tran, J. C., Doucette, A. A., Gel-eluted liquid fraction entrapment electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. Anal. Chem. 2008, *80*, 1568-1573.

[138] Tran, J. C., Doucette, A. A., Multiplexed Size Separation of Intact Proteins in Solution Phase for Mass Spectrometry. Anal. Chem. 2009, *81*, 6201-6209.

[139] Masuoka, J., Glee, P. M., Hazen, K. C., Preparative isoelectric focusing and preparative electrophoresis of hydrophobic Candida albicans cell wall proteins with in-line transfer to polyvinylidene difluoride membranes for sequencing. Electrophoresis 1998, *19*, 675-678.

[140] Parks, B. A., Jiang, L., Thomas, P. M., Wenger, C. D., Roth, M. J., Boyne, M. T., Burke, P. V., Kwast, K. E., Kelleher, N. L., Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers. Anal. Chem. 2007, *79*, 7984-7991.

[141] Assiddiq, B. F., Snijders, A. P. L., Chong, P. K., Wright, P. C., Dickman, M. J., Identification and characterization of Sulfolobus solfataricus P2 proteome using multidimensional liquid phase protein separations. J. Proteome Res. 2008, *7*, 2253-2261.

[142] Pal, M., Moffa, A., Sreekumar, A., Ethier, S. P., Barder, T. J., Chinnaiyan, A., Lubman, D. M., Differential phosphoprotein mapping in cancer cells using protein microarrays produced from 2-D liquid fractionation. Anal. Chem. 2006, *78*, 702-710.

[143] Hamler, R. L., Zhu, K., Buchanani, N. S., Kreunin, P., Kachman, M. T., Miller, F. R., Lubman, D. M., A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification. Proteomics 2004, *4*, 562-577.

[144] Meng, F. Y., Cargile, B. J., Patrie, S. M., Johnson, J. R., McLoughlin, S. M., Kelleher, N. L., Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. Anal. Chem. 2002, *74*, 2923-2929.

[145] Hsu, J. L., Huang, S. Y., Chow, N. H., Chen, S. H., Stable-isotope dimethyl labeling for quantitative proteomics. Anal. Chem. 2003, *75*, 6843-6852.

[146] Hsu, J. L., Huang, S. Y., Shiea, J. T., Huang, W. Y., Chen, S. H., Beyond quantitative proteomics: Signal enhancement of the a(1) ion as a mass tag for peptide sequencing using dimethyl labeling. J. Proteome Res. 2005, *4*, 101-108.

[147] Ji, C. J., Guo, N., Li, L., Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. J. Proteome Res. 2005, *4*, 2099-2108.

[148] Ji, C. J., Li, L., Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS. J. Proteome Res. 2005, *4*, 734-742.

[149] Ji, C. J., Li, L. J., Gebre, M., Pasdar, M., Li, L., Identification and quantification of differentially expressed proteins in E-cadherin deficient SCC9 cells and SCC9 transfectants expressing E-cadherin by dimethyl isotope labeling, LC-MALDI MS and MS/MS. J. Proteome Res. 2005, *4*, 1419-1426.

[150] Melanson, J. E., Avery, S. L., Pinto, D. M., High-coverage quantitative proteomics using amine-specific isotopic labeling. Proteomics 2006, *6*, 4466-4474.

[151] Melanson, J. E., Chisholm, K. A., Pinto, D. M., Targeted comparative proteomics by liquid chromatography/matrix-assisted laser desorption. Rapid Commun. Mass Spectrom. 2006, *20*, 904-910.

[152] Yang, X., Lazar, I. M., MRM screening/biomarker discovery with linear ion trap MS: a library of human cancer-specific peptides. BMC Cancer 2009, *9*, 96.

[153] Kim, K., Kim, S. J., Yu, H. G, Yu, J., Park, K. S., Jang, I., Kim, Y., Verification of Biomarkers for Diabetic Retinopathy by Multiple Reaction Monitoring. J. Proteome Res. 2010, *9*, 689-699.

[154] Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L., Park, B. K., Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. J. Chromatogr. B 2009, *877*, 1229-1239.

[155] Metodieva, G., Greenwood, C., Alldridge, L., Sauven, P., Metodiev, M., A peptide-centric approach to breast cancer biomarker discovery utilizing label-free multiple reaction monitoring mass spectrometry. Proteomics Clin. Appl. 2009, *3*, 78-82.

[156] Giddings, J. C., Concepts and Comparisons in Multidimensional Separation. J. High Resolut. Chrom. Chrom. Comm. 1987, *10*, 319-323.

[157] Wang, H., Hanash, S., Multi-dimensional liquid phase based separations in proteomics. J. Chromatogr. B 2003, *787*, 11-18.

[158] Dowell, J. A., Frost, D. C., Zhang, J., Li, L., Comparison of two-dimensional fractionation techniques for shotgun proteomics. Anal. Chem. 2008, *80*, 6715-6723.

[159] Peng, J. M., Elias, J. E., Thoreen, C. C., Licklider, L. J., Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS. J. Proteome Res. 2003, *2*, 43-50.

[160] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc. Natl. Acad. Sci. U. S. A. 2000, *97*, 9390-9395.

[161] Bjorhall, K., Miliotis, T., Davidsson, P., Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics 2005, *5*, 307-317.

[162] Tran, J. C., Doucette, A. A., Rapid and effective focusing in a carrier ampholyte solution lsoelectric focusing system: A Proteome prefractionation tool. J. Proteome. Res. 2008, *7*, 1761-1766.

[163] Kopaciewicz, W., Rounds, M. A., Fausnaugh, J., Regnier, F. E., Retention Model for High-Performance Ion-Exchange Chromatography. J. Chromatogr. 1983, *266*, 3-21.

[164] Tran, J. C., Wall, M. J., Doucette, A. A., Evaluation of a solution isoelectric focusing protocol as an alternative to ion exchange chromatography for charge-based proteome prefractionation. J. Chromatogr. B 2009, *877*, 807-813.

[165] Van Pelt, C. K., Corso, T. N., Schultz, G. A., Lowes, S., Henion, J., A four-column parallel chromatography system for isocratic or gradient LC. Anal. Chem. 2001, *73*, 582-588.

[166] King, R. C., Miller-Stein, C., Magiera, D. J., Brann, J., Description and validation of a staggered parallel high performance liquid chromatography system for good laboratory practice level quantitative analysis by liquid chromatography. Rapid Commun. Mass Spectrom. 2002, *16*, 43-52.

[167] Briem, S., Pettersson, B., Skoglund, E., Description and validation of a four-channel staggered LC-MS. Anal. Chem. 2005, *77*, 1905-1910.

[168] Lee, H., Griffin, T. J., Gygi, S. P., Rist, B., Aebersold, R., Development of a

multiplexed microcapillary liquid chromatography system for high-throughput proteome analysis. Anal. Chem. 2002, *74*, 4353-4360.

[169] Feng, B. B., McQueney, M. S., Mezzasaima, T. M., Slemmon, J. R., An integrated ten-pump, eight-channel parallel LC. Anal. Chem. 2001, *73*, 5691-5697.

[170] Feng, B. B., Patel, A. H., Keller, P. M., Slemmon, J. R., Fast characterization of intact proteins using a high-throughput eight-channel parallel liquid chromatography mass spectrometry system. Rapid Commun. Mass Spectrom. 2001, *15*, 821-826.

[171] Li, F. A., Wu, M. C., Her, G. R., Development of a multiplexed interface for capillary electrophoresis-electrospray ion trap mass spectrometry. Anal. Chem. 2006, *78*, 5316-5321.

[172] Fang, L. L., Demee, M., Cournoyer, J., Sierra, T., Young, C., Yan, B., Parallel high-throughput accurate mass measurement using a nine-channel multiplexed electrospray liquid chromatography ultraviolet time-of-flight mass spectrometry system. Rapid Commun. Mass Spectrom. 2003, *17*, 1425-1432.

[173] Fang, L., Cournoyer, J., Demee, M., Zhao, J., Tokushige, D., Yan, B., High-throughput liquid chromatography ultraviolet/mass spectrometric analysis of combinatorial libraries using an eight-channel multiplexed electrospray time-of-flight mass spectrometer. Rapid Commun. Mass Spectrom. 2002, *16*, 1440-1447.

[174] Miliotis, T., Kjellstrom, S., Nilsson, J., Laurell, T., Edholm, L. E., Marko-Varga, G., Capillary liquid chromatography interfaced to matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using an on-line coupled piezoelectric flow-through microdispenser. J. Mass Spectrom. 2000, *35*, 369-377.

[175] Li, L., Wang, A. P. L., Coulson, L. D., Continuous-Flow Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry. Anal. Chem. 1993, *65*, 493-495.

[176] Liu, C. L., Zhang, X. M., Multidimensional capillary array liquid chromatography and matrix-assisted laser desorption. J. Chromatogr. A 2007, *1139*, 191-198.

[177] Chen, H. S., Rejtar, T., Andreev, V., Moskovets, E., Karger, B. L., High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. Anal. Chem. 2005, *77*, 2323-2331.

[178] Wei, H., Nolkrantz, K., Powell, D. H., Woods, J. H., Ko, M. C., Kennedy, R. T., Electrospray sample deposition for matrix-assisted laser desorption/ionization (MALDI) and atmospheric pressure MALDI mass spectrometry with attomole detection limits. Rapid Commun. Mass Spectrom. 2004, *18*, 1193-1200.

[179] Young, J. B., Li, L., An impulse-driven liquid-droplet deposition interface for combining LC with MALDI MS and MS/MS. J. Am. Soc. Mass Spectrom. 2006, *17*, 325-334.

[180] Zhang, X., Narcisse, D. A., Murray, K. K., On-line single droplet deposition for MALDI mass spectrometry. J. Am. Soc. Mass Spectrom. 2004, *15*, 1471-1477.

[181] Yoo, C., Zhao, J., Pal, M., Hersberger, K., Huber, C. G., Simeone, D. M., Beer, D. G., Lubman, D. M., Automated integration of monolith-based protein separation with on-plate digestion for mass spectrometric analysis of esophageal adenocarcinorna human epithelial samples. Electrophoresis 2006, *27*, 3643-3651.

[182] Nesbitt, C. A., Jurcic, K., Yeung, K. K. C., Nanoliter-volume protein enrichment, tryptic digestion, and partial separation based on isoelectric points by CE for MALDI mass spectral analysis. Electrophoresis 2008, *29*, 466-474.

[183] Fei, X., Wei, G., Murray, K. K., Aerosol MALDI with a reflectron time-of-flight mass spectrometer. Anal. Chem. 1996, *68*, 1143-1147.

[184] Orsnes, H., Graf, T., Degn, H., Murray, K. K., A rotating ball inlet for on-line MALDI mass spectrometry. Anal. Chem. 2000, *72*, 251-254.

[185] Preisler, J., Foret, F., Karger, B. L., On-line MALDI-TOF MS using a continuous vacuum deposition interface. Anal. Chem. 1998, *70*, 5278-5287.

[186] Zhan, Q., Gusev, A., Hercules, D. M., A novel interface for on-line coupling of liquid capillary chromatography with matrix-assisted laser desorption/ionization detection. Rapid Communications in Mass Spectrometry 1999, *13*, 2278-2283.

[187] Preisler, J., Hu, P., Rejtar, T., Karger, B. L., Capillary electrophoresis-matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using a vacuum deposition interface. Anal. Chem. 2000, *72*, 4785-4795.

[188] Preisler, J., Hu, P., Rejtar, T., Moskovets, E., Karger, B. L., Capillary array electrophoresis-MALDI mass spectrometry using a vacuum deposition interface. Anal. Chem. 2002, *74*, 17-25.

[189] Zhang, B. Y., McDonald, C., Li, L., Combining liquid chromatography with MALDI mass spectrometry using a heated droplet interface. Anal. Chem. 2004, *76*, 992-1001.

[190] Perlman, D. H., Huang, H., Dauly, C., Costello, C. E., McComb, M. E., Coupling of protein HPLC to MALDI-TOF MS using an on-target device for fraction collection, concentration, digestion, desalting, and matrix/analyte cocrystallization. Anal. Chem. 2007, *79*, 2058-2066.

[191] McComb, M. E., Perlman, D. H., Huang, H., Costello, C. E., Evaluation of an on-target sample preparation system for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in conjunction with normal-flow peptide high-performance liquid chromatography for peptide mass fingerprint analyses. Rapid Commun. Mass Spectrom. 2007, *21*, 44-58.

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

[193] Lilley, K. S., Friedman, D. B., All about DIGE: quantification technology for differential-display 2D-gel proteomics. Expert Rev. Proteomics 2004, *1*, 401-409.

[194] Marouga, R., David, S., Hawkins, E., The development of the DIGE system: 2D fluorescence difference gel analysis technology. Anal. Bioanal. Chem. 2005, *382*, 669-678.

[195] Garrett, T. J., Yost, R. A., Analysis of intact tissue by intermediate-pressure MALDI on a linear ion trap mass spectrometer. Anal. Chem. 2006, *78*, 2465-2469.

[196] Garrett, T. J., Prieto-Conaway, M. C., Kovtoun, V., Bui, H., Izgarian, N., Stafford, G., Yost, R. A., Imaging of small molecules in tissue sections with a new intermediate-pressure MALDI linear ion trap mass spectrometer. International J. Mass Spectrom. 2007, *260*, 166-176.

[197] Verhaert, P. D., Conaway, M. C. P., Pekar, T. M., Miller, K., Neuropeptide imaging on an LTQ with vMALDI source: The complete 'all-in-one' peptidome analysis. Int. J. Mass Spectrom. 2007, *260*, 177-184.

[198] Shrestha, B., Nemes, P., Nazarian, J., Hathout, Y., Hoffman, E. P., Vertes, A., Direct analysis of lipids and small metabolites in mouse brain tissue by AP IR-MALDI and reactive LAESI mass spectrometry. Analyst 2010, *135*, 751-758.

[199] Magparangalan, D. P., Garrett, T. J., Drexler, D. M., Yost, R. A., Analysis of Large Peptides by MALDI Using a Linear Quadrupole Ion Trap with Mass Range Extension. Anal. Chem. 2010, *82*, 930-934.

[200] Landgraf, R. R., Garrett, T. J., Calcutt, N. A., Stacpoole, P. W., Yost, R. A., MALDI-linear ion trap microprobe MS/MS studies of the effects of dichloroacetate on lipid content of nerve tissue. Anal. Chem. 2007, *79*, 8170-8175.

[201] Wessel, D., Flugge, U. I., A Method for the Quantitative Recovery of Protein in Dilute-Solution in the Presence of Detergents and Lipids. Anal. Biochem. 1984, *138*, 141-143.

[202] DeSouza, L., Diehl, G., Rodrigues, M. J., Guo, J. Z., Romaschin, A. D., Colgan, T. J., Siu, K. W. M., Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and clCAT with multidimensional liquid chromatography and tandem mass spectrometry. J. Proteome Res. 2005, *4*, 377-386.

[203] Ostin, A., Bergstrom, T., Fredriksson, S. A., Nilsson, C., Solvent-assisted trypsin digestion of ricin for forensic identification by LC-ESI MS. Anal. Chem. 2007, *79*, 6271-6278.

[204] Strader, M. B., Tabb, D. L., Hervey, W. J., Pan, C. L., Hurst, G. B., Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems. Anal. Chem. 2006, *78*, 125-134.

[205] Russell, W. K., Park, Z. Y., Russell, D. H., Proteolysis in mixed organic-aqueous solvent systems: Applications for peptide mass mapping using mass spectrometry. Anal. Chem. 2001, *73*, 2682-2685.

[206] Batra, R., Gupta, M. N., Enhancement of Enzyme-Activity in Aqueous-Organic Solvent Mixtures. Biotechnol. Lett. 1994, *16*, 1059-1064.

[207] Chen, E. I., Cociorva, D., Norris, J. L., Yates, J. R., Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. J. Proteome Res. 2007, *6*, 2529-2538.

[208] Lill, J. R., Ingle, E. S., Liu, P. S., Pham, V., Sandoval, W. N., Microwave-assisted proteomics. Mass Spectrom. Rev. 2007, *26*, 657-671.

[209] Lin, S. S., Wu, C. H., Sun, M. C., Sun, C. M., Ho, Y. P., Microwalve-assisted enzyme-catalyzed reactions in various solvent systems. J. Am. Soc. Mass Spectrom. 2005, *16*, 581-588.

[210] Doucette, A., Craft, D., Li, L., Protein concentration and enzyme digestion on microbeads for MALDI-TOF peptide mass mapping of proteins from dilute solutions. Anal. Chem. 2000, *72*, 3355-3362.

[211] Xu, X. Q., Deng, C. H., Yang, P. Y., Zhang, X. M., Immobilization of trypsin on superparamagnetic nanoparticles for rapid and effective proteolysis. J. Proteome Res. 2007, *6*, 3849-3855.

[212] Jeng, J. Y., Lin, M. F., Cheng, F. Y., Yeh, C. S., Shiea, J. T., Using high-concentration trypsin-immobilized magnetic nanoparticles for rapid in situ protein digestion at elevated temperature. Rapid Commun. Mass Spectrom. 2007, *21*, 3060-3068.

[213] Ma, J., Zhang, L., Liang, Z., Zhang, W., Zhang, Y., Monolith-based immobilized enzyme reactors: Recent developments and applications for proteome analysis. J. Separ. Sci. 2007, *30*, 3050-3059.

[214] Massolini, G., Calleri, E., Immobilized trypsin systems coupled on-line to separation methods: Recent developments and analytical applications. J. Separ. Sci. 2005, *28*, 7-21.

[215] Temporini, C., Perani, E., Mancini, F., Bartolini, M., Calleri, E., Lubda, D., Felix, G., Andrisano, V., Massolini, G., Optimization of a trypsin-bioreactor coupled with high-performance liquid chromatography-electrospray ionization tandem mass spectrometry for quality control of biotechnological drugs. J. Chromatogr. A 2006, *1120*, 121-131.

[216] Svec, F., Less common applications of monoliths: I. Microscale protein mapping with proteolytic enzymes immobilized on monolithic supports. Electrophoresis 2006, *27*, 947-961.

[217] Slysz, G. W., Schriemer, D. C., On-column digestion of proteins in aqueous-organic solvents. Rapid Commun. Mass Spectrom. 2003, *17*, 1044-1050.

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

[219] Hervey, W. J., Strader, M. B., Hurst, G. B., Comparison of digestion protocols for microgram quantities of enriched protein samples. J. Proteome Res. 2007, *6*, 3054-3061.

[220] Schwert, G. W., Takenaka, Y., A Spectrophotometric Determination of Trypsin and Chymotrypsin. Biochim. Biophys. Acta 1955, *16*, 570-575.

[221] Simon, L. M., Laszlo, K., Vertesi, A., Bagi, K., Szajani, B., Stability of hydrolytic enzymes in water-organic solvent systems. J. Mol. Catal. B 1998, *4*, 41-45.

[222] Gupta, M. N., Roy, I., Enzymes in organic media - Forms, functions and applications. Eur. J. Biochem. 2004, *271*, 2575-2583.

[223] Klibanov, A. M., Improving enzymes by using them in organic solvents. Nature 2001, *409*, 241-246.