Urine Proteomics for the Purpose of Biomarker Discovery

by

Diane M. Botelho

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

at

Dalhousie University
Halifax, Nova Scotia
July 2010

DALHOUSIE UNIVERSITY

DATE: July 20, 2010

AUTHOR: Diane M. Botelho

TITLE: Urine Proteomics for the Purpose of Biomarker Discovery

DEPARTMENT OR SCHOOL: Department of Chemistry

DEGREE: PhD CONVOCATION: October YEAR: 2010

Dedicated to

My Husband and Daughter

TABLE OF CONTENTS

LIST OF TABLES.....xi

LIST OF FIGURES.....xiii

ABSTRACT.....xvii

LIST OF ABBREVIATIONS AND SYMBOLS USED.....xviii

ACKNOWLEDGEMENTS.....xx

CHAPTER 1: INTRODUCTION 1

1.1 Proteomics Background……. 1

1.1.2 The Evolution of Proteomics……..1

1.1.3 Mass Spectrometry Comes of Age……2

1.2 MS Analysis of Proteins…….. 2

1.2.1 Ionization of Biomolecules……2

1.3 MS-Based Methods for Protein Identification……. 7

1.3.1 Peptide Mass Fingerprinting and Peptide Mass Sequencing……. 7

1.4 Methods for Protein Quantitation……. 13

1.5 Challenges in Proteomics……..22

1.6 Approaches for Proteome Fractionation…….....24

1.6.1 One and Two Dimensional Gel Electrophoresis…….....26

1.6.2 One Dimensional Polyacrylamide Gel Electrophoresis……27

1.6.3 Two Dimensional Polyacrylamide Gel Electrophoresis……... 29

1.7 Top Down and Bottom Up Proteomics……31

CHAPTER 2: URINE PROTEOMICS FOR BIOMARKER DISCOVERY……33

2.1 Biology of Urine Formation……..... 33

2.2 Disorders of the Kidney and Urinary Tract……. 35

2.3 Proteins Expected to be Found in Urine……..... 36

2.4 Methods for the Analysis of Urine Proteins…….. 37

2.5 Candidate Biomarkers Found in Urine……. 40

2.6 Challenges in Urine Proteomics……. 42

2.7 Research Proposals…….. 45

CHAPTER 3: ANALYSIS OF URINE PROTEINS AND THE IMPORTANCE OF URINE SEDIMENT PROTEINS ……. 48

3.1 Introduction……...48

3.2 Materials and methods……..49

3.2.1 Materials…….....49

3.2.2 Urine Collection and Preparation for 2D-SDS-PAGE-MALDI-MS^2…… 50

3.2.3 2D-SDS-PAGE of Urinary Supernatants…….50

3.2.4 2D-SDS-PAGE of Urinary Sediments…….51

3.2.5 Protein Digestion and Peptide Extraction…… 51

3.2.6 MALDI-MS^2 of Proteins Separated by 2D-SDS-PAGE……. 52

3.2.7 Urine Collection for ESI-MS^2 Analysis of Samples…… 52

3.2.8 Protein Visualization by SDS-PAGE……...52

3.2.9 Sample Manipulations for Qualitative Mass Spectrometry *via* ESI-MS^2…….. 53

3.2.10 Effect of Freezing on Amount of Sediment Proteins…….....54

3.2.11 Effect of Vortexing in Protein Distribution Between Phases…….54

3.2.12 Effect of pH on the Partitioning of Proteins Between Phases……55

3.2.13 Solution Isoelectric Focusing and Protein Digestion……. 56

3.2.14 LC-ESI-MS^2 of Proteins and Data Analysis…….. 56

3.2.15 iTRAQ Analysis of Effect of pH on Partitioning of Proteins…… 57

3.3 Results and Discussion……..59

3.3.1 A Proteomic Investigation of Urinary Sediments…… 59

3.3.2 Examining the Proteins in Urinary Sediments by 2D-SDS-PAGE-MALDI-MS^2..... 63

3.3.3 Effect of a Freeze-thaw Cycle on Sediment Proteins……...…69

3.3.4 Vortexing a Urine Sample Prior to Analysis Resolubilizes Some Sediment Proteins…….. 79

3.3.5 pH of a Sample Influences the Protein profile of Urinary Sediments …….. 81

3.3.6 iTRAQ Quantitation of Sediment Proteins…….. 83

3.4 Conclusions…… 84

CHAPTER 4: MASS SPECTROMETRY TOLERANCE TO SODIUM DODECYL SULPHATE…… 86

4.1 Introduction……..... 86

4.2 Materials and Methods…….89

4.2.1 Solutions, Reagents and Protein Standards……. 89

4.2.2 Protein Precipitation…….89

4.2.3 SDS Assay……90

4.2.4 LC-TOF MS of Intact Proteins…….90

4.2.5 Bottom up Proteome Analysis for SDS-Containing Samples.....……91

4.3 Results and Discussion……..92

4.3.1 Effects of SDS on Proteomics LC-MS Experiments…….. 92

4.3.2 Bottom Up SDS Suppression……...92

4.3.3 Top Down SDS Suppression……96

4.3.4 Removal of SDS from Protein Containing Samples……99

4.4 Conclusions……..101

CHAPTER 5: A PROTEOMIC COMPARISON OF GELFrEE TO GeLC…….. 102

5.1 Introduction……102

5.2 Materials and Methods……...103

5.2.1 Yeast Proteome Extraction…… 103

5.2.2 Acetone and SPE Preparation of Urine Proteins……103

5.2.3 GELFrEE Separation…….. 104

5.2.4 GELFrEE of Urine Proteins and Visualization *via* 1D-SDS-PAGE……. 104

5.2.5 GELFrEE of Urine Proteins for LC-MS^2……...105

5.2.6 Comparison of Acetone and SPE Extracted Proteins Following GELFrEE ……..105

5.2.7 1D-SDS-PAGE of GELFrEE *VS* GeLC Comparison……106

5.2.8 Filter-Aided Sample Preparation of GELFrEE Fractions……..107

5.2.9 Solution Digestion of GELFrEE Fractions……107

5.2.10 Sample Cleanup, Mass Spectrometry and Database Searching……...108

5.3 Results and Discussion……109

5.3.1 Assessment of Proteins Identified *via* GELFrEE and GeLC……109

5.3.2 GELFrEE Versus GeLC Protein Resolution……..115

5.3.3 GELFrEE of Urine Proteins……...119

5.3.4 GELFrEE-MS Analysis of Acetone and SPE Extracted Proteins …… 125

5.4 Conclusions……..127

CHAPTER 6: INTACT ISOTOPE LABELLING OF URINE PROTEINS AND SEPARATION *VIA* GELFrEE……..……129

6.1 Introduction……129

6.2 Materials and Methods……...132

6.2.1 Solvents and Reagents……... 132

6.2.2 Extraction of Urine Proteins…….. 133

6.2.3 Isotope Labelling of Proteins……. 133

6.2.4 GELFrEE of Labelled Urine Proteins……134

6.2.5 1D-SDS-PAGE for Visualization of GELFrEE Fractions……134

6.2.6 Digestion of Samples……. 134

6.2.7 Sample Cleanup by RPLC……. 135

6.2.8 LC-MS^2 and database searching of Fractions…… 135

6.3 Results and Discussion……...…….136

6.3.1 Qualitative Investigation of labelling…….136

6.3.2 A Quantitative Investigation of Labelled Peptides……..... 147

6.4 Conclusions……..152

CHAPTER 7: INTACT LABELLING AND GELFrEE FRACTIONATION OF RAT URINE FOR BIOMARKER DISCOVERY…….. 153

7.1 Introduction……153

7.2 Materials and Methods…….. 153

7.2.1 Collection of Control and Test Urine Samples……..153

7.2.2 Preparation of urine Samples for Intact Labelling and GELFrEE…..... 154

7.2.3 GELFrEE of Labelled Samples……. 154

7.2.4 Digestion of GELFrEE Samples …….....155

7.2.5 MS of Proteins and Database searching……155

7.3 Results and Discussion……155

7.4 Conclusions……..162

CHAPTER 8: CONCLUSIONS…….. 163

8.1 Future Work……165

REFERENCES……. 166

LIST OF TABLES

Table 1.1 Matrix molecules commonly used in MALDI-MS and associated analytes conventionally used with these matrices. ……6

Table 1.2 Chemical and enzymatic species commonly used for the digestion of proteins……..8

Table 1.3 Databases and their website addresses available for PMF of MS spectra……10

Table 1.4 Examples of commonly observed post translational modifications of proteins and the sites which are most probable for these modifications……..……24

Table 2.1 A very small selection of candidate biomarker proteins from urine, as described in the literature ……42

Table 3.1 Sediment proteins identified from 2D-SDS-PAGE followed by MALDI-MS^2…….. 66

Table 3.2 Supernatant proteins identified from 2D-SDS-PAGE followed by MALDI-MS^2……67

Table 3.3 Urinary proteins identified by LC-MS^2……75

Table 3.4 iTRAQ ratios of urinary proteins after subjected to pH adjustments ……..... 83

Table 5.1 Proteins and peptides identified from GeLC and GELFrEE using chloroform-methanol-water, acetone and the spin column procedure for sample preparation ……119

Table 6.1 The tryptic and chymotryptic peptides of alpha-1-microglobulin through MASCOT search engine without specifying any labels ……. 142

Table 6.2 The tryptic and chymotryptic peptides of alpha-1-microglobulin incorporating the IDOG intact feature……...143

Table 6.3 Labelled proteins digested with trypsin and their corresponding peptides identified *via* MASCOT search with and without labels……145

Table 6.4 Labelled proteins digested with chymotrypsin and their corresponding peptides identified *via* MASCOT …….. 147

Table 6.5 Human urinary proteins quantified from intact labelling……. 150

Table 7.1 List of control –control rat urinary proteins meeting all requirements for reliable quantitation……..…… 156

Table 7.2 Quantitation of proteins from the obstructed *vs* unobstructed rat samples…… 158

LIST OF FIGURES

Figure 1.1 Depiction of the charge residue model of ion formation in ESI…… 4

Figure 1.2 The process of MALDI ion formation. ……. 5

Figure 1.3 A peptide mass fingerprint of bovine serum albumin obtained from a mass spectrometer equipped with a MALDI source ……9

Figure 1.4 Schematic of a peptide showing the terminology used when fragmented along the peptide backbone ……11

Figure 1.5 An example of a peptide mass sequence spectrum obtained from the fragmentation of a peptide from a urine protein tryptic digest …….. 12

Figure 1.6 The first ICAT tag developed for protein quantitation……... 15

Figure 1.7 The iTRAQ tag strategy showing the peptide, mass balance and reporter ion components……..... 17

Figure 1.8 MS peaks for iTRAQ labelled peptides ……. 18

Figure 1.9 Isotope labelling of peptides using formaldehyde ……..….. 20

Figure 1.10 A theoretical mass spectrum that would result from formaldehyde and deuterated formaldehyde labelled peptides…….. 21

Figure 1.11 Schematic of SDS-PAGE……... 28

Figure 1.12 Schematic of 2D-SDS-PAGE …… 30

Figure 2.1 A human kidney……..34

Figure 3.1 Urinary sediments isolated from 1 mL and 50 mL of urine that has undergone a freeze-thaw cycle …… 49

Figure 3.2 A photomicrograph of urinary sediments stained with crystal violet ……... 59

Figure 3.3 Urinary sediments remaining after solubilization with PBS…….. 60

Figure 3.4 SDS-PAGE and 2D-SDS-PAGE of urinary sediments. …… 61

Figure 3.5 2D-SDS-PAGE or urinary supernatants and sediments……. 63

Figure 3.6 Image of human urine before and after freezing …… 70

Figure 3.7 Urinary sediment proteins before and after freezing…….. 71

Figure 3.8 Urinary sediment proteins from 4 different donors ……... 72

Figure 3.9 Urinary sediments isolated from the same donor on different days ……..73

Figure 3.10 A summary of proteins identified in urinary sediments and supernatants *via* LC-ESI-MS^2…….. 74

Figure 3.11 Sediment and supernatant proteins isolated from the urine of 4 healthy donors ……80

Figure 3.12 Urinary protein profiles of control samples with no pH adjustment and of samples subjected to pH adjustment …… 82

Figure 4.1 The number of yeast peptides identified by MS^2 as a function of the % SDS in the peptide solution ……. 93

Figure 4.2 Total ion chromatograms of a yeast at 0% and 0.02% SDS……... 94

Figure 4.3 Selected ion chromatograms for two peptides at varying percentages of SDS……. 95

Figure 4.4 Normalized intensity versus % SDS for 6 intact protein standards after LC-TOF MS analysis …… 97

Figure 4.5 Charge envelope and deconvoluted spectra observed from LC-ESI analysis of carbonic anhydrase at 0% SDS and 0.025 % SDS..... 98

Figure 4.6 The percentage of SDS that remains following chloroform-methanol-water and acetone precipitation of a yeast sample ……100

Figure 5.1 Fractions collected from a yeast proteome separated using GELFrEE ……110

Figure 5.2 Yeast proteins fractionated using GELFrEE on a 5% T tube gel, allowing for higher masses to be collected ……... 110

Figure 5.3 Venn diagrams comparing the number of proteins identified by GeLC and GELFrEE when employing chloroform-methanol-water and acetone precipitation following solution fractionation ……... 112

Figure 5.4 Venn diagram comparing the number of peptides identified by GeLC and GELFrEE when employing chloroform-methanol-water and acetone precipitation following solution fractionation …….. 112

Figure 5.5 A mass distribution plot of showing the unique protein identified from each of the GeLC and GELFrEE platforms ……114

Figure 5.6 Mass distribution profiles of proteins identified by the GeLC platform and the GELFrEE platform …… 116

Figure 5.7 The percentage of proteins identified across multiple fractions for GELFrEE/acetone and GeLC separations ……..117

Figure 5.8 1D-SDS-PAGE image resulting from analysis of GELFrEE fractions obtained from separation of urinary proteins……121

Figure 5.9 Images of the urine samples used for SPE and acetone protein extraction after a freeze-thaw cycle before vortexing and after vortexing.……..... 122

Figure 5.10 Image of the SPE extract of the supernatant, the remaining sediments and acetone extracted proteins from urine as they are fractionated by GELFrEE. ……... 123

Figure 5.11 1D-SDS-PAGE of urinary GELFrEE fractions in which the urine proteins were initially extracted using SPE and acetone protocols ……. 124

Figure 5.12 Venn diagrams displaying the urinary proteins and peptides identified *via* acetone and SPE protocols ……126

Figure 6.1 Diagram showing the difference in labelling with formaldehyde when performed at the bottom up level versus the top down …….. 131

Figure 6.2 SDS-PAGE images showing the labelled samples spiked with SDS and urea following fractionation by GELFrEE ……. 137

Figure 6.3 MS selected ion chromatogram from the MS analysis of peptides digested by chymotrypsin after intact formaldehyde labelling and GELFrEE fractionation……139

Figure 6.4 Selected ion chromatogram from the MS analysis of peptides digested by trypsin after intact formaldehyde labelling and GELFrEE fractionation ……..139

Figure 6.5 The amino acid sequence of human alpha-1-microglobulin…….. 141

Figure 6.6 Correlation spectra of two labelled peptides with a correlation value of correlation: 0.91 and 0.7…….....148

Figure 7.1 Correlation spectrum for alpha-2-microglobulin observed at a ratio of 0.87:1 for heavy to light samples ……... 157

Figure 7.2 Correlation spectrum for utrophin, a protein that was 12 fold lower in the healthy sample compared to the obstructed sample.…….. 159

Figure 7.3 Correlation spectra for Olfactory receptor protein, a protein from rat urine identified to be 57 fold higher in the control sample ……. 162

ABSTRACT

Body fluids have gained widespread importance for proteomics based biomarker discovery as they have proved to reveal many candidate biomarkers for a variety of physiological diseases. Urine is a particularly favourable body fluid when profiling diseases associated with proximal tissues (kidney, urinary tract, *etc.*). The collection of urine is also non-invasive compared to other fluids such as blood, plasma and amniotic or cerebral spinal fluids.

The main objective of this work was to determine an optimal protocol for profiling of the urine proteome *via* peptide mass sequencing. Subsequently, the urine proteome was characterized as a potential source for protein biomarker(s) related to kidney obstruction. Of particular relevance to a quantitative investigation, it was found that the conventional urinary proteome workflow inadvertently introduces a sampling bias. Demonstrated here is the fact that the sediment proteins of urine, which are typically discarded prior to analysis, contain important protein constituents which were previously reported in the literature as candidate biomarkers.

A solution-based intact protein separation workflow was demonstrated for the analysis of urinary proteins. The mass-based separation platform, namely gel-eluted liquid fractionation entrapment electrophoresis (GELFrEE), incorporates the use of sodium dodecyl sulphate (SDS). As a known signal suppressant in mass spectrometry (MS), the MS tolerance of SDS in a proteome workflow was determined. Simple and effective protocols for the isolation of proteins from SDS-containing solutions are presented. Also presented is a comparison between GELFrEE and the conventional 'GeLC' protocol for SDS-based protein separation, with similar numbers of proteins identified by the two platforms. Lastly, a novel strategy for isotope labelling of proteins at the intact level is presented. This intact labelling of proteins is therefore compatible with intact proteome prefractionation, as seen with GELFrEE-MS^2 separation. This quantitative workflow was applied to biomarker profiling from urine samples obtained for a model (rodent) kidney obstruction.

LIST OF ABBREVIATIONS AND SYMBOLS USED

1D	one dimensional
2D	two dimensional
AMBP	alpha-1-microglobulin
AMU	atomic mass units
BCA	bicinchoninic acid
BSA	bovine serum albumin
CAT	computed axial tomography
CE-MS^2	capillary electrophoresis followed by MS^2
CID	collision induced disassociation
CMW	chloroform/ methanol/ water
D_2	deuterium
Da	Dalton
DGP	dystrophin glycoprotein complex
DTT	dithiothreitol
ESI	electrospray ionization
FASP	filter aided sample preparation
FFE	free flow electrophoresis
g	gravity
GELFrEE	gel eluted liquid fractionation entrapment electrophoresis
h	hour
HCCA	alpha-hydroxy 4-hydroxycinnamic acid
HILIC	hydrophobic interaction liquid chromatography
HKUPP	human kidney and urine proteome project
HPLC	high performance liquid chromatography
ICAT	isotope coded affinity tag
IEF	isoelectric focusing
IL-7	interleukin-7
ITRAQ	isobaric tagging for relative and absolute quantification
kDa	kilodalton
LC	liquid chromatography

m/z	mass to charge ratio
MALDI	matrix assisted laser desorption ionization
mg/mL	milligrams per millilitre
MH^+	molecular ion
min	minute(s)
mL	millilitre(s)
MRI	magnet resonance imaging
MRM	multiple reaction monitoring
MS	mass spectrometry
MS^2	tandem mass spectrometry
MW	molecular weight
PBS	phosphate buffered saline
pg/mL	picogram per millilitre
pI	isoelectric point
PMF	peptide mass fingerprinting
psi	pounds per square inch
PTM	post translational modification
RPLC	reversed phase liquid chromatography
rpm	rotations per minute
SCX	strong cation exchange
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELDI	surface-enhanced laser desorption ionization
sIEF	solution isoelectric focusing
SILAC	stable isotope labelling by amino acids
SIM	selected ion monitoring
SPE	solid phase extraction
TFA	trifluoroacetic acid
TOF	time of flight
UPJO	ureteropelvic junction kidney obstruction
UV	ultraviolet
Vh	volt hour
YEPD	yeast extract peptome dextrose
µg	microgram
µm	micrometer
^oC	degrees Celcius
µL	microlitre

ACKNOWLEDGEMENTS

I would like to thank with greatest sincerity my supervisor Dr. Alan Doucette whose guidance, leadership and incredible knowledge of proteomics has been invaluable to my research. Not only is Dr. Doucette an exceptional supervisor, he is a sincere and kind individual whom is very dedicated to research and his teaching obligations. He has much to contribute to any student, whether undergraduate or graduate.

I also thank past and present members of Dr. Doucette's research group, namely John Tran, Mark Wall, Fang Liu, Dennis Orton, Douglas Vieira and Shayla Fitzsimmons. I cannot express how wonderful of an experience it has been to work with and learn from all of these students.

I would like to thank Dr. David Byers, Dr. Robert White and Dr. Robert Guy for taking the time to read this thesis as well as all the time they have dedicated towards my research project during my time at Dalhousie University. I thank Dr. Neil Kelleher for agreeing to be my external examiner and I also thank Dr. Dawn MacLellan for providing me with guidance and any additional equipment I required for my research.

I owe sincere gratitude to Dr. Devanand Pinto, Patrick Murphy and Ken Chisholm from the Institute for Marine Biosciences in Halifax for opening their doors to me, allowing me to carry out research in their laboratories as I pleased. I thank Mike Boutlier and Rick Conrad from the chemistry machine shop for all of the mechanical help they have provided to my research lab.

I also thank the ladies in the chemistry office, Giselle Andrews, Deanna Wentzell Cheryl Stanton and Gail Powers. They are truly dedicated to the chemistry students at Dalhousie University. They make the students experience a wonderful one.

Last but certainly not least, I would like to thank my family for all of their support throughout my degree. To my parents and sister, thank you for encouraging me continuously from the beginning of my undergraduate degree. To my husband and daughter, you have given me the motivation and strength I never knew I had to continue to pursue this dream of a PhD, and for that I love you both with all my heart.

CHAPTER 1: INTRODUCTION

1.1 Proteomics Background

1.1.2 The Evolution of Proteomics

Proteomics involves the systematic determination of protein sequence, quantity, modification states, interaction partners, activity, subcellular localization, and structure in any particular cell at any particular time [1]. Without doubt, such a determination is an enormous analytical challenge, requiring the most intricate of experiments, complex instrumentation and vast amount of computational analysis.

The historical evolution of proteomics is a testament to the rapid progression and multidisciplinary nature of modern science. In 1955, the sequencing of a protein was perceived as an immense, if not impossible challenge. During this time, it was inconceivable to project that thousands of proteins could be sequenced in a single experiment, enabling the profiling of a complete proteome. Proteins were initially believed to be amorphous species', with no defined or static sequence. Dr. Frederick Sanger dispelled these beliefs when he sequenced the first protein; insulin [2]. This was quite an accomplishment, considering that today's high throughput intruments were not available and thus his sequencing technology needed to be developed 'from scratch'. His decade-long accomplishment was eventually completed through the pioneering use of simple instrumentation (paper chromatography, electrophoresis) and basic chemical and enzymatic reactions. To no surprise, this work earned Sanger the Nobel Prize in Chemistry in 1958.

Since Sanger's time, researchers have developed improved chemical-based techniques for protein sequencing, including the well-known Edman degradation reaction [3]. It was also realized at this time that the incorporation of separation technologies, not only for the purpose of teasing out the resulting products of enzyme digestion, but also to isolate and purify proteins from their native biological source were critical steps in the analysis.

1.1.3 Mass Spectrometry Comes of Age

It was not until the late 1980's when a new and reliable detection strategy for protein sequencing became available. The ionization of large biomolecules was an incredible breakthrough in mass spectrometry (MS). While earlier developments in MS ionization gradually permitted the analysis of increasingly larger molecules, the development of two particular ionization methods, namely electrospray ionization (ESI) [4] and matrix-assisted laser desorption/ ionization (MALDI) [5,6], exposed a new light for high-throughput sequencing and identification of proteins. Soon thereafter, the conventional strategies for protein sequencing built on protein separation and chemical detection were translated into this instrumentational platform.

1.2 MS Analysis of Proteins

1.2.1 Ionization of Biomolecules

ESI and MALDI are two very different ionization techniques, each contributing significantly to protein analysis *via* MS. ESI is a solution phase process of which the mechanism of ion formation is still being debated in part [7]. However, there are two

theories considered to be credible explanations for the ESI ion formation process. The first theory is derived from the ion evaporation model [7]. In this model, the analyte and solvent, sprayed from a positively charged capillary needle, form an aerosol containing droplets of the solvent-analyte mixture. Through evaporation, these droplets eventually reach a critical radius in which the field strength at the surface of the droplet is enough to aid in desorption of charged ions. The second model, the charged residue model [7], theorizes that ions are formed by the disintegration of charged droplets that form from aerosol droplets (Figure 1.1).

In contrast to ESI, MALDI is a solid phase ionization technique in which the analyte is co-crystallized with an acidic matrix molecule on a sample target plate. Ions are then formed by firing a high energy laser at the sample (Figure 1.2).

Most importantly, these two types of ion formation processes are soft ionization techniques, meaning that they are able to ionize molecules without causing excessive fragmentation. This is particularly important for protein analysis, as it is essential that the protein or peptide being analyzed remains intact until it travels into the mass analyzer. If desired, peptide or proteins can then be fragmented in a controlled fashion for amino acid sequencing. An additional benefit of ESI and to a lesser extent, MALDI is that they both easily permit coupling of liquid chromatography (LC) to MS, being a favoured separation strategy in the proteomic workflow.

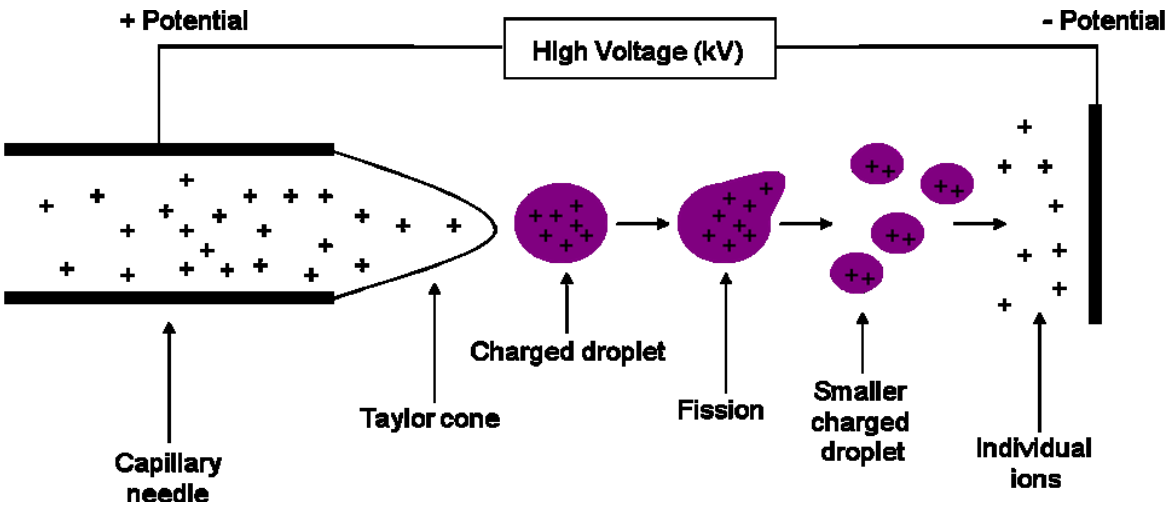


Figure 1.1: Depiction of the charge residue model of ion formation in ESI. The charged droplets formed from a positively charged capillary needle are attracted to an electrode plate held at relative negative potential. As the droplets move toward the electrode plate, the solvent begins to evaporate. Evaporation causes the charges on the drop to concentrate and due to space-charge effects, the droplet eventually breaks up into singly, doubly or triply charged ions [7].

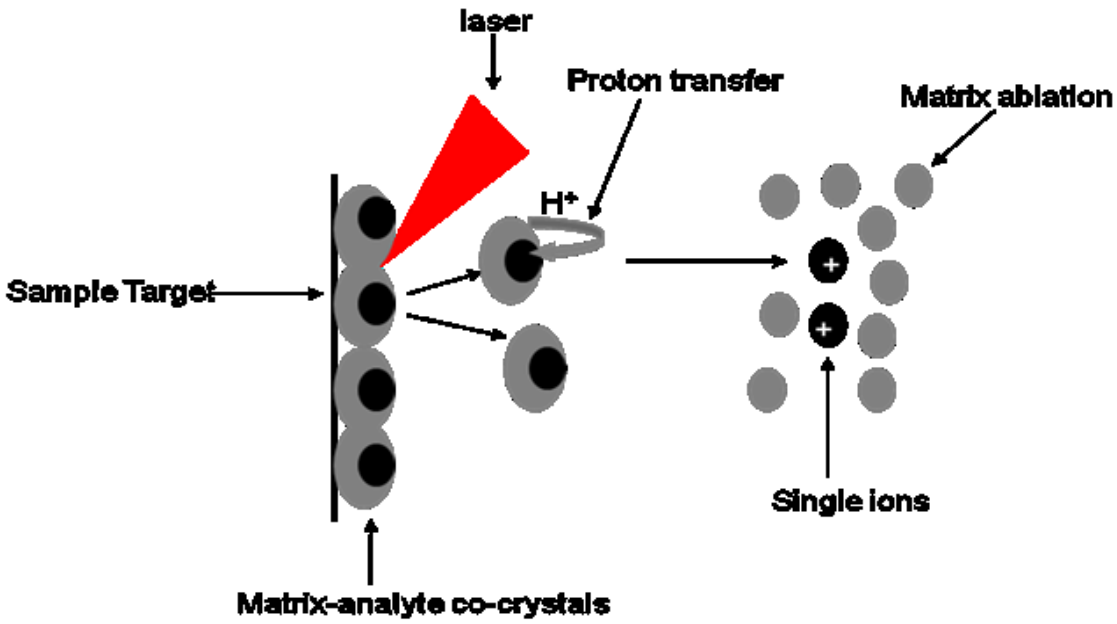


Figure 1.2: The process of MALDI ion formation. A high energy laser is fired at the matrix-analyte co-crystals, deposited on a target plate. Energy provided from the laser is sufficient to create an "explosion" of the matrix-analyte crystals. There is then a proton transfer from the acidic matrix molecule to the analyte, typically resulting in a singly, positively charged analyte ion [8].

Several different matrix molecules are commercially available, each with its own attributes depending on the nature of the sample being analyzed. The most common matrix molecules employed for MALDI can be seen in Table 1.1. A more descriptive discussion of ESI-MS and MALDI-MS can be obtained from Kriwacki *et al.* [9].

Table 1.1: Matrix molecules commonly used in MALDI-MS and associated analytes conventionally used with these matrices.

Matrix Molecule	Optimum Analyte Type
α-cyano-4-hydroxycinnamic acid (HCCA)	peptides, small proteins
3,5-dimethoxy-4-hydroxycinnamic acid	proteins
2-(4-hydroxyphenylazo)benzoic acid	peptides, proteins, glycoproteins
2-mercaptobenzothiazole	peptides, proteins, synthetic polymers
5-chloro-2-mercaptobenzothiazole	peptides, proteins, synthetic polymers
2,6-dihydroxyacetophenone	glycopeptides, phosphopeptides, proteins
picolinic acid	oligonucleotides
3-hydroxyoicolinic acid	oligonucleotides
3-aminopicolinic acid	oligonucleotides
2,5-dihydroxybenzoic acid	oligosaccharides, peptides

Proteomics has proven to have many applications, from disease control and prevention to drug design and development. It has even found a place in development of vegetative crops resistant to extreme weather conditions [10]. Undoubtedly, MS plays an imperative role in each of these processes.

1.3 MS-Based Methods for Protein Identification

1.3.1 Peptide Mass Fingerprinting and Peptide Mass Sequencing

Currently, there are several methods for protein identification using mass spectrometry. Two of the most common methods are peptide mass fingerprinting (PMF) and peptide mass sequencing. In PMF [11-15], a purified protein (typically a gel slice from one or two dimensional gel electrophoresis) is cleaved into smaller peptide counterparts, either by chemical or enzymatic cleavage. Table 1.2 depicts a list of commonly employed chemical and enzymatic reagents as well as their cleavage sites. Once cleaved, the protein fragments (*a.k.a* peptides) are analyzed by MS to determine the mass of each fragment.

Table 1.2: Chemical and enzymatic species commonly used for the digestion of proteins. Also listed is the cleavage site specific to each chemical or enzyme.

Chemical or Enzyme	Cleavage Site
Cyanogen bromide	methionine
2-(2'-Nitrophenylsulfonyl)-3-methyl-3-bromoindolenine	tryptophan
Formic acid	aspartic acid-proline peptide bonds
Hydroxylamine	asparagine-glycine peptide bonds
2-nitro-5-thiocyanobenzoic acid	cysteine
Trypsin	lysine and arginine
Chymotrypsin	tryptophan, tyrosine, phenylalanine
Endoproteinase Arg-C	arginine
Endoproteinase Lys-C	lysine
Pepsin	between hydrophobic and aromatic residues

Owing to the specificity of the cleavage reagents, the resulting list of peptide masses is unique to that protein, and hence constitutes a 'peptide fingerprint' of the protein (Figure 1.3). This mass list can then be compared *in silico* to a database of proteins that have been theoretically digested with the same enzyme. Statistical analysis is required to determine how well the peptide peaks detected by MS match those generated *in silico* where a decision can be made as to whether the candidate match is considered a positive identification of the protein.

While considered a high-throughput approach, the main disadvantage of PMF is that the protein being analyzed must be relatively pure. Of course, the protein sequence must also be known. PMF is not a suitable approach for samples in which the protein is expected to be novel or unsequenced. Several online protein databases are available for searching mass spectrometric data which typically takes advantage of the information made available through genome sequencing (Table 1.3).

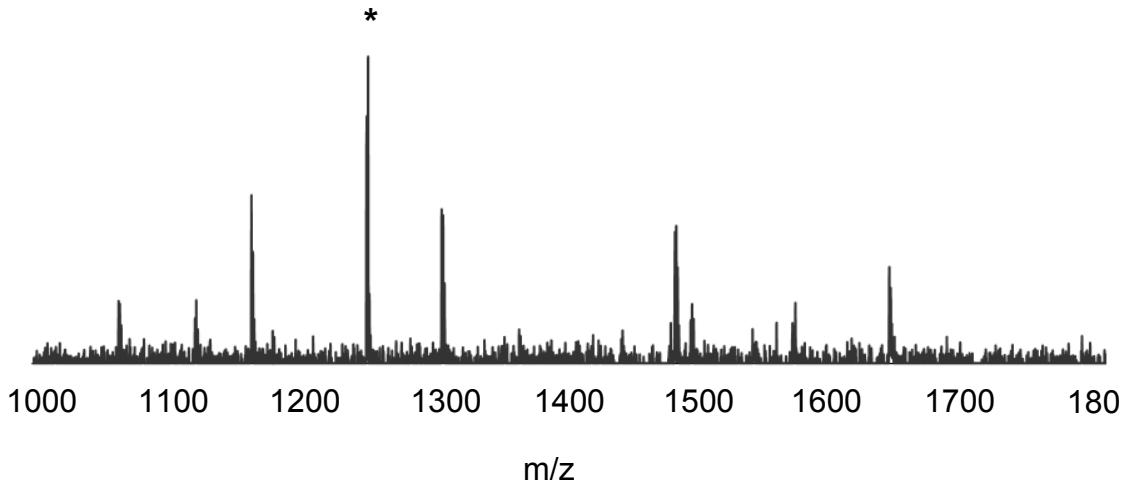


Figure 1.3: A peptide mass fingerprint of bovine serum albumin (BSA) obtained from a mass spectrometer equipped with a MALDI source. Each peak corresponds to a peptide generated from the digested protein. For example, the peptide peak marked with an asterisk corresponds to the peptide sequence VLASSARQRLR. The digestive enzyme used was trypsin and the matrix molecule used was HCCA. This fingerprint was obtained experimentally by Diane Botelho.

Table 1.3: Databases and their website addresses available for PMF of MS spectra.

Database Provider	Application	Website Address
Swiss Institute of Bioinformatics	ExPASy Proteomics Tools	http://ca.expasy.org/tools/findpept.html **(for unspecified cleavages)**
Swiss Institute of Bioinformatics	ExPASy Proteomics Tools	http://expasy.org/tools/findmod/ **(for specified cleavages)**
Matrix Science	MASCOT	http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF
Laboratory of Mass Spectrometry and Gaseous Ion Chemistry	PROFOUND	http://prowl.rockefeller.edu/prowl-cgi/profound.exe
University of California	MS-Fit	http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard

In contrast to PMF, peptide mass sequencing is a tandem MS (MS^2) approach [16]. In this technique, like PMF, the protein is first digested (enzymatically or chemically) to reveal its peptides. However, unlike PMF, digestion of protein mixtures is possible. The peptides are then subjected to separation, typically by LC which is coupled to MS through ESI. Individual ions are then selected in a first stage of MS and subjected to fragmentation, most often by collision induced dissociation (CID) with helium gas within the mass spectrometer. Analysis of these dissociation products in a subsequent

dimension of MS (hence tandem MS) reveals the peptide fragmentation pattern. Fragmentation of the peptide is favoured along the peptide backbone, revealing predominantly b-type and y-type ions [17] due to cleavage of the amide bond (Figure 1.4). The resulting tandem MS spectrum containing multiple fragments (Figure 1.5), with the mass spacing between peaks corresponding to amino acid residues, and so the protein can be sequenced.

Figure 1.4: Schematic of a peptide showing the terminology used when fragmented along the peptide backbone. For example, the X_3 term indicates that fragmentation occurred between the R component of the amino acid to the left and the carbonyl carbon of the amide bond and that the charge remained on the carboxyl terminus following fragmentation. The 3 indicates that the fragment contains 3 amino acid residues (namely R_2, R_3 and R_4).

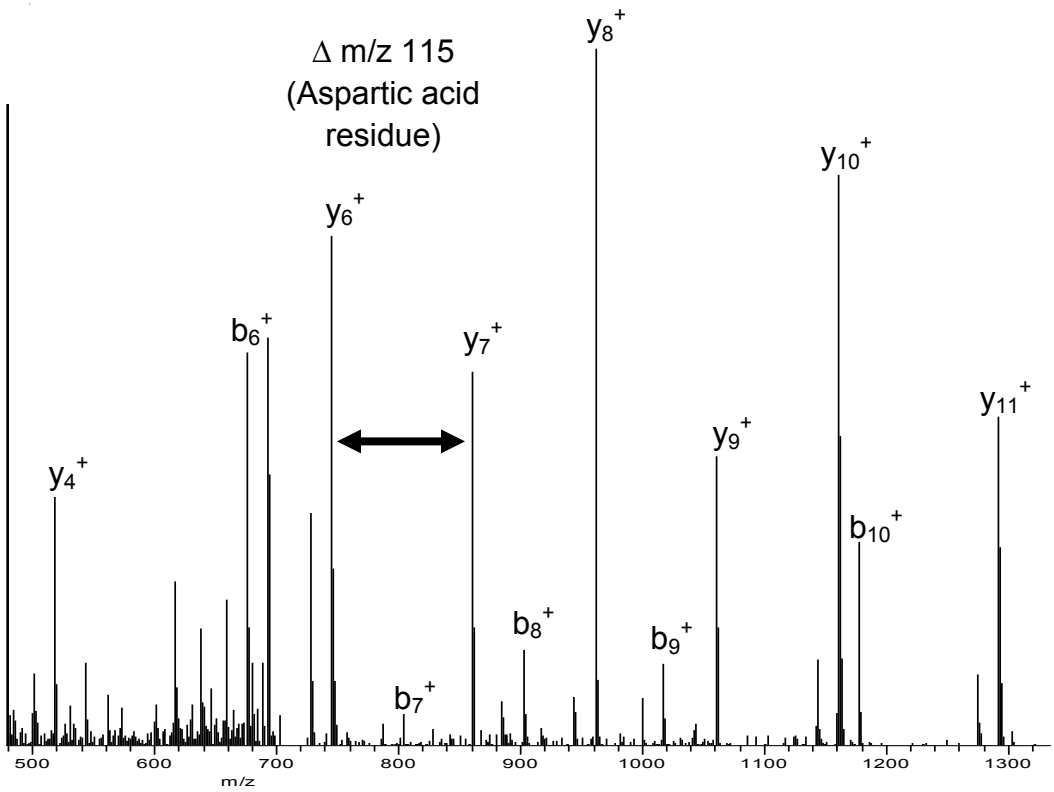


Figure 1.5: An example of a peptide mass sequence spectrum obtained from the fragmentation of a peptide from a urine protein tryptic digest. The peaks resulting from the fragmentation differ by masses corresponding to amino acid residues. The peaks with *m/z* 860.4 and 745.4 differ by an *m/z* value of 115 which corresponds to an aspartic acid residue. The protein was identified as apolipoprotein D. This spectrum was obtained experimentally by Diane Botelho.

Peptide mass sequencing and PMF strategies are collectively referred to as bottom up proteome analysis. A third tandem MS-based strategy that exists for protein identification involves fragmentation of larger undigested proteins. This top down approach to proteomics has only recently become a viable high-throughput approach to proteome analysis [18], owing to developments in high resolution mass spectrometry

[19,20], software for fragment interpretation [21], and, as described in this thesis, MS-compatible separation platforms for intact proteome prefractionation.

1.4 Methods for Protein Quantitation

Inevitably, as soon as MS permitted the sequencing of proteins to determine their identity, the next challenge in proteomics would be to determine the quantity of these proteins. Not only is the identity of a protein crucial for proteomic investigation, including biomarker studies, being able to determine the quantity of each protein is equally significant. Knowing the quantity of proteins and perhaps most importantly, how protein concentrations change between respective biological states, is critical to biomedical advancements.

There are several methods currently used for the quantitation of proteins, which can generally be classified as either relative or absolute quantitation methods. Relative quantitation involves determining the quantity of a given protein from a certain biological state relative to the same protein as expressed in another (control) biological state. When performing relative quantitation of proteins, the exact concentration is not a concern, rather the 'fold' change (*i.e.* the degree of up- or down-regulation) compared to the control is determined. Many biomarker studies use relative quantitation in order to determine candidate biomarker proteins. Such studies often describe quantified proteins as being "ten-fold down-regulated" or "two-fold up- regulated".

Absolute quantitation is a method in which proteins are quantified by determining the specific concentration of the protein. In order to conduct this type of quantitation a

standard must be employed to determine the concentration of the protein in the sample [22]. Other samples can then be compared against the original control in a relative sense to determine specific concentrations.

While selected ion monitoring (SIM) is possible for protein quantitation [23] typically a multiple reaction monitoring (MRM) experiment [24] is employed. In this type of platform, the proteome is digested into peptides, and separated by LC. Knowing the mass and fragment behaviour of the peptide of interest (the one to be quantified), the mass of the peptide is chosen (the parent ion), which is then fragmented and one of its fragment ions is selected for detection and quantitative measurement. `Multiple', in multiple reaction monitoring implies that this process occurs for several peptides, thus the mass spectrometer can rapidly cycle through a list of parent-fragment ions while ignoring all other masses, hence the term multiple reaction monitoring. MRM is favoured among many proteomics/clinical settings due to the fact that this type of experiment can be performed rather rapidly on complex samples. However, the strategy implies that the proteins of interest are already known, and thus represents a downstream component of the protein biomarker discovery process (validation stage).

Another very interesting technique for protein quantitation is the stable isotope labelling by amino acids (SILAC) technique [25]. In such an approach, isotopically labelled amino acids (usually lysine and arginine) are incorporated into the growth medium of an organism (bacteria for example). As a result, the organism incorporates these amino acids into their protein structure and thus may be used as internal standards for comparison against unlabeled proteins.

Proteolytic labelling of proteins is another common technique to protein quantitation [26]. This technique is a fairly simple one, where ^{18}O (from heavy water) is exchanged for ^{16}O at the carboxyl terminus of peptides as they are digested by trypsin. The result is a mass shift of 2 Da for every ^{18}O atom that is incorporated into the peptide, which can be read and quantified following MS analysis.

Aside from the above mentioned quantitative techniques, there are chemical techniques in which a protein or peptide is chemically reacted with a "tag" that can be used to distinguish peptides from one another. An example of this is the isotope coded affinity tag (ICAT) approach in which a chemical moiety is reacted with a protein, creating a mass shift specific to the tag used [27]. The tags are available in a heavy and light version, and so the differences in intensities could be directly seen by MS. There are many commercially available ICAT tags, with the original tag shown in Figure 1.6.

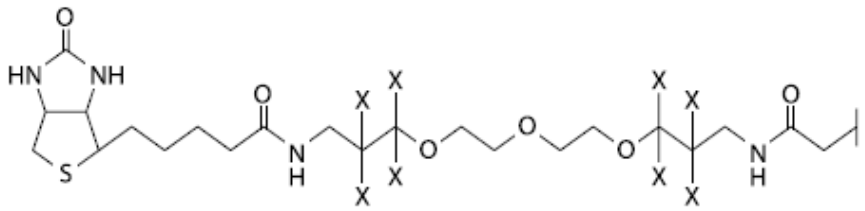


Figure 1.6: The first ICAT tag developed for protein quantitation. X = H (light) or D (heavy tag) This tag is thiol-specific, in which it reacts specifically with cysteine residues [27]. The biotin moiety allows for affinity capture of the tagged peptides.

Isobaric tagging for relative and absolute quantitation (iTRAQ) is also a popular approach for protein quantitation [28]. Each protein tag essentially has two components; a reporter component and a mass balance component. The reporter groups all have a mass that differ by 1, while the mass balance component contains a mass such that the entire tag has the same mass as all the other tags (Figure 1.7). Hence the labels are isobaric, meaning that they have the same mass. The entire component (reporter and mass balance) is chemically reacted with the protein. Thus, during intial MS scanning, all labelled peptides appear as one peak (Figure 1.8a). It is only upon MS^2 that the label is released by the peptide and thus can be distinguished in the mass spectrum (Figure 1.8b). Currently, there are two forms of iTRAQ kits commercially available, the '4-plex' which provides 4 reporter ions and the '8-plex'version.

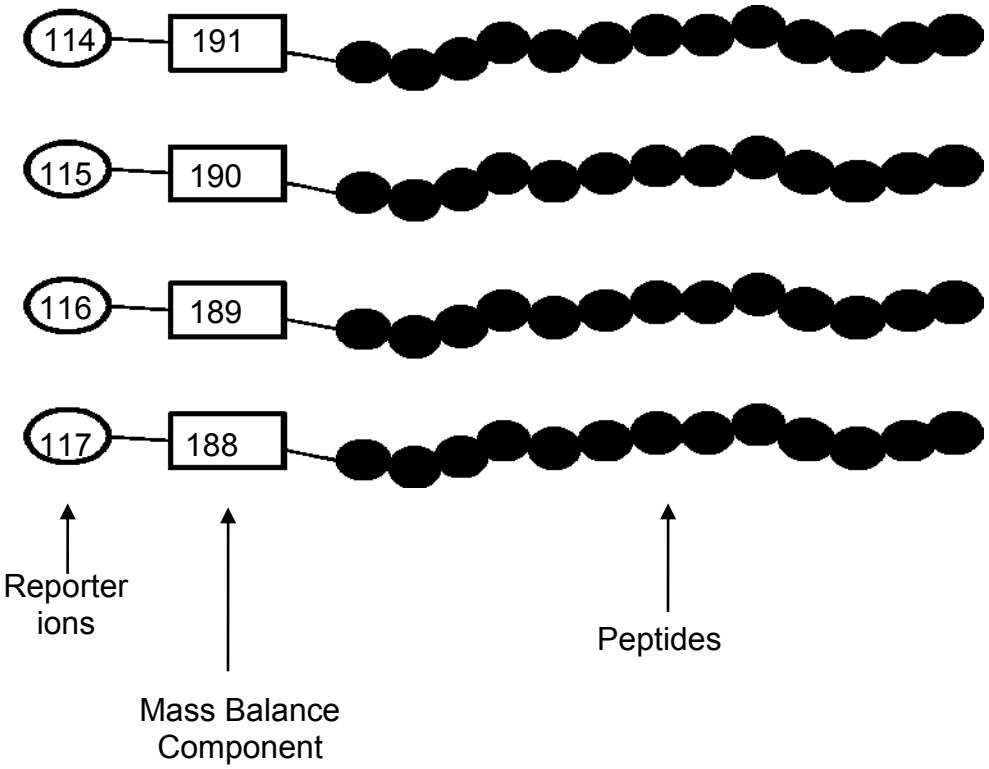


Figure 1.7: The iTRAQ tag strategy showing the peptide, mass balance and reporter ion components. This figure displays the 4-plex strategy, with only 4 reporter ions being generated.

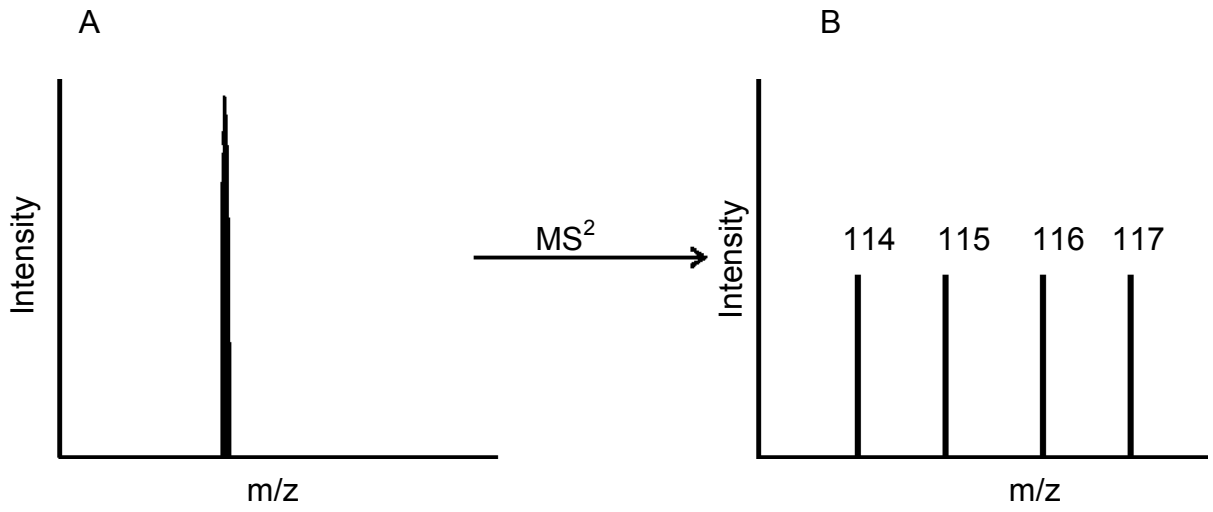


Figure 1.8: MS peaks for iTRAQ labelled peptides. The mass balance component ensure all peptides have the same mass and hence appear in the mass spectrum as one peak (A). Upon fragmentation of the peptides, the reporter components are released (B) and the intensities of these reporter ions can be used to estimate protein quantities.

Several alternative isotopic labelling techniques have been proposed for protein quantitation, and are generally based on the ICAT or iTRAQ strategies, but with differing reactive chemistries for mass separation. For a comprehensice review on these quantitation techniques, one may refer to M. H. Elliot *et al.* [29].

An example of an amine-specific labelling strategy (lysine residues and the N-terminus) incorporates the use of formaldehyde to dimethylate the amines [30]. The methyl groups can either be light, containing regular hydrogen atoms, or heavy, containing deuterium atoms (Figure 1.9). A simple chemical protocol to incorporate formaldehyde labels was provided by Melanson *et al.* [31]. The procedure involves the digestion of proteins from two different samples into their corresponding peptides followed by reductive methylation of 1º peptide amino groups. For one sample, the

reduction is carried out with deuterium labelled formaldehyde, while for the other sample reduction is carried out with normal formaldehyde. Once the amino groups on the peptides from each sample are reduced, the samples are combined and analyzed by mass spectrometry.

The formaldehyde labelled peptides show up in a mass spectrum at a defined *m/z* value whereas the deuterium labelled peptides will show up exactly 4 *m/z* units higher per 1º amino group than the regularly labelled peptide. The resulting mass shift due to the labelling process is 28 amu per labelled residue (including N-terminus and lysine residues) for the light label and 32 amu for the heavy. Thus, upon examination of the mass spectrum, peaks are observed to be 4n mass units apart, where n is the number of lysine residues labelled (plus one accounted for by the N-terminus). This allows for the same peptides to be distinguished in a mass spectrum (Figure 1.10).

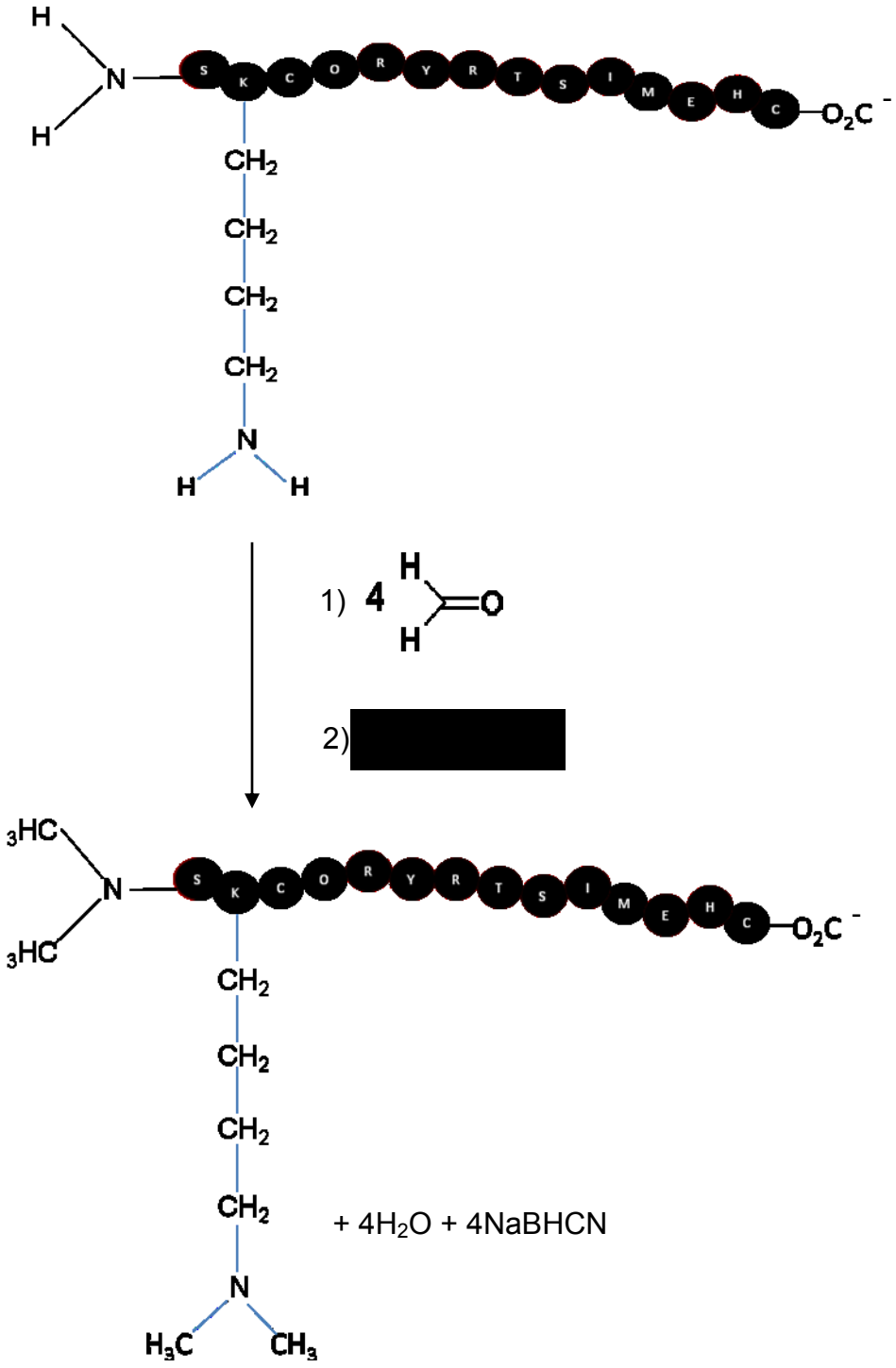


Figure 1.9: Isotope labelling of peptides using formaldehyde (light version shown). Deuterated formaldehyde can also be used to provide a heavy version of the peptide. The amino terminus and lysine side chain are dimethylated.

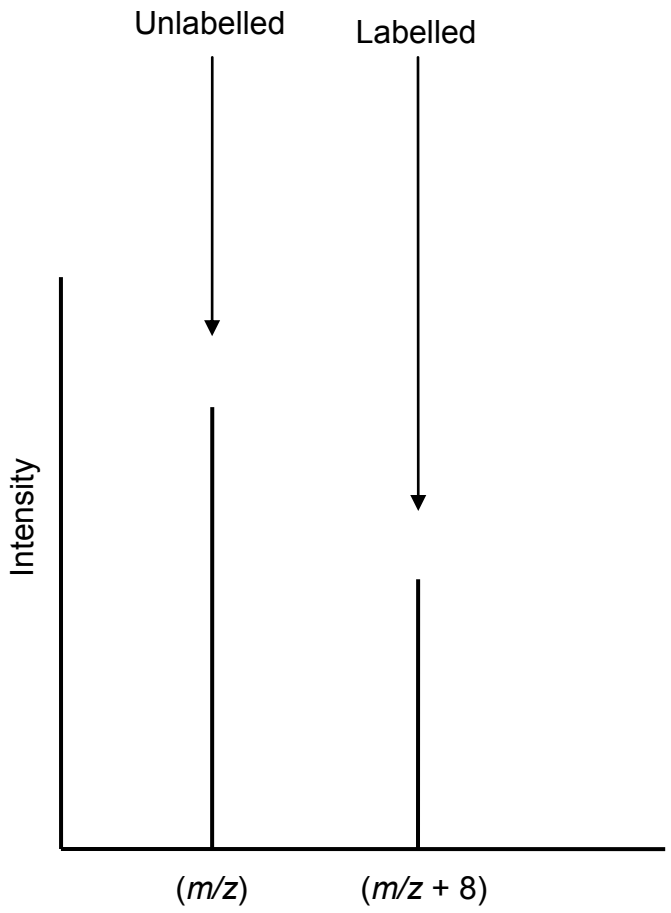


Figure 1.10: A theoretical mass spectrum that would result from formaldehyde and deuterated formaldehyde labelled peptides, where the peptides only contain one labelled lysine residue (mass shift of 4) and a labelled N-terminus (additional mass shift of 4). For simplicity, isotopes from each peptide are omitted and only a +1 peptide is shown.

1.5 Challenges in Proteomics

Proteome analysis presents numerous challenges for sample characterization. First, a typical proteome can contain thousands to hundreds of thousands of proteins (including modifications) and so MS identification of each protein is only possible following extensive separation. Furthermore, protein concentration varies considerably within a sample. Taking blood as an example, the human proteome has an incredibly high dynamic range. The serum albumin protein can be as high as 10^9 pg/mL while interleukin-6 can be as low as 0-0.5 pg/mL [1]. One can imagine that the abundant albumin protein would simply mask detection of interleukin-6, meaning that separation is critical to allow detection of the lower abundance proteins.

Given the need for fractionation to improve detection of low abundance proteins, this can also pose an additional challenge, namely the throughput of detection. Because a typical proteome is so complex, one may generate upwards of one hundred fractions from a single sample, with each of those fractions requiring analysis by MS or more typically, LC-MS. As a result, MS instruments may acquire data for hundreds of hours in order to process an entire set of fractions from this one experiment.

While it is rarely discussed, those who regularly work with analytical instrumentation will realize that the lengthy nature of the experiment poses yet another analytical challenge - that of instrument maintenance. LC-MS instruments are designed to function in a fully automated fashion. Nonetheless, high performance LC (HPLC) systems can easily become clogged and mass spectrometers require routine cleaning. Instruments do occasionally break down midway through an analysis. It is extremely

important that instrumental checkpoints (*e.g.*, running control samples) are in place, and that proper sample treatment (*e.g.*, sample cleanup) is employed to ensure the continued operation of the instrumentation.

An additional challenge in proteomics is that proteins have very different chemical properties from one another. For example, membrane proteins are often only soluble in a non-aqueous solvent system, or by the use of detergents, thereby creating a sampling bias if one were to solubilize the proteome in a purely aqueous buffer. Moreover, structure and function of a protein can vary depending on its environment. Most relevant is the fact that many proteins can exist in more than one form due to post translational modifications (PTM'S).

PTM's are a remarkable example of how nature makes use of chemistry; many physiological disorders are due to a PTM of a particular protein [32]. Perhaps one of the most classic examples of this is the abnormal phosphorylation of the tau protein found in the brains of Alzheimer's patients [33]. There are numerous types of PTM's and a certain protein may posess one or more of these modifications concurrently. Some of the most common PTM's observed in proteomics are listed in Table 1.4. All the above mentioned challenges make proteome analysis extremely difficult. As a consequence, there is currently no single universally accepted method for proteome analysis; proteome researchers can only take advantage of multiple tools and strategies to maximize the likelihood of successful characterization.

Table 1.4: Examples of commmonly observed post translational modifications of proteins and the sites which are most probable for these modifications.

Post Translational Modification	Location of Probable Occurance
phosphorylation	serine, tyrosine, threonine or histidine
alkylation	not specific
methylation	lysine or arginine
acetylation	N-terminus or lysine
amidation	C-terminus
hydroxylation	not specific
oxidation	methionine but may be other residues
glycosylation	Nitrogen or oxygen linked

1.6 Approaches for Proteome Fractionation

There are currently many different separation strategies employed prior to MS analysis of a proteome. Each approach has its merits and limitations, depending on the proteome being investigated as well as the instruments being used. Liquid chromatographic techniques are often used for proteome prefractionation and include anion and cation exchange, hydrophobic interaction liquid chromatography (HILIC) and reversed phase liquid chromatography (RPLC). Column chromatography generally exploits the concept that every protein has a given affinity for certain molecules and this affinity is stronger for some molecules over others. The column contains the stationary phase, with which the protein will interact. When a solvent (mobile phase) is passed

through the column (stationary phase) equilibrium partitioning between stationary and mobile phase can be shifted to favour elution of the protein from the column. In many cases, a mobile phase gradient is employed wherein an increasing strength of eluting solvent is pumped through the column. As the strength increases, protein affinity for the mobile phase increases and eventually the protein will elute.

Size exclusion chromatography is an additional form of liquid chromatography which does not relate to affinity for the stationary and mobile phase. The column is packed with a porous polymer material and the mobile phase is passed through the column. Smaller proteins or peptides can enter the pores of the stationary phase while larger ones do not as easily, and the result is a separation based on analyte size. As a general rule larger analyte molecules elute faster than smaller ones.

Immunodepletion chromatography is also increasingly employed for proteome analysis. In this approach, a column is packed with immobilized antibodies with affinity for specific proteins (often the most abundant of the sample). For example, immunodepletion columns are often used to deplete human serum of the most abundant proteins in the sample prior to further sample workup.

An interesting strategy which integrates separation directly with mass spectrometry is the surface-enhanced laser desorption/ ionization (SELDI) mass spectrometric technique [34]. This approach involves a chip-based array containing a chemically functionalized surface (similar in nature to the packing material of LC columns). The functional groups used on the surface depend on the proteins one wishes to capture. The sample of interest is loaded onto the chip with the idea that only the

proteins of interest will adhere to the functionalized groups on the surface. A series of washing steps eliminates other components from the chip. The chip is then subjected to laser desorption ionization and analyzed by MS.

Stacked sorbents are a fairly new approach for fractionation of proteins. This technique involves several, chemically different stacked solid phase columns. The sample is essentially "poured" onto this stack and as the proteins pass through, they adsorb to the different layers. By the time the protein sample makes it through to the last stack, the sample has been extensively simplified [35].

There are several electrophoretic approaches for fractionation. To name a few, these include preparative zone electrophoresis, continuous electrophoresis in free liquid films such as free flow electrophoresis (FFE) and rotationally stabilized electrophoresis, such as the Rotofor and Gradiflow techniques from Bio-Rad. For a more descriptive explanation of these techniques, Righetti provides an exceptional review [36].

1.6.1 One and Two Dimensional Gel Electrophoresis

Among the electrophoretic approaches, perhaps the most popular strategy of proteome prefractionation is gel electrophoresis, including one dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), and two dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (2D-SDS-PAGE). 2D-SDS-PAGE combines isoelectric focusing (IEF) ahead of SDS-PAGE. The SDS-PAGE approach provides high resolution of proteins according to molecular weight, and also provides a visual image of the proteome (following gel staining), which can be quite

favourable, especially for biomarker discovery when one proteome is being compared to another. The separation medium is composed of polyacrylamide, which is formed from the polymerization of acrylamide and bis-acrylamide monomers. The total amount of acrylamide and bis-acrylamide used is often termed %T which denotes the % total acrylamide. % C is used to refer to the total amount of acrylamide only. It is the pores in the gel that proteins travel through when an electric current is applied to the gel.

1.6.2 One Dimensional Polyacrylamide Gel Electrophoresis

One dimensional SDS-PAGE uses a polyacrylamide gel to separate proteins based on their molecular weight. The % T chosen depends on the molecular weight range of the proteins to be separated. For low molecular weight ranges (masses <50 kDa), a large %T is used (18-22%). Broad molecular weight ranges typically require a smaller %T on the order of 12-15%. Gradient gels are available which attain high resolution across an extremely broad MW range. In SDS-PAGE, the protein sample is first combined with an SDS-containing buffer. SDS is necessary in order to give the protein a net negative charge. This occurs due to the fact that SDS binds to proteins in a ratio of 1.4 g of SDS for every 1 g of protein. Samples are loaded into individual wells of the gel, which are located at the cathode end, and a voltage is applied so that proteins migrate through the gel toward the anode. Large proteins migrate slower and so travel less distance through the gel compared to the smaller proteins, enabling separation based on MW (Figure 1.11). While many buffer compositions are possible, the most widely employed protocol for SDS-PAGE is based on the Laemmli method, which uses a Tris/glycine buffer system [37].

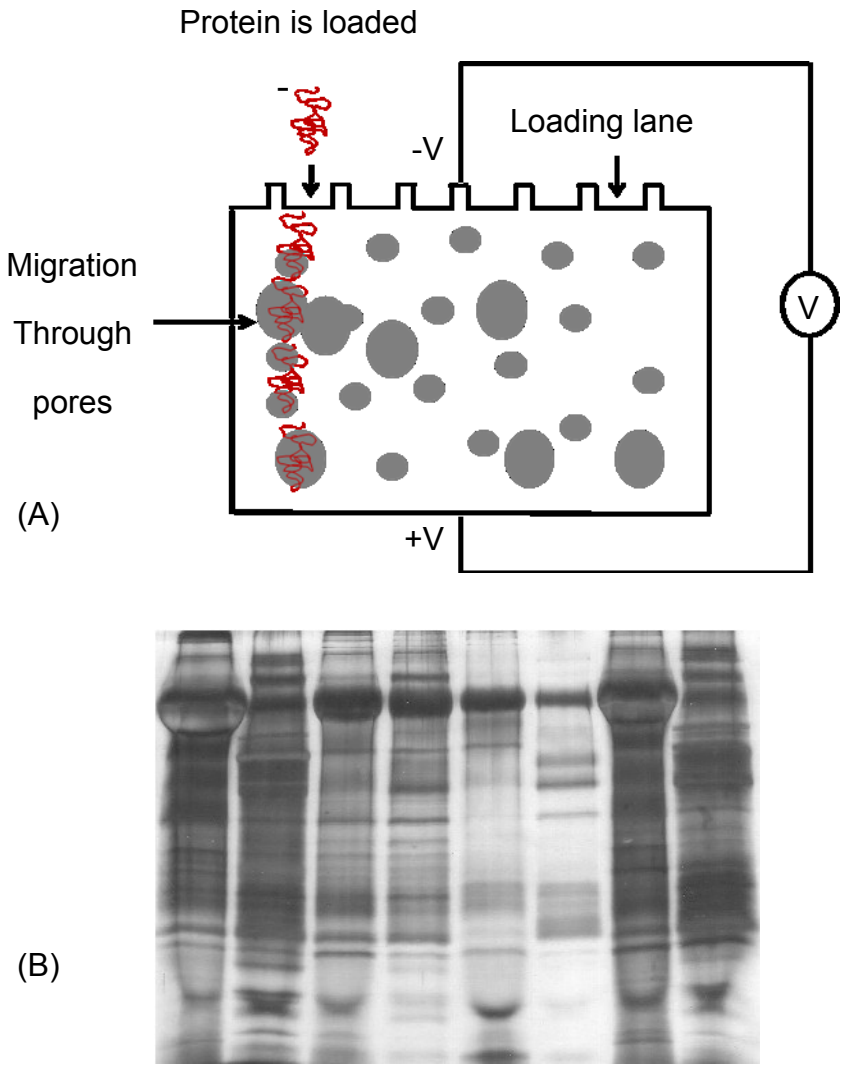
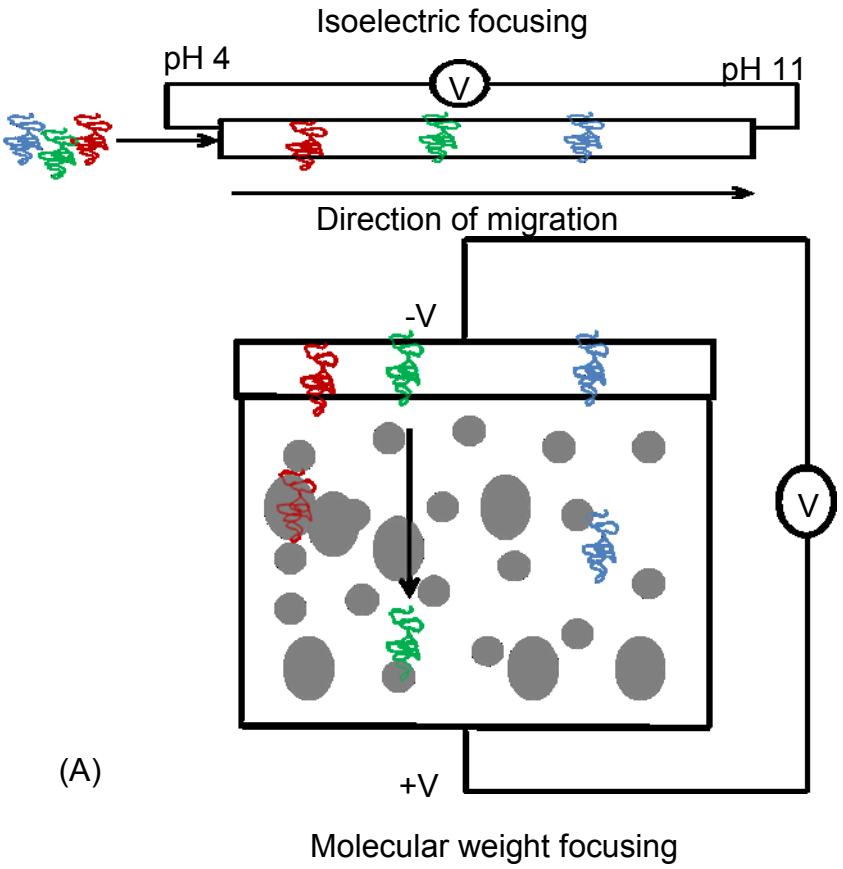


Figure 1.11: (A) Schematic of SDS-PAGE. A negatively charged SDS-bound protein is loaded into a gel lane. As a voltage is applied the protein migrates through the gel pores toward the positive anode. (B) A real image of a SDS-PAGE of urinary proteins. This image was obtained experimentally by Diane Botelho.

1.6.3 Two Dimensional Polyacrylamide Gel Electrophoresis

Two dimensional SDS-PAGE combines an orthogonal dimension of separation, isoelectric focusing, with SDS-PAGE. The net result is that the total peak capacity (a measure of resolution) for a 2D gel becomes the product of the two modes of separation. The isoelectric point (pI) refers to the pH at which the total net charge of a protein is zero. As molecules containing several acid/ base functional groups, proteins exist in solution in multiple charge states depending on the pH. For example, lysine can be positively charged at low pH, while a carboxyl group from aspartic acid can be negatively charged. Of course, the pKa of the respective acid/ base groups determines the degree of protonation at a given pH. During the isoelectric focusing dimension in 2D-SDS-PAGE, a pH gradient is established within the gel strip using either ampholytes, which are small molecular species with varying pH, or immobilized pH strips, in which a pH gradient is chemically immobilized through derivatization of the polyacrylamide gel. Samples are then loaded onto the strip after which a voltage is applied, which causes proteins to migrate until they reach the pH which matches the pI of the protein. At this point the protein becomes neutral and stops migrating. Following isoelectric focusing, the strip is rehydrated with SDS buffer and applied to the top of a polyacrylamide gel. Voltage is applied perpendicular to the strip causing protein separation by molecular weight (Figure 1.12).



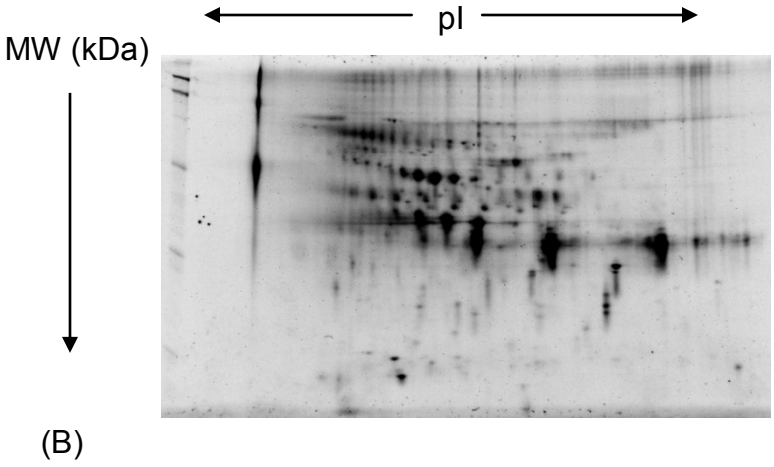


Figure 1.12: (A) Schematic of a 2D-SDS-PAGE. (B) Actual 2D-SDS-PAGE of urinary proteins (obtained by Diane Botelho).

1.7 Top Down and Bottom up Proteomics

As previously mentioned, MS techniques for proteome analysis can be classified as either "top down" or "bottom up" approaches. Each strategy has its merits and demerits. The advantages to the bottom up approach are numerous. As a well characterized and widely applied method, bottom up proteomics makes use of several possible software routines for spectral interpretation. The front-end liquid chromatographic separation of peptides is also well established, achieving high resolution and recovery for the many components of the mixture. The possibility of multiplexed LC platforms also exists (*e.g.* ion exchange coupled with RPLC), and the RPLC technique employs solvents which are directly compatible with MS instruments. With respect to protein quantitation, a bottom up strategy is by far the most favoured approach as it is well known that intact proteins are very difficult to label due to incomplete exposure of all labelling sites [38].

Although bottom up MS platforms possess several attractive qualities, there are disadvantages to such an approach. Most importantly, not all peptides from a given protein are identified. Also, because separation is performed post digestion, peptides from numerous proteins are inevitably mixed together in each fraction. Thus, not only is complete sequence coverage of a protein virtually impossible, low abundant proteins can essentially be missed entirely due to masking from peptides of more abundant proteins. Moreover, due to the fact that low protein sequence coverage is often obtained in a bottom up experiment, PTM's are often completely missed. As previously mentioned, PTM's are often very important for determining the role of proteins in biological

processes such as a physiological disease and so utilizing a method that compromises their detection is indeed a weakness of the experiment.

Top down proteomics offers new insight to the area of proteomics. It is a newer approach and so one disadvantage is that currently only limited software options are available for the interpretation of top down MS spectra. An increase in development of software for interpretation of this type of data is becoming more evident. Kelleher *et al*. [39] is an example of a research group developing intact mass spectrometric software. Generally speaking, only high resolution (high mass accuracy instruments) are suitable for top down MS. However, with developments including the Orbitrap, and other MS systems, these instruments are increasingly becoming available to the average proteomics facility.

A significant advantage to top down proteomics strategies is that essentially complete sequence coverage can be obtained. This is due to the fact that the entire protein is fragmented and offers a strong advantage for PTM analysis. This is especially important for biomarker studies as PTM's are often the key to understanding the role of proteins in a particular disease. In a top down approach, PTM's, and hence possible biomarker candidates are not overlooked. For top down proteome analysis however, perhaps the most limiting aspect is the need for sample preparation directed at intact proteins which is much more difficult than its peptide-level counterpart. Protein solubilization is far more difficult than peptide solubilization; there is typically at least a part of the protein which will be soluble following digestion. In a top down experiment, if the entire protein is not kept in solution, it will not be detected.

CHAPTER 2: URINE PROTEOMICS FOR BIOMARKER DISCOVERY

2.1 Biology of Urine Formation

Considering that this work primarily concerns the analysis of urinary proteins for candidate biomarker discovery, it should be well understood as to the origin and nature of such proteins and thus a comprehensive understanding of the urinary tract is beneficial to understanding the nature of proteins found in urine. This, along with the proposed research directions, is the subject of the present chapter.

Every healthy individual is born with two kidneys. Kidneys are bean shaped organs about the size of an average person's fist. They are located in the back just below the rib cage. Each kidney plays a critical role in filtration of the blood, known as glomerular filtration, ensuring that toxic elements are removed prior to returning the filtered, cleaned blood back to the body. The kidneys are also responsible for maintaining proper pH balance of the blood. The kidney has six main components (Figure 2.1) which all contribute to the glomerular filtration process.

The most critical units within the kidney are the nephrons. These are essentially the 'filters' of the kidney. Each healthy kidney contains approximately $1x10^6$ of these filtering units [40]. Within a nephron is an additional set of units essential to urine formation.

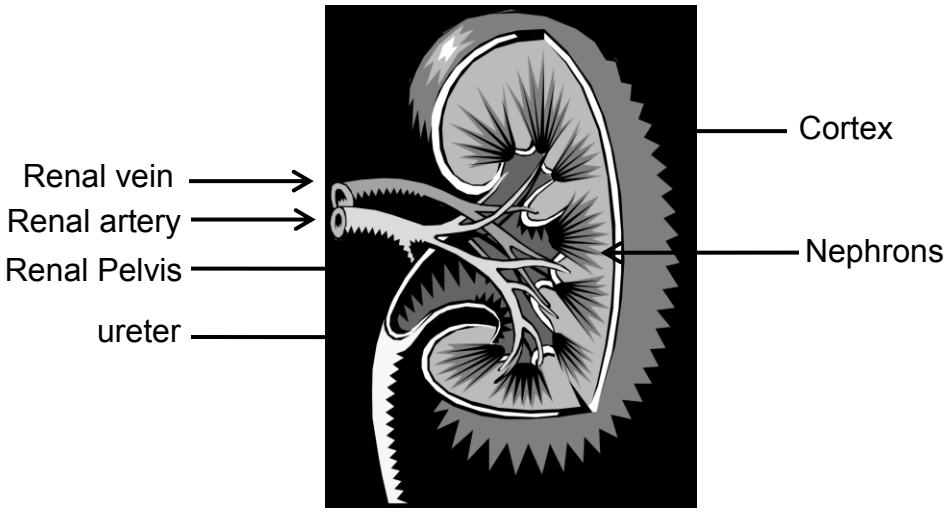


Figure 2.1: A human kidney. The six main parts of the kidney are indicated by arrows [41].

The initial formation of urine begins with blood entering a component of the nephron called the glomerulus. This entry is *via* the renal artery (Figure 2.1). The blood then undergoes ultrafiltration as it passes through a series of semi permeable capillary networks. Ultimately, small solutes such as ions and small molecules are able to pass through the capillary walls, while larger molecules such as proteins are retained. The structure of the capillaries is alone enough to provide the proper environment and pressures needed for the ultrafiltration process to take place.

The filtered blood then travels through the capillary networks and back to the circulatory system *via* the renal vein. The ultrafiltrate (*i.e.* urine) travels through nephron architecture where it then travels into the bladder to be excreted as urine.

2.2 Disorders of the Kidney and Urinary Tract

There are many disorders of the kidney and urinary tract, with varying degrees of severity. Examples of such disorders include Fanconi syndrome, various cancers, kidney failure, various infections, glomerulonephritis, cystinuria, kidney obstructions and many others.

A focus of this work is ureteropelvic junction kidney obstruction (UPJO). This is a disorder in which there is a blockage at the top of the ureter, in an area called the renal pelvis (Figure 2.1), which prevents the flow of urine into the bladder [42,43]. When there is such a blockage, the kidney ceases to function. Amazingly, if one kidney stops functioning there is not much consequence to the body as the other kidney will compensate for this loss. This does not mean however that UPJO is not a concern, it is still vital that the problem be corrected in order to have a completely healthy individual.

Unfortunately, UPJO often goes unnoticed until pain is experienced by the individual (often a child) or until there is blood detected in the child's urine. By this point extensive damage has already occurred to the kidney or urinary tract. Current methods of diagnosing UPJO often involve a computed axial tomography (CAT) scan, ultrasound or magnet resonance imaging (MRI) [43]. In some cases, a more invasive kidney biopsy is required.

The advantages of finding protein biomarkers for UPJO are numerous. First, if one or more protein biomarkers were to be uncovered, children could be tested at the time of birth for the disorder. This would mean that parents and doctors would know

immediately of the problem instead of finding out only when further damage had occurred. Secondly, by having a urinary protein biomarker for the disorder, a simple urine test would be sufficient for diagnosis rather than the expensive or invasive alternatives.

2.3 Proteins Expected to be Found in Urine

When one asks the question of what proteins are expected to be found in urine, it makes sense to think of the path that urine takes from initial entry into the kidney (as blood) to the exiting of the body (as urine) *via* the bladder. In undergoing this path, urine comes into direct contact with all the components of the kidney described above. Thus, in order to understand what proteins may be in urine, it is critical to understand what these components are made up of.

The unfiltered blood first enters the glomerulus. Examination of the glomerulus has shown that it is composed of a layer of endothelial cells which are separated from epithelial cells by a basement membrane [40]. A basement membrane is a sheet of cells and/or fibres that underlies two other kinds of cells [44]. It is composed of extracellular matrix proteins containing collagen type IV, nidogen, agrin, perlecan, glycoproteins and laminin [40]. Thus, some proteins expected from this initial step of urine formation would be epithelial and endothelial proteins as well as the above mentioned basement membrane proteins.

Interestingly, proteins from all organs of the body can essentially enter the bloodstream, making it possible for them to make it through the filtrate barrier and into

the proximal tubule. Although these proteins are usually mid to low molecular weight proteins, it cannot be excluded that fragments of higher molecular weight proteins or perhaps in some instances, intact high molecular weight proteins may make it through the glomerular barrier. Thus, proteins from all organs of the body may be found in urine which explains its source for candidate biomarkers of non-proximal disorders such as sleep apnea in children and lung cancer [45,46].

In addition, proteins from the proximal tubule, ureter, bladder and urethra will also make it into urine. In males, proteins secreted from the vas deferens as well as the prostate gland may also be found in urine. Since these components of the urinary tract are after the nephron, and thus do not undergo glomerular filtration, proteins originating from these components may have a wide range of molecular weights and are not necessarily restricted to smaller masses.

2.4 Methods for the Analysis of Urine Proteins

There are many different approaches currently used for urine proteomics analysis. Given that urine proteins are present in a rather complex matrix, and that the protein concentration is rather low (~0.1 mg/mL) [47], it is important that proper sample preparation be employed to enrich proteins prior to analysis [48].

One very common technique is precipitation of proteins *via* acetone or other organic solvents such as by using tricholoracetic acid, ethanol, acetonitrile or methanol [49]. Other researchers use ultrafiltration or a ligand library bead approach for urine protein preparation [50,51]. Many methods combine molecular weight cut off membranes

with centrifugation to concentrate proteins from urine [47]. Alternatively, solid phase extraction (SPE) has been employed to capture proteins [52].

Following extraction, the resulting proteins are then solubilized and processed by 1D or 2D-SDS-PAGE (most common) or LC-MS^2 strategies [47,48]. If conducting SDS-PAGE, each gel spot is excised and subjected to sequencing *via* MS^2 using either MALDI or ESI. Most often, MALDI is used for a 2D-SDS-PAGE experiment because these techniques are very compatible. It is quite easy to cut out a 2D-SDS-PAGE gel spot, digest the protein and extract the peptides, which are then easily spotted onto a MALDI target. Using this technique, Tantipaiboonwong *et al*. [46] were able to identify potential urinary biomarkers for lung cancer. Moreover, Kumar *et al*. [53] showed by this approach that there were several candidate urine protein biomarkers for a variety of renal disorders. Capillary electrophoresis followed by MS^2 (CE-MS^2) is also another approach utilized by scientists to analyze urine proteins. Meier *et al*. [54] were able to identify over 1000 proteins including candidate biomarkers for Type 1 diabetes using a CE-MS^2 strategy.

A variation of the 2D-SDS-PAGE approach to comparative proteome analysis is the difference gel electrophoresis technique [55]. In this approach, differentially labelled samples are combined into a single gel, and the proteins are distinguished through fluorescence analysis. The advantages are that the proteins from two populations are present in a single gel, and thus alignment issues are no longer a concern. Sharma *et al.* [56] were able to show that there were ninety-nine differentially expressed proteins between healthy individuals and patients who had diabetic nephropathy using this technique.

One of the most comprehensive lists of urinary proteins generated to date has been that from Matthias Mann *et al.* [47] in which over 1500 proteins were identified. The list was established by the accumulation of protein lists generated by SDS-PAGE followed by LC-MS^2 (the so-named 'GeLC' approach) as well as an alternative RPLC separation of intact proteins followed by LC-MS^2. Tandem mass spectra were obtained using a combination of LTQ-FT and LTQ-Orbitrap mass spectrometers. What favoured this technique was that the accuracy of the MS experiments was so high (ppm level) that false positives were essentially eliminated from the list of candidate proteins. The comprehensive analysis of all gel slices in the GeLC strategy also eliminates the need for visualization of gel spots (by gel staining) prior to MS. Thus, low-abundance proteins, which would normally be missed in classical 2D-SDS-PAGE strategies, were detected in this experiment. The list of proteins generated from the research presented by Mann *et al*. [47] has served as a benchmark for the detection and identification of urinary proteins for all researchers.

Most recently, in September of 2009, Alex Kentsis and his research group at the Boston Proteomics Centre identified 2362 proteins in human urine [57]. This impressive list of proteins was generated using ultracentrifugation, gel electrophoresis, ion exchange and reversed phase chromatography as prefractionation techniques. Each fraction was then analyzed by high accuracy linear orbitrap MS^2. Like that of Mann's experiment described above, Kentsis' experiment is indeed impressive, considering that it undoubtedly took hundreds, if not thousands of hours of MS instrument time to process such an incredible number of fractions generated by extensive fractionation. The ever-expanding list of identified proteins in urine points to the richness of this fluid as a source

of candidate biomarkers. Unfortunately, this also points to a sample complexity which challenges proteome analysis of this fluid.

2.5 Candidate Biomarkers Found in Urine

The term 'candidate' in this heading is used with ultimate conviction. The challenge in urine biomarker discovery has not been necessarily the detection of a biomarker, rather *validation* of these biomarkers between different laboratories.

To provide a strict definition, a urinary protein biomarker, or biomarker panel (multiple species) is a protein species that reliably indicates a particular physiological disorder to a given degree of specificity. This biomarker should thus be present across a population of individuals with this same disorder. False positive and/or false negative rates are to be expected, and thus a more useful biomarker would have clear indications as to the reliability of the test. Moreover, this biomarker should be detectable by any laboratory across the world given that the sample being analyzed is indeed of the particular biological state of interest.

Interestingly, urine has been a fluid used for biomarker discovery for approximately twenty-four hundred years. It was in 400 BC when the Greek physician, Hippocrates, made a connection between the foamy appearance of urine and kidney disease [40]. In present day, it is now known that this foamy appearance is due to a condition called proteinuria. This condition is one in which there is more protein secreted into the urine than there should be. Causes for proteinuria can vary from intense exercise and fevers to more serious illnesses such as kidney malfunctions [40]. Although

Hippocrates did not know the origin of this foamy appearance it is albeit quite impressive that urine was used as a biomarker fluid such a long time ago.

Presently, urine has become an extremely valuable body fluid for biomarker discovery. In the Web of Science database alone, there are almost 2300 results when searching the keywords "urinary biomarkers" (as of February, 2010). There is no debate; urine is indeed a fluid that is able to provide important diagnostic information about certain biological disorders and/or conditions. For example, human chorionic gonadotropin in urine indicates pregnancy [58], blood in urine indicates kidney malfunctions or renal injuries [59], and excessive protein in urine indicates glomerular malfunctions [60] or possible blood pressure complications [61]. This list can go on for the importance of urine as a diagnostic fluid. Moreover, not only is urine a powerful diagnostic fluid in a 'dip stick' fashion, as already mentioned, biomarkers for other urinary tract disorders, various cancers, and even sleep disorders have become evident in urine [45,46,53].

To mention specifically all of the candidate urinary protein biomarkers for various disorders would be too extensive. However, an example of some candidate biomarkers can be seen in Table 2.1. It is noted that the results of these studies still classify these proteins as 'candidate' biomarkers, in that they have yet to be validated across a wider population. It is expected that many such studies are ongoing, and that only a small percentage of these proteins will stand up to the rigours of clinical testing.

Table 2.1: A very small selection of candidate biomarker proteins from urine, as described in the literature. The disorders are listed alongside the corresponding reference.

Urinary Protein Biomarker	Biomarker Significance [ref]
mannose-6-phosphate modified	lysosomal disorders [62]
uroplakins	urinary tract malformations [63]
54 proteins	unique to type 1 diabetes [54]
59 different proteins	immunoglobulin nephropathy [64]
34 proteins	membranous nephropathy [65]
interleukin 17	renal allograft rejection [66]
calgranulin B/MRP-14	prostate cancer [67]
19 possible proteins	ureteropelvic obstruction [43]
4 proteins	acute renal injury [68]
exosomal feutin A	acute kidney injury [69]
gelsolin and perlecan	obstructive sleep apnea [45]
alpha-1-microglobulin	vesicoureteral reflux [70]
transforming growth factor-beta 1	upper urinary tract obstruction [71]
3 urinary proteins	lung cancer [46]
N-acetyl-β-D-glucosaminidase	tubular injury [72]
interleukin-6	reflux nephropathy [73]
ADAM 12	breast cancer [74]
5 urinary proteins	different renal disorders [53]

2.6 Challenges in Urine Proteomics

In addition to the challenges outlined in chapter 1, *urine* proteomics can be particularly challenging for a number of reasons. First, urine protein concentration is quite low. In a healthy individual, the protein content is not higher than 150 mg per day. This would translate into approximately 0.1 mg of protein per millilitre of urine. Furthermore, of this 150 mg, approximately 20 mg is serum albumin [40]. Often, this

abundance of serum albumin is a nuisance as it interferes with the detection of other less abundant proteins. By far the most abundant protein in urine is uromodulin, which constitutes approximately 50% of the total protein concentration. The biological significance of this protein remains speculative, though it is thought that this protein has influence on defending the urinary tract against bacterial infections [75,76].

In addition to low protein concentrations, urine proteomics is challenging due to the fact that the urine proteome profile can vary significantly from one healthy individual to the next. Moreover, urinary protein profiles will even vary within the same individual, depending simply on the time of day the sample was collected as well as what the person may have eaten or drank that day [77]. As one can imagine, this great variability poses a challenge for the creation of a standard urine proteome map which all researchers can consider valid and reliable.

Another challenging factor is that urine samples have high salt content, thus electrophoretic methods such as SDS-PAGE and 2D-SDS-PAGE can be quite complicating. Protein migration is compromised due to the fact that salts and other ions are migrating as well. As already described above, researchers attempt to purify the proteins from these salts either by solid phase extraction (SPE) techniques or by precipitating the proteins using organic solvents such as acetone, methanol, acetonitrile or chloroform. However, precipitation of the proteins induces another problem, which is that of incomplete recovery of all proteins. This is particularly challenging if one is trying to do biomarker comparisons, as one sample may appear altered from the other simply due to differences in the degree of precipitation of the proteins from each sample.

Even more of a challenge is the fact that lists of identified urinary proteins varies from one laboratory to another depending on the method of analysis. This is common to proteomics in general, and is a consequence of incomplete proteome detection. An example of this can be seen by Mann and co-workers [47] whose lab was able to generate a list of over 1500 identified proteins with high confidence. It was indicated that this list is a very good start to the "standard proteome list". However, most research labs have not been able to reproduce this same list. This is primarily due to the fact that every lab has their favourite proteomics workflow, confounded by biological differences which are intrinsic to urine. It remains a question as to which method is best for the analysis of urinary proteins. All too often it is seen that different laboratories make their own assessments of sample preparation and analysis, with little or no regards to other available laboratories techniques. A "universal" proteomics platform would represent an ideal situation, but given the current landscape this is certainly something which would be met with considerable debate.

In an attempt to universalize urine proteomics approaches, a group of scientists spanning several countries developed the human kidney and urine proteome project (HKUPP). This organization strives to universalize urine proteomics. Specifically, HKUPP has the following initiatives [78]:

1. "To facilitate proteome analysis and foster proteomics researchers in nephrology."
2. "To understand kidney functions and pathophysiology of kidney diseases."

3. "To provide kidney and urine proteome databases for discovery of biomarkers and drug targets of kidney diseases."

In a report published in the journal Proteomics [79], the HKUPP correspondents stated that the most pressing goal for the HKUPP is "the establishment of standard protocols and guidelines for urine proteome analysis". Thus, it has finally been recognized that without a universal protocol for urine protein preparation, there is ultimately no hope for the generation of a validated list of candidate biomarker proteins. After all, it is this list that all researchers in urine proteomics are striving to achieve.

2.7 Research Proposals

There are a series of research objectives presented in this thesis, each of which is considered to be a critical element contributing to urine proteomics. First, the question of how to collect and initially process urine proteins was a very important one to investigate. Specifically, it is proposed that the sediment protein of urine be considered just as important as the urinary supernatant in proteome analysis.

For a long period of time researchers viewed urinary sediments as salts, urinary crystals and cellular waste that interfere with protein analysis. It is a hypothesis of this thesis that urinary sediments do not just consist of the above mentioned species, but also a large amount of proteins belonging to the urinary proteome. In *all* current protocols for urine protein analysis, sediments are discarded prior to analytical workup. This practice was initiated upon SDS-PAGE practices, as the sediments interfered with the electrophoretic process. However, this practice has extended to all avenues of urine

proteomics, whether gel-based or solution-based. It is believed that discarding sediments is detrimental to urine protein profiling. Work presented in this thesis will provide convincing evidence that complete proteome characterization of urine is only possible upon characterization of urinary sediments along with the supernatant.

Secondly, it is suggested that SDS-based separation platforms, more specifically SDS-PAGE and GELFrEE, are both very comparable techniques for proteome characterization. SDS has a poor reputation amongst proteomics laboratories, as researchers are generally under the impression that it is a severe signal suppressant in mass spectrometry. From a biological point of view, SDS is a very powerful chemical as it aids in protein solubility and has great benefits as the main component in electrophoretic reagents. An SDS-based separation platform, GELFrEE is shown in this thesis to have advantages as it is an intact separation approach. Due to the fact that GELFrEE fractions can contain up to 2% SDS, it was a goal of this thesis to establish the LC-MS tolerance to this species and to assess methods for effective reduction of SDS to the determined threshold level.

Lastly, it was a goal of this research to develop an approach for quantitative proteome analysis. This new strategy required unbiased analysis, and SDS-PAGE and GELFrEE techniques are two such platforms. However, labelling at the peptide level is insufficient as these two separation strategies will not work for peptides. Thus, a new intact protein labelling and separation platform was devised. This method consists of intact dimethylation of proteins followed by GELFrEE-MS^2 for the purpose of protein separation and quantitation. However, a problem exits of incomplete labelling when

considering an intact protein. A possible solution to this is addition of SDS to aid in denaturing the protein prior to labelling. This thesis presents evidence that intact labelling of proteins followed by GELFrEE separation is a possible strategy for the quantitation of proteins.

CHAPTER 3: ANALYSIS OF URINE PROTEINS AND THE IMPORTANCE OF URINE SEDIMENT PROTEINS[1]

3.1 Introduction

As previously mentioned, a variety of proteome detection protocols have been applied to urinalysis. Interestingly, despite the analytical diversity of these protocols, certain actions are universally accepted. For example, to preserve the urine proteome over extended periods of time, the urine is frozen after its collection [80,81]. This is to prevent bacterial growth and protein degradation. Upon thawing, sediments from the sample are typically removed by filtration or centrifugation. Sediment removal is considered a necessary step in proteome visualization by gel electrophoresis [82] and has since been extended to modern gel and gel-free proteome analysis techniques, as it is assumed that the sediments may interfere with subsequent analysis. However, the presence of urinary sediments is quite intriguing, leaving the question of what these sediments are composed of. Figure 3.1 shows the urinary sediments isolated from two thawed urine samples, one from 1 mL of urine (left) and the other isolated from 50 mL of urine (right). Both of these samples have undergone a freeze-thaw cycle.

---

[1] This is in part the pre-peer reviewed version of the following article: Diane Mataija-Botelho, Patrick Murphy, Devanand M. Pinto, Dawn L. MacLellan, Chantale Langlois and Alan A. Doucette. A Qualitative Proteome Investigation of the Sediment Portion of Human Urine: Implications in the Biomarker Discovery Process. *Proteomics, Clin. Appl.*, **2009**, 3, 95-105.

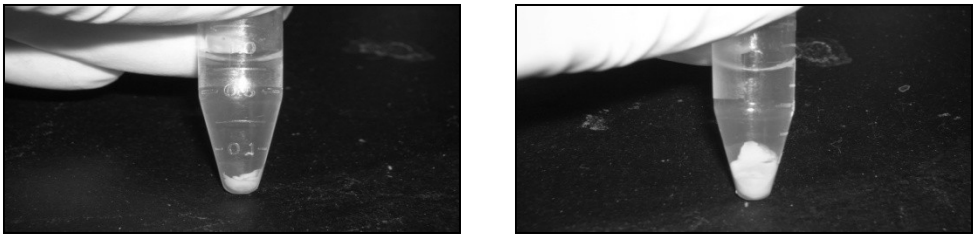


Figure 3.1: Urinary sediments isolated from 1 mL (left) and 50 mL (right) of urine that has undergone a freeze-thaw cycle.

Upon examination of the literature, urinary sediments are defined as being composed of salts and urinary crystals such as calcium oxalate and uric acid crystals. In addition to salts and crystals, urinary sediments are known to contain epithelial cells. However, a proteomic examination of the sediments obtained from urine has never been conducted. An examination of the existence of proteins in this phase of urine is warranted and is discussed in this chapter.

3.2 Materials and Methods

3.2.1 Materials

Reagents for SDS-PAGE and 2-DE were purchased from Bio-Rad (Missisauga, Canada). Solvents were of HPLC grade and were obtained from Fisher (Ottawa, Canada). Protease inhibitor cocktail (cat. # P2714-1BTL) was from Sigma (Oakville Canada), along with trypsin, TPCK treated to reduce chymotrypsin activity (cat. YSC2), and the MALDI matrix, α-cyano-4-hydroxycinnamic acid (HCCA, cat. C2020). HCCA was recrystallized from 60/20/20 ethanol/methanol/water prior to use. The iTRAQ reagents

were generously supplied from the Institute for Marine BioSciences in Halifax, NS. All other chemicals were obtained from Sigma and were used without further purification.

3.2.2 Urine Collection and Preparation for 2D-SDS-PAGE-MALDI-MS^2

Urine samples were collected from four healthy donors (three male, one female), and were frozen at -20^{o}C immediately following collection. Prior to freezing, protease inhibitor cocktail was added to the urine sample at a 1:10 ratio by volume (inhibitor: urine) in order to reduce proteolysis. Urine samples were allowed to thaw at room temperature, without shaking. As required, the sediment phase was separated from the supernatant by centrifugation at 13,000 g for 10 min at room temperature. The supernatant was further filtered through a Whatman 0.22 μm PTFE filter (Fisher) to ensure all sediments were removed.

3.2.3 2D-SDS-PAGE of Urinary Supernatants

Fifteen milliliters of filtered urine supernatant was processed through a 50 mg Bond Elute SPE column (Varian, Mississauga, Canada) according to the method of Smith *et al.* [83]. Protein concentration was assayed using a Bio-Rad assay. For 2D-SDS-PAGE, 100 μg of protein were suspended in rehydration buffer (8 M urea, 4% CHAPS, 2 mM tributylphosphine, 0.2% Bio-Rad Biolyte Carrier Ampholytes pH 5-7, 0.0002% bromophenol blue), and rehydrated overnight onto a Bio-Rad pH 4-7 11 cm strip. Isoelectric focusing (IEF) was performed by applying 250 V (1 h), 500 V (30 min), 8000 V (2 h), 8000 V (50 mA/gel for 35000 Vh). The strips were incubated in two changes of equilibration buffer (6 M urea, 0.0375 M Tris HCl, pH 8.8, 20% glycerol, 2% SDS), with

2 mM tributylphosphine for the first equilibration, and 2.3% iodoacetamide for the second equilibration. The strip was transferred to the top of a Bio-Rad 8-16 %T precast polyacrylamide gel, for the second dimension of electrophoresis (80 min at 160 V). A BenchMark Fluorescent Protein standard ladder was also applied to the gel (Invitrogen, Burlington, Canada), and proteins were stained with Flamingo Fluorescent Gel stain (Bio-Rad). Protein spots were visualized using a Versa Doc Imaging System 3000 (Bio-Rad) using Quantity One 1D Analysis software version 4.5 (Bio-Rad).

3.2.4 2D-SDS-PAGE of Urinary Sediments

The sediment pellet corresponding to the same urine sample was subjected to 2-DE and washed three times with 1 mL of phosphate buffered saline (PBS). The pellet was suspended in 400 µL of rehydration buffer, centrifuged at 5000 rpm for 5 min and desalted by dialyzing overnight with a slide-A-Lyzer mini dialysis unit (Pierce # PI69550, Rockford, USA). 100 µg of protein was subjected to 2-DE according to the running conditions described above for the urine supernatant.

3.2.5 Protein Digestion and Peptide Extraction of 2D-SDS-PAGE Gel Spots

Protein bands were excised from the gels and digested according to standard protocol [84]. Briefly, the samples were reduced and alkylated in gel with dithiothreitol (DTT) and iodoacetamide. Proteins were digested for 24 h at 37 ^{o}C with the addition of 17 µg/mL trypsin in 50 mM NH_4HCO_3, in sufficient amount to just cover the gel slice. Reactions were stopped with the addition of 1% TFA to lower the pH of the sample. The solution layer of each digest was subjected to MS analysis.

3.2.6 MALDI-MS^2 of Proteins Separated by 2D-SDS-PAGE

Digested proteins from 2D gels were analyzed on a vMALDI linear ion trap mass spectrometer (Thermo Fisher, San Jose, CA, USA). Briefly, 5 μL of sample was spotted on a target and allowed to dry. Once dried, 1 μL of HCCA matrix solution (1mg/mL in 50% acetonitrile/water and 50% 0.1% TFA/water) was spotted on top of the sample and allowed to dry. A full MS scan was acquired from *m/z* 1000 to *m/z* 4000. MS^2 data were acquired manually, and subsequently searched using the Bioworks software version 3.1 SR1 (Thermo Fisher) utilizing the SEQUEST algorithm with Uniprot-Human database. A protein was considered to be positively identified if the peptide mass sequence had a score of at least 1.9, a ΔCn value of at least 0.1 and the peptide was found to be fully tryptic.

3.2.7 Urine Collection for ESI-MS^2 Analysis of Samples

Urine samples were collected from healthy male and female donors with approval of the Dalhousie University Ethics Committee (ethics approval #2007-1520). Individuals were determined to be healthy from previous blood tests and had no indications of kidney abnormalities. Immediately following collection, protease inhibitors (cat. P2714-1BTL) were added to the sample (10:1 v/v urine to cocktail) and frozen at – 20°C.

3.2.8 Protein Visualization by SDS-PAGE

If not already suspended in gel solubilization buffer, 40 μL of a processed urine protein sample was combined with 10 μL of 5 × gel solubilization buffer (0.5 M Tris-

HCl, pH 6.8, 10% *w/v* SDS, 0.5% *w/v* bromophenol blue, 26% glycerol). Following 5 min heating at 95^{o}C, 20 μL aliquots were loaded onto individual lanes of a 15% acrylamide gel, The proteins were separated by electrophoresis and visualized *via* silver staining.

3.2.9 Sample Manipulations for Qualitative Mass Spectrometry *via* ESI-MS^2

Four samples (50 mL/sample) from individual healthy donors (two male, two female) were thawed undisturbed at room temperature and centrifuged at 3716 × *g* to pellet the sediments. The supernatants were decanted and the sediments washed with ~5 mL of ice cold water. Each pellet was suspended in 0.5 mL of solubilizing buffer (50 mM DTT, 2 M thiourea, 8 M urea, 1% CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate). Of this, 50 μL was reserved for SDS-PAGE. The remaining ~ 450 μL from each of the sediment samples was pooled and brought to a total volume of 15.0 mL with solubilizing buffer, and subjected to solution isoelectric focusing (sIEF) (section 3.2.13).

The supernatants obtained from these samples were desalted and concentrated by SPE using a C18 cartridge (Varian, Palo Alto, CA, USA). Briefly, ~50 mL of the supernatant was loaded on a SPE column, washed with 20% acetonitrile/ water + 0.1% formic acid, and eluted with 250 μL of 30% acetonitrile, followed by 250 μL of 50% acetonitrile, each in 0.1% formic acid/ water. All samples were pooled, brought to a total volume of 8 mL with solubilizing buffer and subjected to solution isoelectric focusing (section 3.2.13).

3.2.10 Effect of Freezing on Amount of Sediment Proteins

Immediately upon collection, four 1 mL urine samples, obtained from four donors, were centrifuged (3716 × *g*, 6 min) to obtain the sediments present prior to freezing. The supernatant was removed and both the sediments and supernatants were frozen overnight at – 20 ºC. Samples were thawed at room temperature and the sediments which accumulated in the solution fraction during the freeze-thaw cycle were again collected by centrifugation. Sediments obtained before and after freezing were individually solubilized in 60 µL of 1 × gel solubilization buffer (section 3.2.8), heated for 5 min at 95ºC and 30 µL of each sample was subjected to SDS-PAGE (section 3.2.8).

3.2.11 Effect of Vortexing on Protein Distribution Between Phases

Samples of urine, obtained from four donors (three male, one female), were thawed and aliquoted into 10 mL volumes, with gentle agitation to suspend the sediments between transfers. Samples were then either vortexed for 2 min at full power on a VWR mini vortexer (Equipland, Fremont, CA, USA), or left undisturbed. The samples were centrifuged at 3716 × *g* for 5 min, the supernatant decanted and the proteins in this phase were acetone precipitated overnight at 4^{o}C using a 3:1 ratio of acetone: urine. All pellets (supernatant and sediment) were solubilized in 5 mL of PBS, and subjected to SDS-PAGE (section 3.2.8).

3.2.12 Effect of pH on the Partitioning of Proteins Between Phases

Two urine samples (~50 mL/sample) from separate donors were thawed, vigorously vortexed for 5 min, pooled, and divided into ten equivalent 1 mL aliquots. Five of these aliquots were used as controls (no pH adjustment) to assess the reproducibility of the visualization protocol. For the remaining five samples, the pH was adjusted over the range 0.5 to 8.0 by addition of 500 µL HCl or NaOH (prepared at different concentrations). All samples were agitated on a vortex mixer, allowed to equilibrate for 15 min, and centrifuged to pellet the insoluble material. The solution phase was carefully removed from the precipitate, and the proteins in these solutions were precipitated with cold acetone as described previously. Along with the original sediment fraction, the acetone-pelleted proteins were suspended in 20 µL of 1 × gel solubilization buffer (section 3.2.8) for visualization by SDS-PAGE.

For quantitative analysis by iTRAQ, a urine sample obtained from a single donor was divided into 4 × 50 mL portions immediately following collection. The pH was adjusted to 8.8 for one sample, pH 3.9 for another, and unaltered at pH 5.9 for the remaining two, by addition of 2 mL of 1 M HCl, 1 M NaOH, or water respectively. Samples were aliquoted into 10 mL fractions and frozen at –20ºC. One fraction of each sample type was allowed to thaw at room temperature, and centrifuged at 3716 × *g* in order to collect the sediments. Sediments were suspended in 1 mL of 1 M urea with gentle heating at 30 ºC to encourage solubilization. Proteins were digested with trypsin as described in section 3.2.5. Digestion was stopped with 100 µL of 10% formic acid.

Samples were centrifuged to pellet any insoluble material, the soluble portion was evaporated to dryness, and subjected to iTRAQ labeling, as described in section 3.2.15.

3.2.13 Solution Isoelectric Focusing and Protein Digestion

Solution isoelectric focusing was performed using a custom designed eight-chamber solution isoelectric focusing device. Details of the apparatus are provided elsewhere [85]. Prior to loading, 200 μL of Bio-Rad carrier ampholytes (pI 3-10) were added per 4 mL of sample. Four hundred microlitres were loaded into each of the 8 chambers, and the device was subjected to 2 W constant power for 1.5 hours to effect protein focusing. The eight resulting fractions were subjected to protein precipitation with cold acetone (overnight, 4^{o}C). These fractions were centrifuged at 3716 × *g* for 10 min to obtain the protein pellets. Pellets were dissolved in 100 μL of water, adjusted to pH 8 with 1 M ammonium bicarbonate and digested with trypsin. Briefly, the proteins were reduced with 200 mM DTT, alkylated with 200 mM iodoacetamide and digested at 37^{o}C overnight at an approximate 10:1 ratio of protein: trypsin.

3.2.14 LC-ESI-MS^2 of Proteins and Data Analysis

Digested proteins were separated with a 1200 HPLC system (Agilent Technologies, Mississauga, Canada) using a Biobasic-C18 column (Thermo Fisher). Solvent B consisted of 0.1% formic acid in acetonitrile, while solvent A was 0.1% formic acid in water. Following a 5 min initial hold at 2% B, the gradient consisted of a ramp from 2 to 40% B over 40 min, then to 80% B in 1 min. The injection volume was 1 μL with a constant flow rate of 2 μL/min. Nanospray mass spectrometry was performed on a

LTQ linear ion trap mass spectrometer (Thermo Fisher). Each fraction was analyzed twice (*i.e.* 2 × 1 μL injections with LC-MS^2), the first run being performed with dynamic exclusion disabled, and in the second, with dynamic exclusion set to 1.5 min. Data were obtained in data dependent mode (MS scan followed by MS^2 of top 5 ions). MS^2 data were searched using the Bioworks Browser software package (Thermo Fisher) against human Uniprot database (14,804 entries). Criteria for positive protein identification were as follows: For all peptides, ΔCn had to be at least 0.1 and Rsp had to be less than or equal to 4. For triply charged peptides, X_{corr} had to be at least 3.75, doubly charged at least 2.2 and singly charged at least 1.9. Peptides also had to have a peptide probability no greater than 0.05. Reversed database searches were performed on each of the runs using identical search criteria to obtain an assessment of false positive scores.

3.2.15 iTRAQ Analysis of Effect of pH on Partitioning of Proteins

The digested sediment proteins were first subjected to reversed phase column cleanup on a Betasil C18 Dash HTS column (20 × 2.1 mm, 5 μm particle size, Thermo Fisher). The fractions containing peptides were pooled, evaporated and subjected to iTRAQ labelling according to the manufacturer's instructions. Label 114 represented the acidified sample (pH 3.9), label 115 and 116 were the pH 5.9 samples, and 117 was the pH 8.8 sample. Following labelling, all samples were combined and again subjected to reversed phase cleanup as described above. Fractions were dissolved in 100 μL of strong cation exchange (SCX) mobile phase A (10 mM ammonium formate, 25% acetonitrile, pH 3) and separated using a polysulfoethyl A 100 × 2.1 mm column with 5 μm particle size (PolyLC, Columbia, MD, USA). The gradient was selected as follows: hold at 0% B

for 5 min (600 mM ammonium formate, 25% acetonitrile, pH 3), to 10% B at 20 min, 20% B at 30 min, 60% B at 40 min, hold to 45 min. The flow rate was 0.2 mL/min and fractions were collected in 2 min intervals from 0 – 45 min. SCX fractions 4-8 were selected for LC-MS^2 analysis due to the fact that these fractions were the only fractions that showed peptide peaks in the SCX chromatogram.

The SCX fractions were dried down and resuspended in 30 µL 5% acetonitrile/0.2% formic acid and subjected to RPLC-MS^2 using a nanoAcquity UPLC system (Waters, Milford, MA, USA) with a Q-TOF premier mass spectrometer (Waters, Milford, MA, USA) equipped with a nanospray source. The column used was a 1.7 µm BEH130 C18 75 µm × 100 mm column (Waters, Milford, MA). The injection volume was 6 µL and the gradient was as follows: 5% B (acetonitrile/0.1% formic acid) to 40% B over 30 min followed by a ramp to 95% B over 2 min. Solvent A was water/ 0.1% formic acid. Mass spectrometry parameters were set to scan from *m/z* 450-1600 for the MS scan followed by MS^2 of the top three ions at a mass range of 100-1500. Data analysis was performed using the MASCOT search engine with SWISSPROT-human database. Peak lists were exported using ProteinLynx version 2.3 (Waters, Milford, MA). Exported lists were combined and searched using the MASCOT search engine. Search criteria were set using iTRAQ modifications for 4 labels, a peptide mass tolerance of 0.8 Da, fragment mass tolerance of 0.6 Da and one missed cleavage. Quantitation of iTRAQ reporter ions was also performed using MASCOT. False positive rates were obtained by searching a reversed database using the exact same search criteria.

3.3 Results and Discussion

3.3.1 A Proteomic Investigation of Urinary Sediments

Upon microscopic examination of urinary sediments stained with crystal violet, urinary crystals and epithelial cells are evident (Figure 3.2). However, it is clear in the photomicrograph that these are not the only species present. Also seen in the photomicrograph are multiple aggregates of an unknown species, which was easily stained by the crystal violet. It was presumed that these species were protein aggregates or clumps that had precipitated from the urinary supernatant.

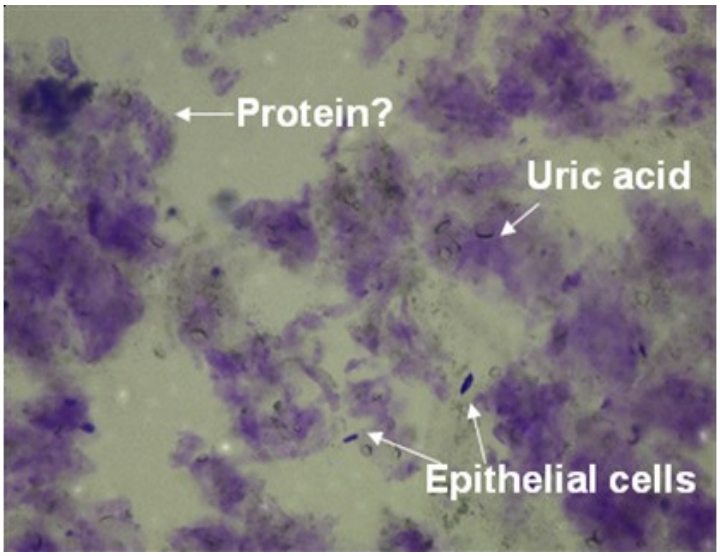


Figure 3.2: A photomicrograph (25 × magnification) of urinary sediments stained with crystal violet. Clearly indicated are uric acid crystals and epithelial cells. Also indicated is the suspected protein.

To see how easily the questionable species dissolved, PBS was added to the sediments and the sample was vortexed. Following the attempted solubilization, the remaining sediments were examined (Figure 3.3). It is clear from the figure that the suspected protein is easily dissolved, leaving behind only uric acid and calcium oxalate crystals.

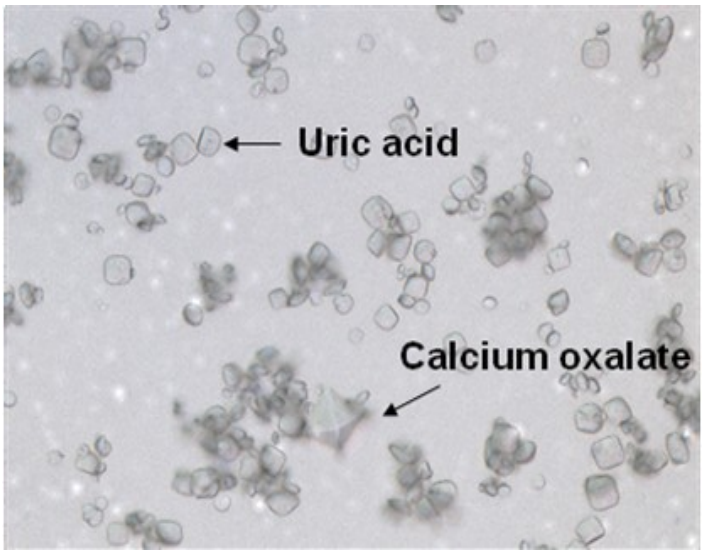


Figure 3.3: Urinary sediments that remained after solubilization with PBS. As seen, the only species remaining are urinary crystals.

The simplest way to determine the origin of the purple clumps (suspected protein) observed in Figure 3.2 is to conduct an SDS-PAGE analysis of the sediments. Presumably, if there are indeed proteins present in the sediments, then they will be evident in a polyacrylamide gel. Figures 3.4 A and B are 1D-SDS-PAGE and 2D-SDS-PAGE images respectively of urinary sediments isolated from urinary supernatants.

(A) (B)

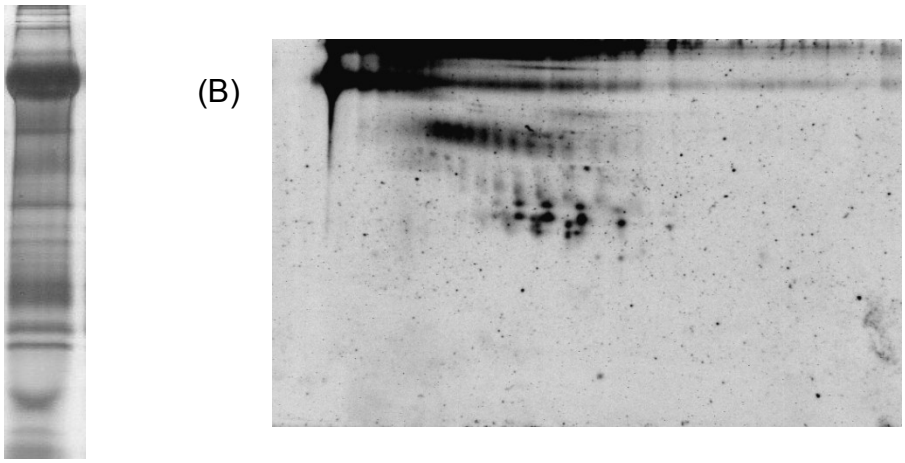


Figure 3.4: (A) SDS-PAGE of urinary sediments. (B) 2D-SDS-PAGE of urinary sediments.

It is clear from the figures that the sediments contain a copious amount of protein. To reinforce this finding, it was reported as early as 1986 that a thawed urine sample should be thoroughly vortexed but *not* centrifuged prior to SDS-PAGE, as this practice would lead to protein loss [86]. The identities of the so-called *lost* proteins were not discussed in that report. However this clearly indicates that even as early as 1986, it was believed that urinary sediments contained protein. More recently, Zhou *et al.* [81] determined that the sediment phase of urine accounts for approximately 50% of the total protein content by mass and is primarily comprised of uromodulin. Furthermore, Zhou *et al.* clearly demonstrated (through SDS-PAGE images) that several other proteins contribute to the sediment phase of urine. However, a comprehensive study to identify these proteins was not conducted.

As a simple prefractionation technique, there is no doubt that removal of the urinary sediment phase can be considered beneficial for the analysis of low-abundant protein components in the urine supernatant [48]. However, it is unnerving that although urinary sediments are routinely analyzed by microscopy as a means of diagnosing physiological conditions [87], current proteomic investigations for biomarkers focus only on the solubilized portion of the sample. It is evident from the SDS-PAGE images that discarding the sediment phase undoubtedly removes other potentially important protein components from the sample.

It should be noted that any variation in protein profiles imposed by the analytical procedure (such as discarding sediments) will result in an unnecessary bias upon detection or quantitation of proteins. That is, a bias induced by a sample handling protocol such as filtering sediments may have significant implications in the biomarker discovery process. Clearly, seemingly differentially expressed proteins may simply correlate to relative changes in protein partitioning between the supernatant and sediments. This partitioning could be due to minor differences in the sample (*e.g.,* pH, protein concentration, extent of sample agitation). Thus, a comprehensive analysis of the urinary proteome, including urinary sediments, would lead to the greatest probability of biomarker discovery and validation.

3.3.2 Examining the Proteins in Urinary Sediments by 2D-SDS-PAGE-MALDI-MS^2

As demonstrated in section 3.3.1, urinary sediments do contain many proteins. As to the types of proteins present or how these proteins compare to the urinary supernatant proteins still remains unclear. Perhaps the most obvious question is how the sediment proteins compare to that of the supernatant proteins. Figure 3.5 displays 2D-SDS-PAGE images of urinary supernatants (top) and corresponding sediments (bottom) isolated from the same sample.

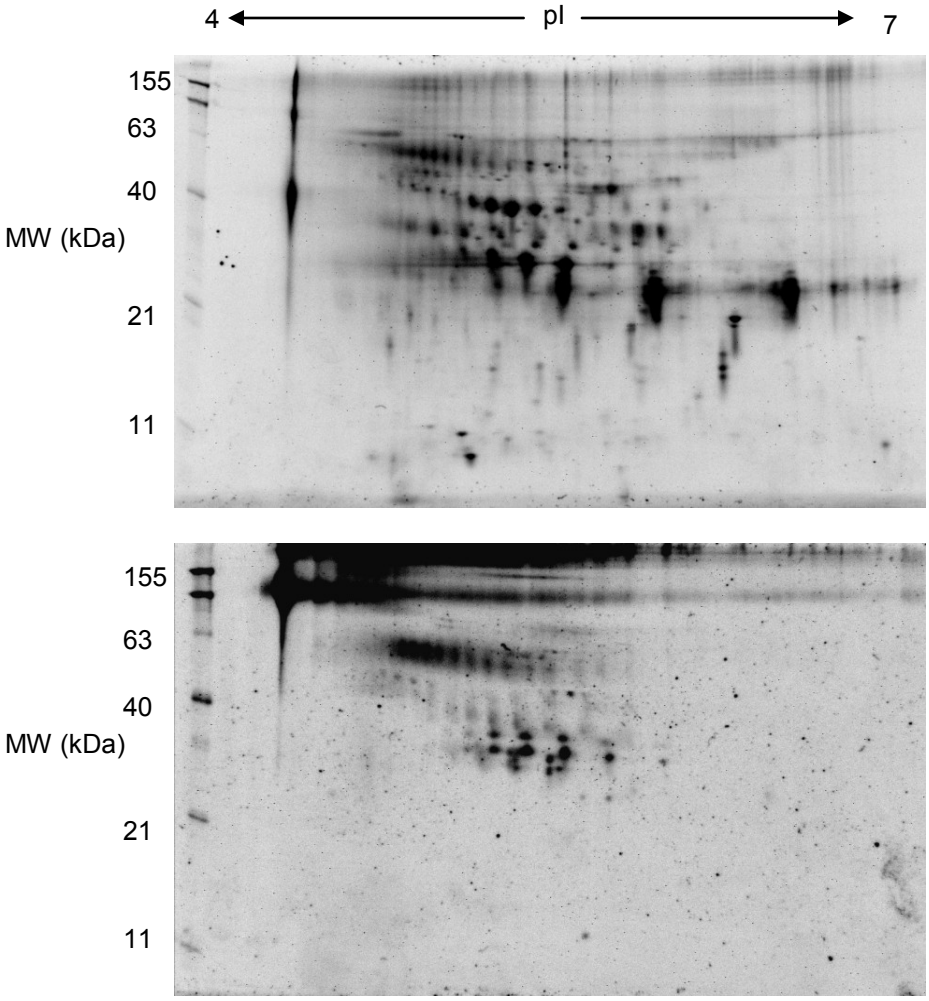


Figure 3.5: 2D-SDS-PAGE or urinary supernatants (top) and sediments (bottom) isolated from the same urine sample.

The 2D-SDS-PAGE images provide a better relative look at the diversity of proteins in urinary sediments, compared with those from urinary supernatant. Many of the protein spots in the sediment sample do not align with spots from the supernatant, indicating that there are different proteins, or at the very least, possible differences in protein isoforms which exist in each phase. These images provide evidence that filtering a urine sample to rid the sediments not only removes salts, epithelial cells and other crystals, but also removes proteins, and in doing so, some urine proteins will go undetected by mass spectrometry. To gain comprehensive knowledge of the urine proteome as a whole, proteins from the sediments should be analyzed along with the supernatant proteins. It is realized that some researchers would prefer to remove the sediments in order to rid the sample of salts and other interfering compounds. However, it must be accepted that the analysis of isolated urinary sediments permits an additional analysis of a subset of proteins present in urine. Such a practice has important implications in biomarker analysis. Detection of a larger number of proteins affords the best possibility of detecting potential biomarker candidates for a particular physiological state.

A total of 71 spots were excised from the 2D-SDS-PAGE gel of urinary sediments, along with 134 spots from that of the supernatant proteins. These spots were subjected to tryptic digestion and MALDI-MS^2 analysis. Of the collected spots, a positive identification was assigned to 46 of these. Accounting for redundant identification of proteins, a total of 37 unique proteins were identified across the two gels, and the results are presented in Table 3.1 (sediments) and Table 3.2 (supernatants). From the 2D gels, 19 unique proteins were identified from the sediments, while 20 proteins were identified in

the supernatant. Of the identified proteins, only 2 were common to both the supernatant and the sediments, with all other identified proteins being unique to each phase.

The number of proteins identified from the 2D-SDS-PAGE gels relative to the number of spots detected, is consistent with previous reports on 2D-SDS-PAGE analysis of urine. For example, Spahr *et al*. were able to identify 38 of 254 excised spots [88]. In a separate study, of 450 distinct spots in a 2D-SDS-PAGE of human urine, 150 proteins were identified. Several other spots were assigned to protein isoforms and degradation products [89].

Table 3.1: Sediment proteins identified from 2D-SDS-PAGE followed by MALDI-MS^2.

Sediment Protein Identified via 2D-SDS-PAGE-MALDI-MS^2	**SwissPROT Accession Number**
Apolipiprotein B-100	P04114
Decraprenyl-diphosphate synthase subunit 2	Q86YH6
DNA topoisomerase 2-alpha	P11388
Gamma-butyrobetaine dioxygenase	O75936
Interleukin-7 Precursor	P13232
Keratin type 2 cytoskeletal 1	P04264
Kininogen-1 Precursor	P01042
Malate dehydrogenase, cytoplasmic	P40925
Mannan-binding lectin serine protease 2	O00187
Multidrug resistance-associated protein 4	O15439
Myosin-5B	Q9ULV0
Serine/threonine protein kinase R1O1	Q9BRS2
Serine/threonine protein kinase LATS2	Q9NRM7
Serine/threonin-protein kinase D1	Q15139
Type 1 cytoskeletal 10	P13645
U5 small nuclear ribonucleoprotein 200 kDa helicase	O75643
Ubiquitin carboxyl-terminal hydrolase	Q9UPU5
Uncharacterized protein C15	Q96M60
Uromodulin precursor	P07911

Table 3.2: Supernatant proteins identified from 2D-SDS-PAGE followed by MALDI-MS^2.

Supernatant Protein Identified via 2D-SDS-PAGE-MALDI-MS^2	**SwissPROT Accession #**
Basement membrane-specific heparan sulphate proteoglycan	P98160
AMBP protein precursor	P02760
Calpain-9	O14815
Calumenin precursor	O43852
Endothelial protein C receptor precursor	Q9UNN8
Gastric inhibitory polypeptide precursor	P09681
Ig kappa chain C region	P01834
Inhibin beta C chain precursor	P55103
Kininogen-1 precursor	P01042
Lethal(3)malignant brain tumor-like protein	Q9Y468
Malate dehydrogenase, cytoplasmic	P40925
NAD kinase	O95544
Probable ATP-dependent RNA helicase	Q9Y2R4
Prostaglandin-H2 D-isomerase precursor	P41222
Secreted and transmembrane protein 1 precursor	Q8WVN6
SH3 domain-binding glutamic acid-rich-like protein 3	Q9H299
Thimet oligopeptidase	P52888
Titin	Q8WZ42
Tubulointerstitial nephritis antigen	Q9UJW2
Zinc-alpha-2-glycoprotein precursor	P25311

An example which illustrates the significance of analyzing the sediment phase for biomarker discovery can be seen with Interleukin-7 (IL-7). This protein was only identified by 2D-SDS-PAGE in the sediment phase of the urine sample. A very interesting fact is that elevated levels of IL-7 in the serum of children have been shown to correlate with elevated platelet counts in children with nephritic syndrome (NS) [90]. However, authors of the study reported that although the IL-7 levels correlated with elevated platelet counts, they do not correlate with urine protein [90]. This latter point is particularly interesting since the above results indicate the presence of IL-7 in the sediments of urine and not the supernatant. It is possible that in the previous study, IL-7 was partially or entirely discarded during urinalysis upon discarding the sediments, thus giving no correlation to urine samples. Thus, if IL-7 levels in serum may be a possible biomarker for NS, it may be analyzed in urine sediments. This is a particularly favourable finding since urine sediment analysis would be a much less invasive procedure than serum analysis.

It is worth noting that of all the proteins identified by 2D-SDS-PAGE-MALDI-MS^2, human serum albumin, being one of the most abundant proteins in urine, was not detected. It is believed that the albumin may have been lost during the solid phase extraction, used for initial sample preparations and so it is likely that other proteins were lost during this stage of extraction. This reinforces the outcome of different sample preparation protocols on the resulting list of identified proteins.

As reported by others, sample preparation, including protein extraction and concentration, is critically important to the success of the urine protein analysis. For this

reason, it was decided that the proteins isolated from urinary sediments and supernatant would be analyzed employing a separate approach, involving acetone precipitation, solution isoelectric focusing and LC-ESI MS^2. Also investigated was the effect of the freeze-thaw cycle on inducing sediment proteins.

3.3.3 Effect of a Freeze-Thaw Cycle on Sediment Proteins

As previously mentioned, with a high risk of proteolysis, urine is inevitably frozen upon collection to help preserve the sample. Although necessary, this manipulation step undesirably alters the native state of the sample. Indeed, Figure 3.6 demonstrates that a simple freeze-thaw cycle significantly increases the sediment portion of urine (up to 50 fold increase by mass, depending on the sample).

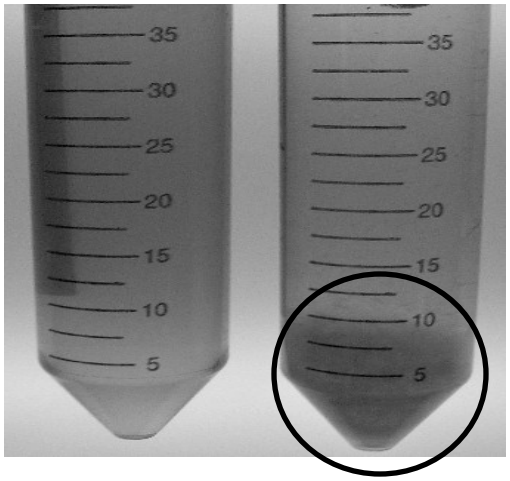


Figure 3.6: Image of human urine before freezing (left) and after freezing (right). The sediments are indicated by the circle.

The sediment portion of urine collected before freezing, along with those which result from a freeze-thaw cycle were subjected to SDS-PAGE for protein visualization. It can be seen from Figure 3.7 that many proteins are present in the sediment phase of urine prior to freezing the sample. This was surprising, as the sediments are not easily seen prior to freezing and it was presumed that they were induced entirely from the free-thaw cycle. Alternatively, sediments were clearly present in a fresh urine sample prior to freezing.

Perhaps more significant is the fact that additional proteins are observed in the sediment portion of urine following a freeze-thaw cycle, despite having completely removed the sediments obtained immediately following collection of the sample (Figure 3.7). Such proteins represent those which are not intrinsic to the sediment portion of urine, upon initial collection of the sample. Rather precipitation of these proteins is

induced by the freeze-thaw cycle. Moreover, it can be seen from the gel image that the sediment proteins vary greatly according to the individual, both before and after a freeze-thaw cycle.

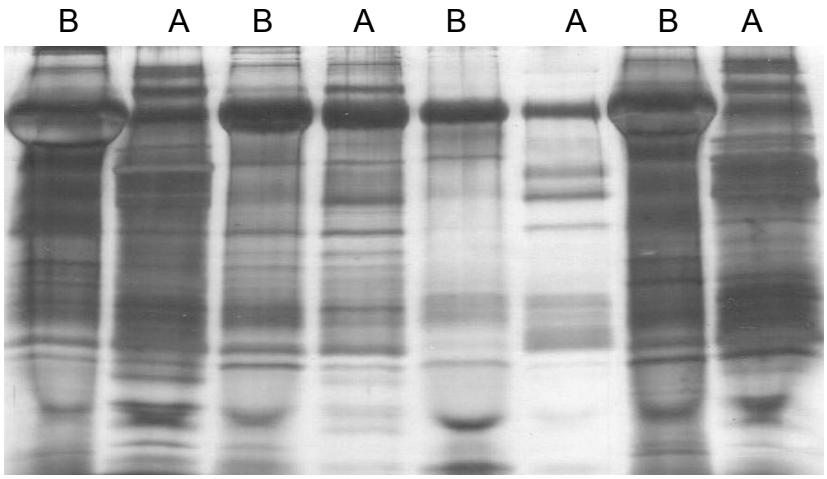


Figure 3.7: Urinary sediment proteins isolated from 4 donors before freezing the sample (B) and again after freezing the sample (A).

Urinary sediments from four individual donors were subjected to SDS-PAGE to provide a qualitative look at the protein content present in these samples. The sediments obtained following a freeze-thaw cycle from ~ 50 mL of urine varied from as little as 3 mg to as much as 120 mg. Consistent with Figure 3.7, the gel image shown in Figure 3.8 shows extremely variable protein profiles from the urinary sediments of the four donors. Here, equal volumes of samples were taken from each donor, thus the most concentrated protein lanes correspond to samples with a larger portion of sediments. It is apparent that the differences in these gel lanes cannot be accounted for by simple variations in sample

loading. Individual protein bands have significant variability in intensity, indicating that the sediment portion of urine, much like its soluble counterpart, varies according to the donor.

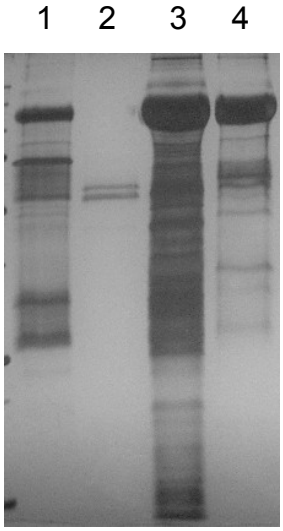


Figure 3.8: Urinary sediment proteins obtained from 4 different donors (denoted 1-4).

Also observed was considerable variation in the protein content of urinary sediment obtained from the same individual donor, collected at different times (Figure 3.9). Although these gel images provide only a qualitative view of the urinary sediments proteome, they firmly establish that the sediment portion of urine is highly variable. This suggests that discarding sediments will bias the urinary proteome prior to analysis.

Figure 3.9: Urinary sediments isolated from the same donor on different days (denoted as 1 and 2).

Because urinary sediments contain many proteins, they inevitably must also contain potential biomarkers. Therefore, sediments should be combined with the supernatant or at least analyzed in parallel with the supernatant. This is particularly important when doing comparative biomarker studies, since one sample may have a larger portion of sediments than another (as clearly demonstrated above), resulting in apparent differences in protein concentration. As clearly shown, these variations may simply be due to variations in the degree of precipitation of proteins induced by a freeze-thaw cycle rather than a pathogenic process. It is realized, however, that the consequence of discarding the sediments may vary depending on the analytical protocol. For example, using the ligand library bead method, as described by Castagna *et al.* [51], less emphasis is placed on protein concentration, and thus discarding the sediments may not significantly alter the resulting protein profile. However, discarding the sediments while

using this method may still impose consequences with respect to *unique* sediment proteins since these would ultimately be discarded.

LC-MS^2 analysis was conducted on urinary proteins recovered from the sediment and supernatant of a urine sample, pooled from four donors. Table 3.3 provides a comprehensive listing of all proteins identified in the sediment portion of the pooled urine sample. A similar MS investigation was conducted using the supernatant portion of urine (sediments discarded), and the complete results are summarized in Figure 3.10. Briefly, 60 individual proteins were identified from urine sediments, of which 44 were common to the supernatant. As a point of reference, 148 proteins were identified from the urinary supernatant. Reversed database searches revealed a false positive rate of 5% for peptides of the sediments and 9% for peptides of the supernatant.

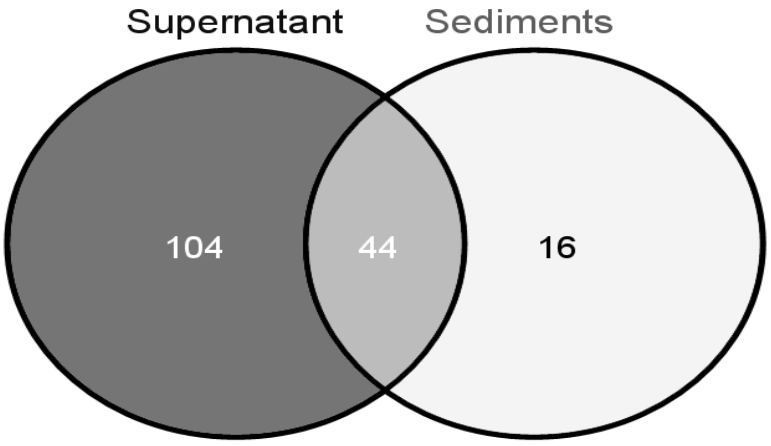


Figure 3.10: A summary of proteins identified in urinary sediments and supernatants *via* LC-ESI-MS^2.

Table 3.3: Urinary sediment proteins identified by LC-MS^2. Proteins unique to the sediments are indicated.

Protein Name	SwissPROT Accession #	Unique to Sediment?
Acylglycerol kinase	Q53H12	no
Alpha-1 acid glycoprotein precursor	P02763	no
Alpha-1-acid glycoprotein	P19652	no
Alpha-1-microglobulin	P02760	no
Alpha-2-HS-glycoprotein	P02765	no
Apolipoprotein D	P05090	no
ATP binding cassette sub familiy	P61221	yes
Brevican core protein	Q96GW7	yes
Cadherin 13	P55290	no
CD44 antigen	P16070	no
CD59 glycoprotein	P13987	no
Crooked neck like protein	Q9BZJ0	no
Double C2 like domain containing protein	Q14184	yes
Dynein intermediate chain 1	Q9UI46	yes
Extracellular sulphatase	Q8IWU5	no
F box like/WD repeat containing protein	Q9BZK7	yes
F-box only protein	Q5XUX0	yes
Granulins	P28799	no
Immunoglobulin gamma 1 chain	P01857	no
Immunoglobulin kappa chain	P01834	no
Immunoglobulin lambda chain	P01842	no
Inter-alpha trypsin inhibitor	Q14624	no
Interleukin-18 binding protein	O95998	yes

Table 3.3 continued

Protein Name	SwissPROT Accession #	Unique to Sediment?
Kallikrein-1	P06870	no
Keratin type 1 cytoskeletal	P13645	no
Keratin type 1 cytoskeletal	P02533	no
Keratin type 1 cytoskeletal	P08779	yes
Keratin type 1 cytoskeletal	P35527	no
Keratin type 1 cytoskeletal 13	P13646	no
Keratin type 2 cytoskeletal	P04264	no
Keratin type 2 cytoskeletal	P35908	no
Keratin type 2 cytoskeletal	P04259	yes
Keratin type 2 cytoskeletal	P13647	no
Kininogen	P01042	no
Leukocyte-associated immunoglobulin	Q6ISS4	yes
Lithostathine 1-alpha	P05451	no
Matrix metalloproteinase	P51512	no
Microtubule associated protein	P11137	yes
Non secretory ribonuclease	P10153	no
Nuclease sensitive element binding protein	P67809	yes

Table 3.3 continued

Protein Name	SwissPROT Accession #	Unique to sediment?
Oteopontin	P10451	no
Peptidoglycan recognition protein	O75594	no
Polymeric immunoglobulin receptor	P01833	no
Proactivator polypeptide	P07602	no
Prostaglandin-H2-D isomerase	P41222	no
Prostate stem cell antigen	O43653	no
Protein phosphatase 1D	O15297	yes
Protein S100-A8	P05109	yes
Protein YIPF3	Q9GZM5	no
Prothrombin	P00734	no
Roundabout homologue	Q8WZ75	no
Serum albumin	P02768	no
TATA box binding protein	P62380	yes
Trypsin-1-precursor	P07477	no
Uromodulin	P07911	no
Vesicular integral membrane protein	Q12907	no
Vitellin membrane outer layer protein	Q7Z5L0	no
WAP four disulphide core domain protein	Q14508	no
WD repeat containing protein	Q8IWG1	yes
Zinc-alpha-2 glycoprotein	P25311	no

The fact that a large percentage of sediment proteins (73%) were also observed in the supernatant phase is not surprising. As shown above, the freeze-thaw cycle causes proteins to partition between the solution and sediment phases. Some proteins undoubtedly precipitate to a larger extent than others. The proteins which were exclusively observed in the sediment portion of the sample do not provide direct evidence that they completely partitioned to this phase, but may be taken as an indication that their relative concentration is perhaps highest in this phase. Again, this illustrates the importance of analyzing both portions of the sample for a more comprehensive investigation of the urinary proteome as well as for reliable biomarker discovery and validation.

Upon closer examination of the unique sediment proteins (Table 3.3), it can be seen that several of these originate from squamous epithelia, leukocytes or aggregate on urinary stones. This is consistent with the type of unique proteins one would expect to find in the sediments. In addition, the larger portion of keratins in the sediment phase is consistent with sloughed tissues. Given the type of proteins found in this sample, sediment proteins should be considered part of the *urinary proteome* and may prove important for the survey of various disorders involving the entire body, or in particular, the urinary tract. Undoubtedly, sediment proteins may ultimately be used as an additional source of candidate protein biomarkers.

Many of the proteins identified in the sediments have been reported in the literature as potential biomarkers for various physiological disorders. For example, alpha-1-acid glycoprotein, alpha-1-microglobulin, prostaglandin-H2-D-isomerase, inter-alpha

trypsin inhibitor, kinninogen and zinc-alpha-2-glycoprotein were all identified in the sediments, and have been reported in the literature as candidate biomarkers for various diseases [91-95]. These diseases range from liver fibrosis, bladder cancer, prostate cancer and Hodgkin's disease to urinary tract infections and more.

It should be noted that the challenge of identifying protein isoforms and assigning proper identification to sequenced proteins is certainly realized. To try to avoid this problem, if the same peptides were assigned to multiple proteins, the protein with the highest molecular weight was chosen for identification. In addition, proteins suspected to be redundant identifications (e.g; P08779 and P35527 in Table 3.3) each had a set of unique peptides identified and thus both proteins were considered to have been identified during the experiment. This approach was taken throughout this entire thesis.

3.3.4 Vortexing a Urine Sample Prior to Analysis Resolubilizes Some Sediment Proteins

In some instances of the proteome sample workup, it may be impractical to analyze the urinary sediments. In such a case, it is critical to minimize any sample bias induced by discarding the sediment portion of urine. Figure 3.11 displays a silver stained SDS-PAGE gel image of proteins from the urine of four different healthy donors (denoted as 1-4) isolated from the sediment (top) and supernatant (bottom) phases when the sample was either vortexed or not vortexed prior to separation of the two phases. As would be expected, the vortexed samples from each donor show a decrease in protein intensity from the sediment portion and a corresponding increase in intensity from the supernatant.

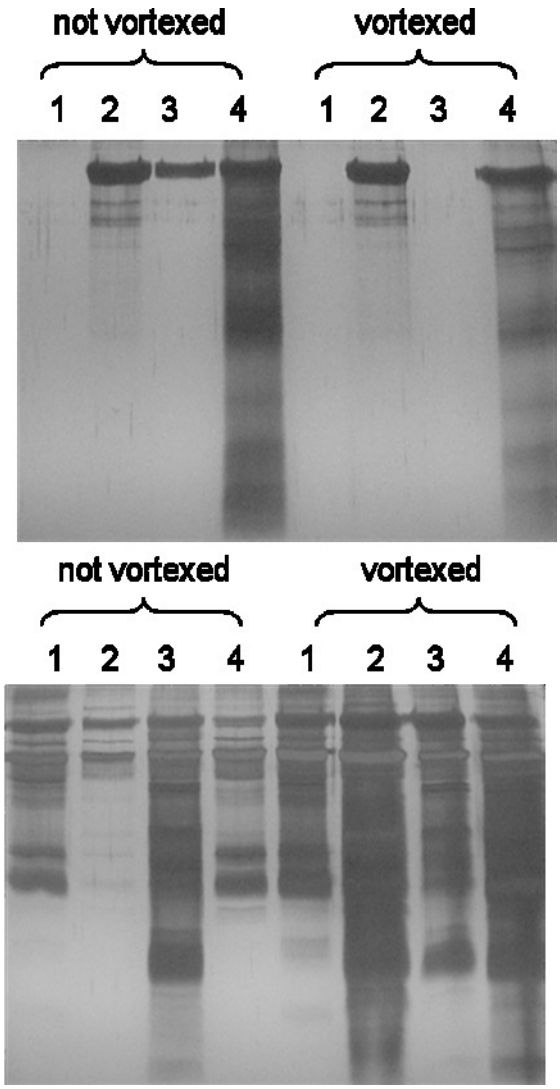


Figure 3.11: Sediment (top) and supernatant (bottom) proteins isolated from the urine of 4 healthy donors. Indicated are the urine protein profiles before vortexing and after vortexing.

Figure 3.11 clearly shows that some of the sediment proteins were solubilized upon vortexing. The extent of sample agitation imposed as urine is initially processed upon thawing is difficult to control. Here, simple agitation of the sample is sufficient to redissolve a portion of the precipitate. It is therefore recommended that as part of all proteome analysis protocols, the urine samples be vigorously vortexed for *at least* two min prior to separating the sediment portion of the sample (if the sediments are to be discarded). This simple step of vortexing affords a greater opportunity of minimizing protein loss due to discarding the sediments.

3.3.5 pH of a Sample Influences the Protein Profile of Urinary Sediments

The relative solubility of urinary proteins is dependent on several variables, such as protein concentration, salt content, temperature, and solution pH [96]. These conditions vary from one physiological state to the next. The pH in particular could have a drastic effect on the partitioning of proteins in urine, since protein solubility is greatly reduced as sample pH approaches the isoelectric point of the protein. It has been stated in the literature that the pH of urine in a healthy individual varies from 4.4 – 8 [97], and this variability increases when considering an unhealthy individual.

Upon examining the partitioning of urinary proteins between supernatant and sediment phases by SDS-PAGE (Figure 3.12), it is clear that changing the pH of the sample will greatly affect the solubility of certain proteins. The control lanes (top) serve as illustration that the sample analysis protocol used here is highly reproducible when considering identical samples of both the soluble and precipitated portions. However, clear differences are observed in the gels as the pH of the urine is adjusted from its

original value of 6.5 (Figure 3.12). The distribution of certain proteins between the sediment and supernatant of urine is significantly altered. Again, since the sediment portion is typically discarded during sample preparation, if such a practice is to be employed it is critical that an *equivalent* portion of the sediments is discarded from all samples being compared. Thus, when sampling only a portion of the urine (*i.e.* the supernatant) it is critical that all variables which may influence the partitioning of proteins to the sediment are controlled, including sample pH and degree of agitation.

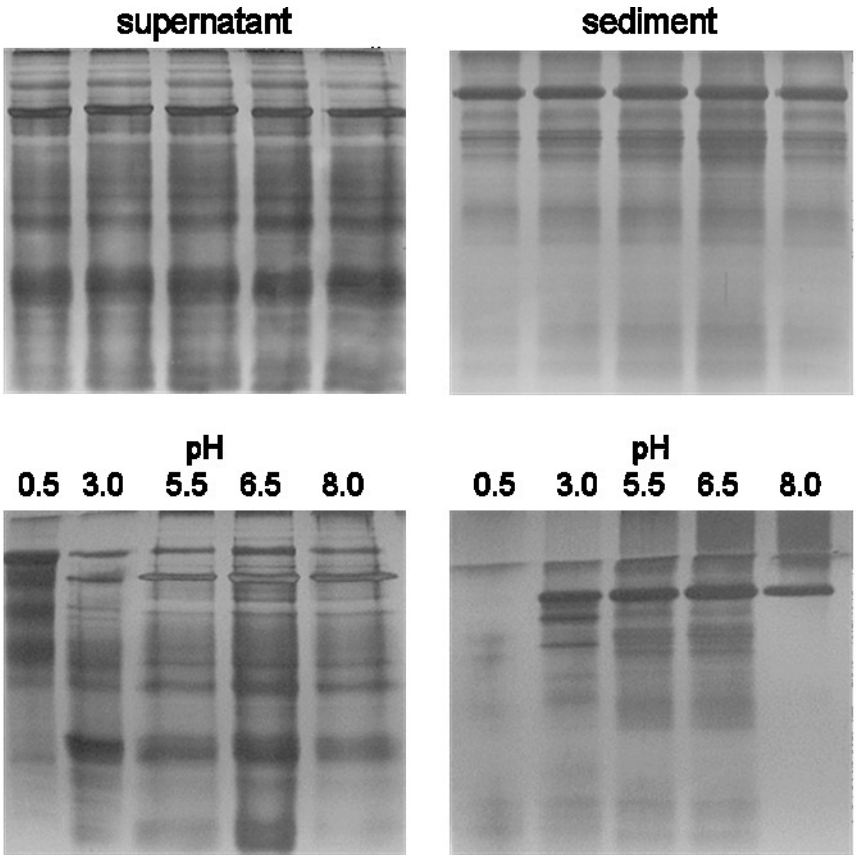


Figure 3.12: Urinary protein profiles of control samples with no pH adjustment (top) and of samples subjected to pH adjustment (bottom).

3.3.6 iTRAQ Quantitation of Sediment Proteins

In addition to the qualitative visualization of the effect of pH, iTRAQ labelling was conducted in order to quantify some of the differences in protein distribution between sediments and supernatants which result from altering the pH of the sample. These results are summarized in Table 3.4. Reversed searches revealed a false positive identification rate of 4%. As expected, differences in protein concentration in the sediments were observed at different pH values. As would be predicted, the proteins detected in the samples at identical pH values (5.9) all had ratios close to 1. This was not the case for the samples whose pH was modified. To provide an example, the acidic sample revealed that the clusterin precursor protein was decreased by a factor of almost three compared to the control, while kallikrein-1-precursor protein was increased by over a factor of six compared to the control.

Table 3.4: iTRAQ ratios of proteins subjected to pH adjustments. The acidic sample had a pH of 3.9, the basic a pH of 8.8 and the control a pH of 5.9.

Protein Name	SwissPROT accession #	Control2/control1	acid/control 1	base/control 1
Alpha-1-microglobulin	O02760	0.92	0.6	0.87
Clusterin Precursor	P10909	1.03	0.27	1.4
Deoxyribonuclease-1 precursor	P24855	0.96	0.58	0.58
Human serum albumin precursor	P02768	0.96	0.88	0.65
Kallekrein-1-precursor	P06870	1.31	6.29	1.16
Pro-epidermal growth factor precursor	P01133	1.03	0.85	1.11
Prostaglandin H-2 D Isomerase precursor	P41222	0.73	0.71	0.68
Uromodulin precursor	P07911	1.34	1.37	1.19

Interestingly, these variations in protein concentration are observed despite the fact that they originate from a single identical urine sample. It is important to note that the pH ranges employed for this experiment are within normal biological variation for a urine sample; thus the pH of urine from an unhealthy donor may have even larger variation. These results are consistent with the gel images in Figure 3.12 and serve to demonstrate that a change in sample pH can have a significant influence on the partitioning of proteins between sediments and the supernatant portion of urine. It is thus essential in a biomarker discovery experiment to ensure that all variations between samples be minimized.

3.4 Conclusions

Human urinary sediments contain a significant number of proteins whose relative distribution between sediment and supernatant is highly variable from sample to sample within an individual and among individuals. The sediment protein profile is influenced by the degree of agitation (vortexing/ non vortexing), and pH. In this chapter, it has been established that the sediment portion of urine is rich in protein, and implications for this finding have been discussed. It is likely that other variables (concentration, salt content, temperature) will also change the relative distribution of proteins between sediments and supernatants. From this work, it is clear that variations in a sample, as well as minor differences in the analytical protocol will affect the detection or quantitation of proteins in urine. In order to achieve an unbiased analysis, urinary sediments should be analyzed along with the supernatant of the sample. As an alternate strategy, where it is desired that the urinary sediments be discarded, it is suggested that the sample pH is normalized and

that urine be thoroughly vortexed prior to centrifugation or filtration, in order to resolubilize the largest portion of proteins possible from the sediment phase. This step will ensure that any sample bias introduced by discarding the sediment portion of urine is, *at best*, minimized.

CHAPTER 4: MASS SPECTROMETRY TOLERANCE TO SODIUM DODECYL SULPHATE [2]

4.1 Introduction

Less than two decades ago, SDS-PAGE was considered *the* leading technology for proteome profiling. With reasonable mass accuracy, this technology allowed for a rapid assessment of protein components from extremely complex biological mixtures. As previously mentioned, soft ionization techniques for mass spectrometry (MS), namely MALDI [5] and ESI [4] gave scientists a new and powerful tool for protein characterization. Through MS, not only did mass determination improve, but more importantly, tandem MS sequencing brought high-throughput proteome characterization to reality. Though these ionization platforms provided a superior technology for protein characterization, both MALDI and ESI possess non-ideal characteristics, which partially limit their use in biological analysis.

As with almost all detection platforms, MALDI and ESI-MS are prone to matrix effects. Signal suppression from ionic compounds is a particular concern with biological samples. Perhaps no source of signal suppressor causes more hesitation in the MS community than ionic detergents, of which sodium dodecyl sulphate (SDS) is the classic and favoured detergent among protein biologists.

---

[2] Reproduced in part with permission from Diane Botelho, Mark J. Wall, Douglas B. Vieira, Shayla Fitsimmons, Fang Liu and Alan A. Doucette. Top-Down and Bottom-Up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J. Proteome Res.*, **2010**, 9, 2863-2810. Copyright 2010 American Chemical Society.

The view that SDS is incompatible with MS was quickly established following adoption of MALDI and ESI for protein analysis. In 1990, Beavis *et al.* presented work which expanded MALDI towards protein mixture analysis [98]. Referring in this report to ionic detergents, it was stated that the detergent "*must be completely removed from a sample before good mass spectra can be obtained*". At approximately the same time, Ikonomou *et al.* [99] presented fundamental investigations on the ESI process, concluding that background ions above 10^{-5} M will induce signal suppression. Inferring from this concentration, SDS levels higher than $\sim 3\times10^{-4}$ % would suppress ESI signals. Evidence for SDS suppression has been extended throughout the literature, discussing the mechanism and extent of suppression for this and other detergents [100]. Though signal enhancement induced by the addition of SDS has occasionally been reported, most are limited to laser ionization approaches [101-103]. In the ESI experiment, however, SDS is inevitably viewed as a signal suppressor. To biologists, such a conclusion presents a problem; while MS is generally intolerable to ionic detergents, SDS is often required for protein solubilization, as well as for establishing separation through SDS-PAGE.

The benefit of SDS for protein solubilization and separation can still be realized if the detergent is removed prior to ionization. While significant effort is devoted to finding alternatives to SDS (organic solvents for solubilization, non-ionic [104] and "MS-compatible" reagents such as cleavable surfactants [105] for example), equivalent effort is devoted to finding effective methods to reduce SDS from the sample. Methods for SDS removal include protein precipitation with organic solvents, column based approaches (ion exchange, HILIC, gel filtration), and dialysis [106-109]. These methods vary in terms of their ability to eliminate SDS, with levels spanning ~2 orders of magnitude

(from 10 to 1000 fold reduction) [106-110]. Other factors weigh into the overall effectiveness of an approach, including protein yield as well as robustness of the protocol (ease of use, reproducibility, cost). As a consequence, many groups adopt their "personal favourite" strategy for SDS removal prior to MS analysis.

MS analysis following SDS reduction has been extensively reported throughout the literature. For example, SDS-assisted protein digestion has been shown to enhance the detection of membrane proteins [111]. As a solubilizing additive, SDS also facilitates comprehensive proteome characterization. Recently reported by Mann *et al*., a high concentration of SDS (4%) was shown to facilitate protein solubilization [112]. The detergent was removed just prior to digestion by using a molecular weight cutoff membrane. While the technique is only amenable to bottom up analysis, the benefits of SDS are clearly illustrated. Given the importance of SDS in proteomics, a quantitative understanding of the influence of SDS on proteome detection would clearly be beneficial. The influence of SDS established in direct infusion experiments cannot translate to an LC-MS experiment. It is also possible that SDS suppression will vary between the analyses of small peptides *vs* larger proteins. The question therefore addressed is to what level the detergent can be tolerated before MS proteome characterization suffers.

In this chapter is a detailed investigation of the effects of SDS on both a bottom up and top down proteome workflow. Results reveal that LC/ESI-MS is surprisingly tolerable to high concentrations of SDS. In addition, methods for reduction of SDS were evaluated in protein solutions, demonstrating that a modification of the well-known acetone [110] and chloroform/ methanol/ water [108] precipitations can reduce SDS to permit MS analysis, while maintaining high protein recovery.

4.2 Materials and Methods

4.2.1 Solutions, Reagents and Protein Standards

All protein standards, including trypsin, as well as HPLC solvents were purchased from Sigma (Oakville, Canada). The reagents used for casting and running acrylamide gels, as well as urea and SDS were obtained from Bio-Rad (Mississauga, Canada).

4.2.2 Protein Precipitation

Samples were precipitated overnight at -20^{o}C following addition of acetone in a 4:1 ratio over the sample. Samples were centrifuged for 15 min at 13,000 rpm and the top layer was decanted, leaving behind ~20 μL in the vial. Additional washing of the pellet was executed using 400 μL aliquots of cold acetone, with immediate centrifugation at 13,000 rpm (no mixing) before removing the acetone layer. To avoid a sampling bias in optimizing the number of washes, the entire batch of samples were precipitated and subsets were randomly selected to continue with added washes. A total of 8 replicates for each wash cycle (0, 1 or 2) was employed.

The protocol used for chloroform/ methanol water precipitation is an adaptation of the method of Wessel and Flugg [108]. Briefly, 400 μL of methanol, followed by 100 μL chloroform, then 300 μL water was added to 100 μL of sample, with brief vortexing after each solvent addition. The sample was centrifuged for 15 min at 13,000 rpm and the top layer was removed. A 400 μL aliquot of methanol was added, with gentle mixing to encourage the solvents to mix. The vial was centrifuged for 15 min at 13,000 rpm and the solution was fully decanted. Optional washing was executed with 400 μL of methanol, with an additional 15 min centrifugation. During the optimization of the precipitation

protocol, a randomized approach was employed to select the vials for the additional washing steps.

4.2.3 SDS Assay

A methylene blue spectrophotometric assay was used to assess the level of SDS remaining in the pellet following protein precipitation [113]. An 8 point standard curve was generated from an SDS stock solution prepared by weight. The results are reported from replicate analysis of 8 yeast samples (100 μL initial volume, containing 10 μg protein) initially diluted in water containing 2% SDS.

4.2.4 LC-TOF MS Analysis of Intact Proteins

Six standard proteins (cytochrome C, myoglobin, ubiquitin, carbonic anhydrase, lysozyme and bovine serum albumin), prepared in solution with varying concentrations of SDS, were subjected to intact protein analysis on a MicroTOF LC mass spectrometer equipped with an electrospray source (Bruker, Toronto, Canada). A self packed column with dimensions 10 cm x 200 μm was prepared in a silica capillary, using 5 μm × 300 Å C4 beads (Michrom, Sacramento, CA). A flow rate of 2 μL/min was used, and 10 μL of sample (corresponding to 1 μg total protein) was injected. The gradient from 0.1% formic acid in water (solvent A) to acetonitrile with 0.1% formic acid (solvent B) was as follows: hold at 5%B for 10 min, increase to 80%B by 30 min and then to 100%B at 35 min. Data were processed with the DataAnalysis version 3.3 software (Bruker, Toronto, Canada). Results reported are from averaging intensities obtained from extracted ion chromatograms from triplicate injections.

4.2.5 Bottom up Proteome Analysis for SDS-Containing Samples

A 0.1 μg/μL yeast sample was digested according to standard protocol. Briefly, samples were reduced with DTT, alkylated with iodoacetamide and digested overnight with trypsin. 5 μL of the digest was diluted with 15 μL of SDS in water to the appropriate concentration. 4 μL of each sample (corresponding to 0.1 μg total protein) was subjected to LC-MS^2 in triplicate. The column was a self-packed 20 cm × 75 μm spray tip (New Objective, Woburn, MA, USA) using 4 μm C18 beads (Phenomenex, Torrance, CA, USA). The flow rate was 0.3 μL/min. Solvent A was water with 0.1% formic acid and solvent B was acetonitrile with 0.1% formic acid. The gradient was initially set to 5% B and held for 10 min. The %B was increased to 35% by 70 min, 100% by 80 min and lowered to 5% at 81 min. MS was on an LTQ linear ion trap instrument (Thermo Fisher Scientific, San Jose, CA, USA) employing a 'triple play' data dependent scan. This method cycles from a full MS scan to a zoom scan to determine charge state, followed by MS^2 of the top three ions. Charge state screening was enabled to ignore singly charged ions, ions with a charge of 4 and greater, or ions where the charge state could not be assigned. Data was searched using the Bioworks browser software package from Thermo, which uses the SEQUEST search engine. Data was filtered as follows: ΔCn less than or equal to 0.1, RSP greater or equal to 4, number of top matches equals 1, peptide probability less than or equal to 5.00×10^{-3} and different peptides. +2 ions required an X_{corr} of at least 2.2 and +3 ions at least 3.75. These criteria establish a peptide false positive rate of 1%, through reverse database searching.

4.3 Results and Discussion

4.3.1 Effects of SDS on Proteomics LC-MS Experiments

With little debate, the ionic surfactant SDS is recognized to cause significant analyte suppression in electrospray MS. As a popular additive for solubilization and separation of proteins, SDS concentrations on the order of 0.1% to 2% are typical. Such a high percentage inevitably means that direct ESI MS of these detergent-containing solutions would be difficult, if not impossible. With a proteome MS experiment, a clearly distinct set of operating conditions is established, which will invariably influence the extent of SDS suppression. A distinguishing feature of a proteome experiment is the coupling of low flow HPLC (sub-μL per min) with MS through nanospray interfaces. One can also consider two types of proteome experiments, namely bottom up and top down analysis. Here, the effects of SDS have been quantified in these two distinct proteome experiments.

4.3.2 Bottom Up SDS Suppression

A tryptic digest of a solubilized yeast extract was chosen as a representative proteome for bottom up MS. Such a sample contains a variety of peptides ranging in size, hydrophobicity, and concentration. As a marker of SDS signal suppression, not only were changing signal intensities quantified, but also assessed was the number of identified peptides obtained through conventional data dependent sequencing. The results from triplicate analysis of SDS-spiked samples are presented in Figure 4.1.

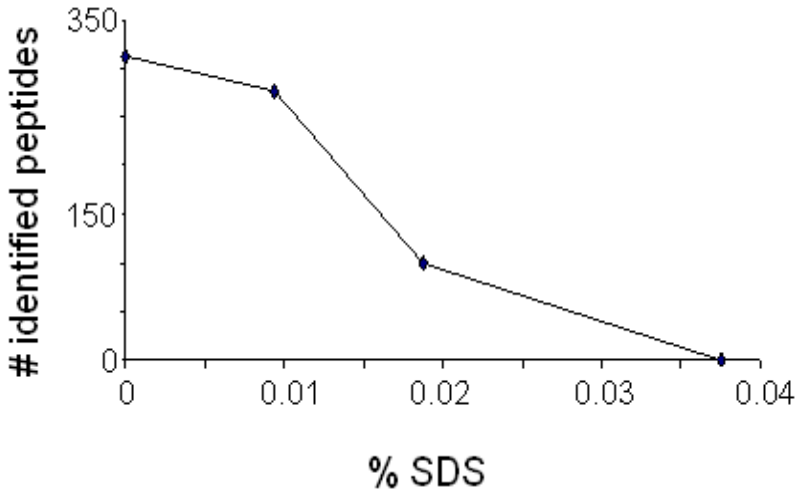


Figure 4.1: The number of yeast peptides identified by MS^2 as a function of the % SDS in the peptide solution.

What came as a surprise is the relatively high tolerance of SDS observed in a bottom up experiment. In fact, initial experiments involved spiking SDS at concentrations ranging from 10^{-6} to 10^{-3}%. Replicate analysis of these low level SDS-spiked samples revealed no change in the number of identified peptides, nor in the appearance of the total ion chromatograms (data not shown). It was ultimately found that 0.01% SDS caused a slight drop in signal intensity, as observed in the total ion chromatogram, and a corresponding drop in the number of identified peptides (Figure 4.1). At 0.02% SDS, signal suppression was evident, both in the TIC (Figure 4.2), as well as by the observed drop in identified peptides.

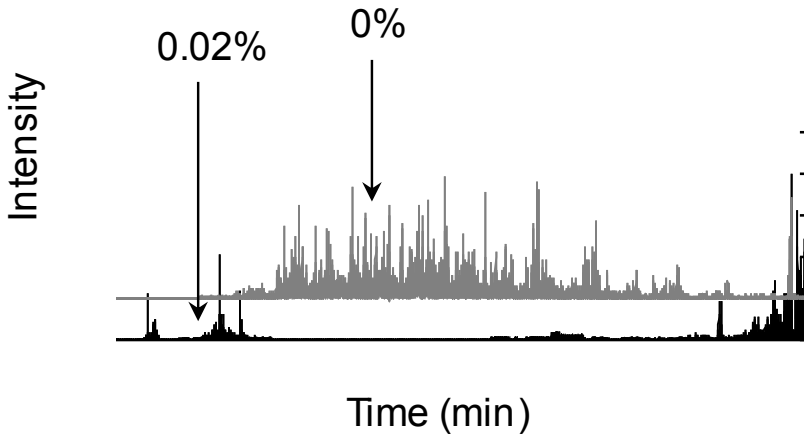


Figure 4.2: Total ion chromatograms of a yeast digest in 0% SDS and 0.02% SDS.

The effects of SDS on chromatographic separation can be seen in Figure 4.3. This image provides selected ion chromatograms for two representative peptides (MH^+ 1416, MH^+ 1288), of known identity. Consistent with the number of peptide identifications, 0.01% SDS in the initial sample had no significant effect on chromatographic peak shape or retention time of the peptide. However, at double this level of SDS, a dramatic change in the chromatogram was revealed. Not only have peaks broadened significantly, which would contribute to the lowering of MS signal intensity, but also the retention times were dramatically altered.

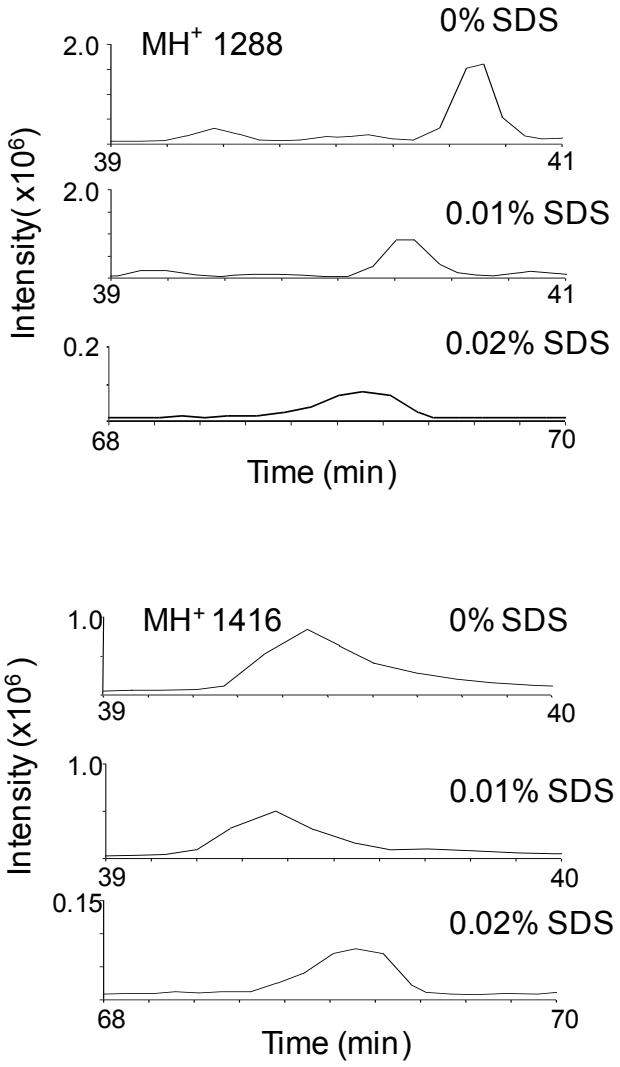


Figure 4.3: Selected ion chromatograms for two peptides (MH^+ 1288 and MH^+ 1416) at varying percentages of SDS.

It is important to note that the total peptide concentration (0.1 μg per injection), as well as the injection volume employed (4 μL) were not altered through these experiments. Such values, along with the column dimensions, flow rates, and MS operating conditions were previously determined as optimal running conditions in the laboratory. It cannot be discounted that changing the peptide concentration or the total injection volume, *i.e.* changing the "protein to SDS" ratio, will have an effect on the level

of signal suppression. Nonetheless, the determination of a high threshold value for SDS serves as an important guideline for proteome experiments in the presence of SDS, and illustrates the dramatic deterioration of proteome signals beyond this concentration.

4.3.3 Top Down SDS Suppression

As previously mentioned, top down MS has been popularized with increasing availability of high-resolution MS, and commercial tools for fragment interpretation. Six standard proteins (MW ~8 to ~70 kDa) were chosen to represent a simple intact proteome mixture. Signal intensities extracted from the top 3 multiply charged ions per protein were used to construct the plot shown in Figure 4.4 (averaged from triplicate injections).

What is evident from this figure is the similar pattern which emerges with each of the 6 proteins. At SDS concentrations below 0.01%, only slight suppression is observed (if any). In fact, a slight increase in signal intensity is often seen at these low percentages of SDS. At higher levels of SDS (0.025%) a dramatic decrease, or in some cases, complete loss in signal intensity is observed. Figure 4.5 illustrates the charge envelope and deconvoluted spectra observed from LC/ESI analysis of carbonic anhydrase at an initial SDS concentration of 0% (top) and 0.025 % (bottom). At 0.025% SDS the deconvoluted mass for carbonic anhydrase (29,027 Da) is still observed, although the mass appears to be distorted and the signal intensity is nearly two orders of magnitude lower than for a “clean” sample. It is clear that at this concentration of SDS, reliable MS detection of intact proteins is no longer possible.

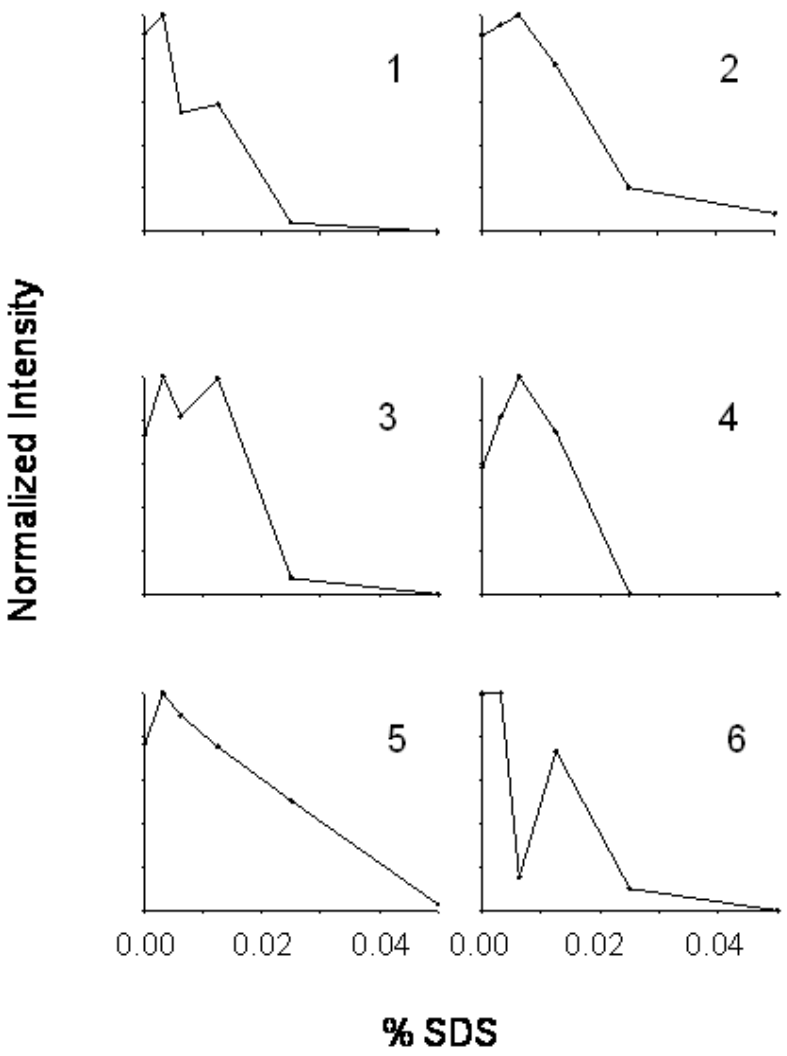


Figure 4.4: Normalized intensity versus % SDS for 6 intact protein standards after LC-TOF MS analysis. The proteins are as follows; 1: Myoglobin, 2: Cytochrome C, 3: Carbonic Anhydrase, 4: Bovine Serum Albumin, 5: Ubiquitin, 6: Lysozyme.

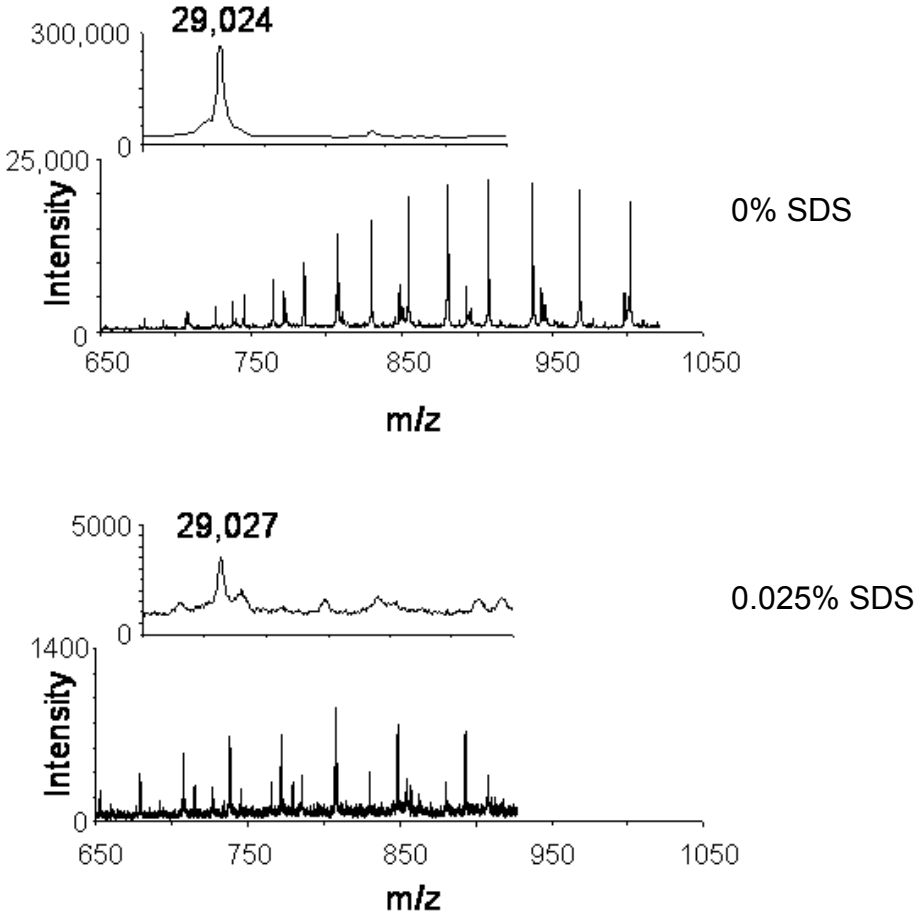


Figure 4.5: Charge envelope and deconvoluted spectra observed from LC-ESI analysis of carbonic anhydrase at 0% SDS (top) and 0.025 % SDS (bottom).

It should be noted that the LC-MS operating conditions used for intact protein analysis were entirely different than those for the bottom up experiment. In addition to a difference in the concentration and volume of sample injected *vs* the bottom up experiments, intact protein analysis was conducted on a time-of-flight (TOF) system, using a conventional ESI source (*vs* nanospray on a linear ion trap). The column dimensions, composition and flow rates were also varied. Interestingly, despite these

many differences, the threshold level for top down proteome experiments remains at 0.01% SDS.

4.3.4 Removal of SDS from Protein Containing Samples

With knowledge of a consistent threshold tolerance of SDS in a proteome MS experiment, next addressed were methods to reduce the SDS content from biological samples. A suitable method for SDS reduction is one which reliably reduces the SDS content to below 0.01%, while maintaining high protein yields. Puchades [110] has quantified methods for reduction of SDS, including organic solvent precipitation. Evaluated here is the acetone and chloroform-methanol-water (CMW) precipitation protocols, choosing an initial SDS concentration of 2%. Figure 4.6 shows the percentage of SDS that remains following precipitation of a 10 μg yeast sample contained in 100 μL. These results represent replicate measures of 8 samples, the small error bars illustrate the high level of reproducibility that is achieved with these precipitation methods.

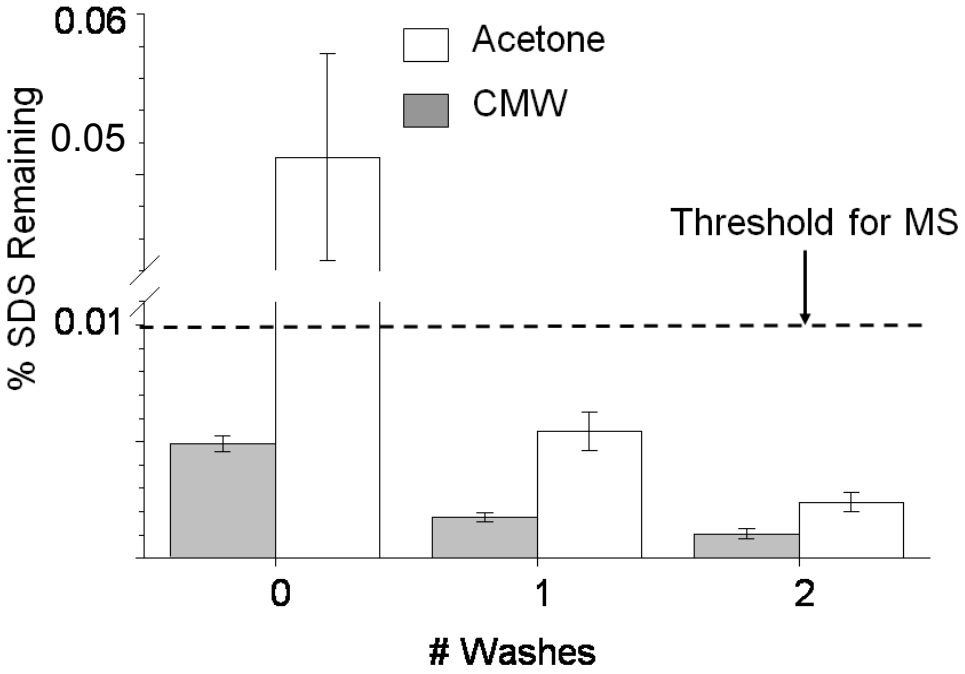


Figure 4.6: The percentage of SDS that remains following chloroform-methanol-water (CMW) and acetone precipitation of a yeast sample. As a point of reference , the previously determined MS tolerance of 0.01% is indicated

It has been shown previously [110] that the CMW protocol clearly outperforms conventional acetone precipitation in terms of reducing SDS concentration. Thus, beginning with a 2% solution of SDS, the conventional acetone protocol (*i.e.* 0 washes in Figure 4.6) will not reduce SDS to tolerable levels to permit proteome analysis.

One of the major variables influencing protein yield during a precipitation protocol is the extent to which the supernatant is pipetted from the protein pellet. A common temptation is to remove as much solvent as possible to maximize the reduction of SDS, but in doing so, one also increases the likelihood of aspirating a portion (or all)

of the protein pellet. Moreover, at protein levels below 5 μg, the risk of protein loss increases as the pellet becomes extremely difficult to visualize. Thus, also explored was a less 'risky' strategy for protein precipitation. In this strategy, additional washing steps were incorporated following the initial precipitation. In doing so, a larger volume of supernatant was allowed to remain in the vial, thus minimizing the risk of accidental aspiration of the pellet. These added washing steps do not reduce protein yields, as minimal resolubilization is expected. As shown in Figure 4.6, the benefits of the added washes are clear in terms of further reduction of SDS levels. In relation to the MS tolerance threshold of 0.01% SDS, a modified acetone protocol incorporating 1 or 2 washes becomes a viable alternative to the CMW protocol. To be confident of sufficient removal, it is recommended that 2 washes be employed for acetone, and 1 wash for CMW precipitation.

4.4 Conclusions

This chapter has shown that MS tolerance to SDS is surprisingly high in both a top down and bottom up proteome workflow. While this concentration is still considerably lower than what is typically used in protein sample preparation, it does lend the use of SDS reduction protocols to enable LC-MS analysis. More specifically, following GELFrEE separation, acetone precipitation employing one wash or chloroform precipitation employing two washes is sufficient for removal of SDS prior to MS analysis. These methods are favoured over column/ cartridge protocols for SDS removal in that they are easily scaled to a large number of fractions and are also amenable to top down and bottom up MS analysis. With this in mind, the common view that SDS should be avoided in a solution-based proteome workflow should be reconsidered.

CHAPTER 5: A PROTEOMIC COMPARISON OF GELFrEE TO GeLC [3]

5.1 Introduction

In a process analogous to SDS PAGE, Tran and Doucette's [114] reported solution-based electrophoretic platform, the GELFrEE system incorporates SDS to impart molecular weight proteome fractionation, and has been applied to both bottom up and top down MS [115,116]. Undoubtedly, the most widely employed proteome workflow incorporating SDS is the coupling of SDS-PAGE to LC-MS. Established over a decade ago, Shevchenko's robust in-gel digestion strategy [84] has extended to the popular GeLC approach [117], involving digestion and analysis of the full set of gel bands recovered from SDS PAGE. The GeLC protocol provides nearly unmatched proteome coverage, and is essentially unbiased towards protein class, concentration, and size [118,119].

While native-type electrophoresis experiments are possible, SDS is used to impart separation throughout this thesis. Given the similarities of the GELFrEE and GeLC platforms, a comparison of these proteome workflows was warranted.

---

[3] Reproduced in part with permission from Diane Botelho, Mark J. Wall, Douglas B. Vieira, Shayla Fitsimmons, Fang Liu and Alan A. Doucette. Top-Down and Bottom-Up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J. Proteome Res.*, **2010**, 9, 2863-2810. Copyright 2010 American Chemical Society.

5.2 Materials and Methods

5.2.1 Yeast Proteome Extraction

S. cerevisiae was grown in yeast extract peptone dextrose (YEPD) media at 25ºC and harvested at an optical density of 0.5 by centrifugation at 3500 x g for 5 min. Cells were washed in phosphate buffered saline (PBS) buffer and collected by centrifugation at 3200 × g for 15 min at 4 ºC. Yeast cells were suspended in 3.6 mL of 25 mM Tris-HCl buffer (pH 7.65) to which 250 µL of protease inhibitor cocktail (Sigma catalogue number P2714, Oakville, Canada) was added. The suspension was passed three times through a French Press at 20,000 psi, and the protein concentration, as determined by a bicinchoninic acid (BCA) protein assay was 10.5 mg/mL. Samples were aliquoted and stored at -20^{o}C until immediately prior to use.

5.2.2 Acetone and SPE Preparation of Urine Proteins

50 mL of urine (never frozen) was collected with 5 mL of protease inhibitor. 1.5 mL was precipitated at -20 ºC overnight with 6 mL of cold acetone. The sample was centrifuged for 10 min; the supernatant was decanted and the pellet was dried and prepared for GELFrEE. For SPE, 10 mL of the same urine was processed an SPE column. Proteins were eluted in 0.5 mL of 30% acetonitrile/0.1% TFA in water, followed by 0.5 mL of 50% acetonitrile/0.1% TFA in water. The eluents were combined and assayed using a Bio-Rad protein assay. The resulting concentration of eluted proteins was 0.88 µg/µL.

5.2.3 GELFrEE Separation

Further details on operation of the GELFrEE system can be obtained from Tran *et al*. [115]. Yeast proteins were suspended in gel loading buffer at a concentration of 2 mg/mL, and heated to 95^{o}C for 5 min. 100 µg of protein was loaded onto a 1 cm gel column, cast to 12% T (with 3 cm resolving gel). The initial voltage was set to 250 V for the first 10 min and then increased to 300 V for the remainder of the separation. Following elution of the dye front (fraction 'zero') additional collection times were as follows: 5 × one min fractions, 5 × two min fractions, 3 × five min fractions and 3 × 15 min fractions, for a total of 17 fractions (including fraction 'zero'). Higher molecular weight fractions were separated by GELFrEE by loading an equivalent 100 µg yeast sample onto a 5% T resolving gel column. Fractions were obtained as follows: 1 × 5 min fraction, 1 × 10 min fraction, 3 × 15 minute fractions. The first fraction was discarded.

5.2.4 GELFrEE of Urine Proteins and Visualization *via* 1D-SDS-PAGE

1.5 mL of human urine (with protease inhibitor) was acetone precipitated overnight at -20 ºC with 6 mL of cold acetone. The acetone pellet was centrifuged at 13,000 rpm and the supernatant was decanted. The pellet was allowed to air dry. The pellet was dissolved in 96.5 µL of gel loading buffer (0.5 M Tris-HCl, pH 6.8, 10% *w/v* SDS, 0.5% *w/v* bromophenol blue, 26% glycerol), heated at 95 ºC for 5 min and 85.5 µL (corresponding to a theoretical loading of 100 µg) was loaded onto a 12% T GELFrEE tube gel. GELFrEE separation was conducted exactly as described in section 5.2.3. For 1D-SDS-PAGE of fractions, 10 µL of each fraction was mixed with 40 µL of gel loading

buffer and loaded into a 12% T slab gel and SDS-PAGE was performed. The gels were silver stained for visualization.

5.2.5 GELFrEE of Urine Proteins for LC-MS^2

Acetone precipitated proteins were dissolved in 75 µL of 1 × gel loading buffer. 67 µL (corresponding to a theoretical loading of 100 µg, assuming 75% recovery from precipitation) was loaded onto a 12% tube gel column for GELFrEE separation. For the SPE extracted proteins, 113 µL of sample was added to 30 µL of 5 × gel loading buffer and 143 µL (corresponding to 100 µg of protein based on protein assay) was loaded onto a 12% tube gel column. The initial GELFrEE voltage was held at 240 V for 10 min and was then increased to 300 V for the remainder of the run. 17 fractions were individually collected per gel column. Following fractionation, all samples (100 µL volumes) were precipitated with 400 µL acetone, overnight at -20 ºC. After precipitation 16 of the 17 fractions were analyzed by LC-MS^2.

5.2.6 Comparison of Acetone and SPE Extracted Proteins Following GELFrEE

Two 1.5 mL urine samples, previously frozen at -20 ^{o}C, were thawed at room temperature and vortexed for 5 min. One of the thawed samples was precipitated overnight at -20 ºC through addition of 6 mL of cold acetone. The remaining sample was centrifuged at 13,000 rpm for 5 min to pellet any remaining sediments and the supernatant was decanted from the sediments. The entire supernatant was loaded on a Bond Elute SPE column (Varian, Mississauga, Canada), washed with 20% acetonitrile/water + 0.1% formic acid, and eluted with 500 µL of 30% acetonitrile,

followed by 500 μL of 50% acetonitrile, each in 0.1% formic acid/water. These extracts were pooled. The SPE eluent was concentrated by completely evaporating the solvent in a Speedvac. This fraction (designated 'supernatant'), along with the sediment fraction of the sample (sediment), and the acetone-precipitated urine sample, were subjected to GELFrEE separation, as described in section 5.2.2. Fractions were individually precipitated at -20ºC overnight with 400 μL of cold acetone. Next, 13 of the fractions were chosen for visualization (fractions 0-9, 12, 14 and 16) and dissolved in 20 μL of gel loading buffer. 15 μL from each fraction was loaded into an SDS-PAGE slab gel. The gels were run at 120 V until the run was complete. Gels were silver stained for visualization.

5.2.7 1D-SDS-PAGE of GELFrEE *vs* GELC Comparison

For visualization of GELFrEE fractions, 40 μL aliquots were mixed with 5 × gel buffer [37], heated at 95 ˚C for 5 min and briefly centrifuged at 13,000 rpm. 20 μl (*i.e.* ~1/6th of a given GELFrEE fraction) were loaded into a 1 mm, 12% T polyacrylamide gel and run at 120 V. Protein bands were silver stained for visualization.

For GeLC, 40 μg of yeast proteins, from the same vial as used for GELFrEE separation, were separated on a 12% T gel. An equivalent amount of this sample was loaded in another lane, permitting staining by coomassie. For bottom up analysis, GeLC slices were not stained, though the entire gel was subjected to 3 × 30 min washes in water to reduce the level of SDS in the gel. 20 identical-sized bands were manually excised,

placed in separate vials, and subjected to in-gel digestion using the method of Shevchenko *et al*. [120].

5.2.8 Filter-Aided Sample Preparation of GELFrEE Fractions

The spin column procedure employed was adapted from Mann *et al.* [112]. Briefly, 100 µL of each fraction obtained from GELFrEE was centrifuged with 200 µL of 8M urea in 0.1 M Tris/HCl (UA), pH 8.5 and 1.5 µL of 1 M DTT for 40 min at 13,000 rpm. Following centrifugation, an additional 200 µL of UA was added and the samples were centrifuged for an additional 40 min at 13,000 rpm. At this point the flow-through was discarded and 100 µL of 1 M iodoacetamide solution was added and the samples were allowed to sit at room temperature for 5 min before they were centrifuged at 13,000 rpm for 30 min. Following centrifugation 100 µL of 8 M urea in 0.1 M Tris/HCl pH 8.0 was added to each sample and they were centrifuged at 13,000 rpm for 40 min. This step was repeated twice and the flow-through was discarded. Trypsin was subsequently added to each sample in a ratio of 1:100 by mass of trypsin to protein and allowed to incubate overnight at 37ºC. Following digestion, the samples were centrifuged for 40 min at 13,000 rpm. 50 µL of 0.5 M NaCl was added to each sample followed by centrifugation at 13,000 rpm for 20 min. The filtrate was then acidified with 10 µL of 10% TFA and desalted using liquid chromatography.

5.2.9 Solution Digestion of GELFrEE Fractions

Following acetone precipitation, samples were redissolved in 20 µL of 8 M urea, with dispensing of the solution through a pipette tip 40 times to aid in mixing. The

solution was diluted with 80 µL of 50 mM ammonium bicarbonate. Proteins were reduced with 0.5 µL of 1 M DTT (56 ºC for 30 min) then alkylated with 1.5 µL of 1 M iodoacetamide (room temperature, dark for 30 min) and digested following addition of 100 ng of trypsin (overnight at 37 ºC).

5.2.10 Sample Cleanup, Mass Spectrometry and Database Searching

Following separation and digestion, the GELFrEE and GeLC fractions were subjected to automated sample cleanup using reversed-phase chromatography (0.5 × 5 mm, 5 µm C4 beads) on an Agilent 1200 HPLC system with UV detection at 214 nm. A 20 min hold following injection allowed contaminants to wash through which was followed by a step gradient of 5% to 60% ACN over 5 min. Eluted peptides were collected and dried in a Speedvac, then stored at -20ºC until just prior to LC-MS^2 analysis. Peptide LC-MS^2 analysis was performed using the conditions described above, except that the following gradient was used: initial solvent condition was 5% ACN in 0.1% formic acid, water, which was followed by a gradient that involved instantaneous increase to 8% B following loading, and subsequent increase to 28% ACN over 80 min. Proportional with the amount of material separated in GeLC relative to GELFrEE, two and a half times more sample was loaded from the GeLC fractions compared to those from GELFrEE (corresponding to 3.3% of a GELFrEE fraction per injection). MS was on an LTQ linear ion trap instrument (Thermo Fisher, San Jose, CA, USA) employing a 'triple play' data dependent scan. This method cycles from a full MS scan to a 'zoom scan' to determine charge state, followed by MS^2 of the top three ions. Charge state screening was enabled to ignore singly charged ions, ions with a charge of 4 and greater, or ions where the charge state could not be assigned. MS^2 spectra were searched against

the *Saccharomyces cerevisiae* yeast proteome database (downloaded November 1, 2007), or the Human Uniprot database using the SEQUEST algorithm in the Bioworks v3.2 software package (Thermo Fisher). Data were filtered as follows: ΔCn greater than or equal to 0.1, RSP greater or equal to 4, number of top matches equals 1, peptide probability less than or equal to 1.00×10^{-2} and different peptides. +1 ions required an X_{corr} of at least 1.9, + 2 peptides at least 2.2 and +3 ions at least 3.75. For yeast, decoy searches were performed against a reversed yeast database to maintain a peptide false discovery rate of less than 1%. A false discovery rate of 1.6% for peptides from acetone precipitated proteins, and of 2.7% for peptides from SPE extracted proteins, was found.

5.3 Results and Discussion

5.3.1 Assessment of Proteins Identified *via* GELFrEE and GeLC

Figure 5.1 illustrates the performance of the GELFrEE device for separation of 100 μg of protein extracted from a yeast proteome. A 15% T resolving column was used in this experiment. This image shows the collection of 17 fractions over the approximate mass range of 10 to 250 kDa. The entire separation takes 90 min, following initial collection of fraction zero (elution of dye front). To illustrate the capability of GELFrEE to capture even higher mass proteins, an additional separation was performed. As shown in Fig 5.2, the use of a 5% T column easily permits entrapment of larger proteins, far exceeding 250 kDa, during the same 90 min separation interval.

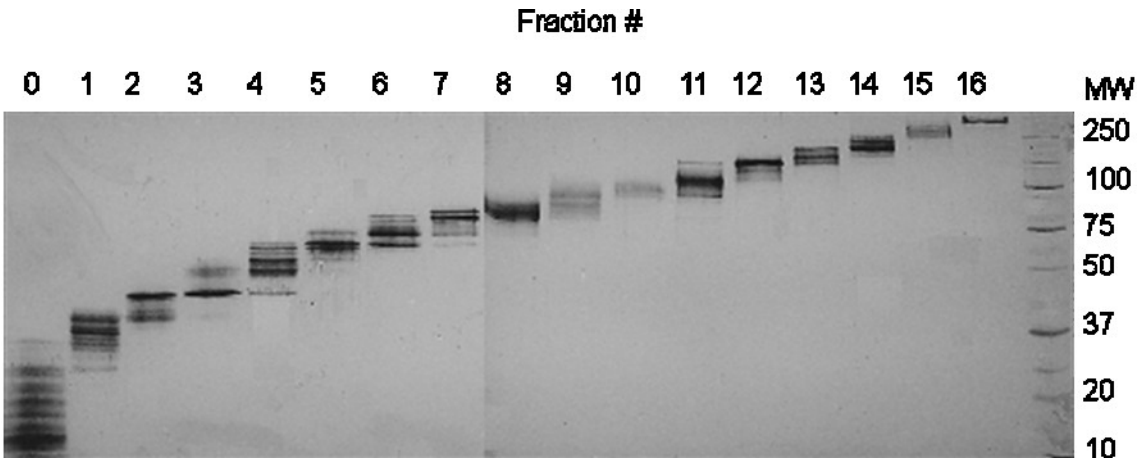


Figure 5.1: Fractions collected from a yeast proteome separated using GELFrEE.

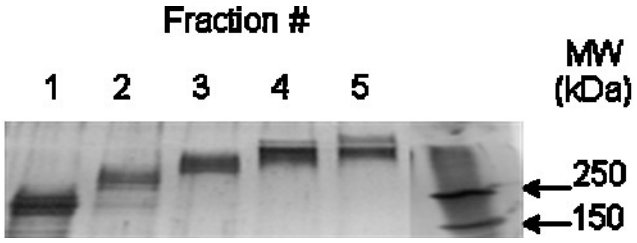


Figure 5.2: Yeast proteins fractionated using GELFrEE on a 5% T tube gel, allowing for higher masses to be collected.

The use of GELFrEE in a top down proteome workflow has recently been reported [116]. From Figures 5.1 and 5.2, it is clear that intact protein separation and

recovery can be achieved over an extremely broad mass range with the GELFrEE experiment.

Figure 5.3 summarizes the number of proteins identified in a bottom up workflow from triplicate LC-MS^2 analysis of each of the GeLC and GELFrEE fractions, herein employing two SDS removal protocols, acetone and chloroform-methanol-water (CMW) precipitation. The corresponding number of peptides is depicted in Figure 5.4.

Several important points can be made from Figures 5.3 and 5.4. First, comparison of the precipitation methods reveals a slightly larger number of proteins (and peptides) identified from acetone over the CMW precipitation (1320 *vs* 1213 proteins). A large number of overlapping proteins are observed (>70% in common). Each method provides distinct advantages: acetone precipitation is technically simpler to perform, while the CMW protocol is much quicker to complete (<1 hour total for CMW *vs* overnight precipitation needed with acetone).

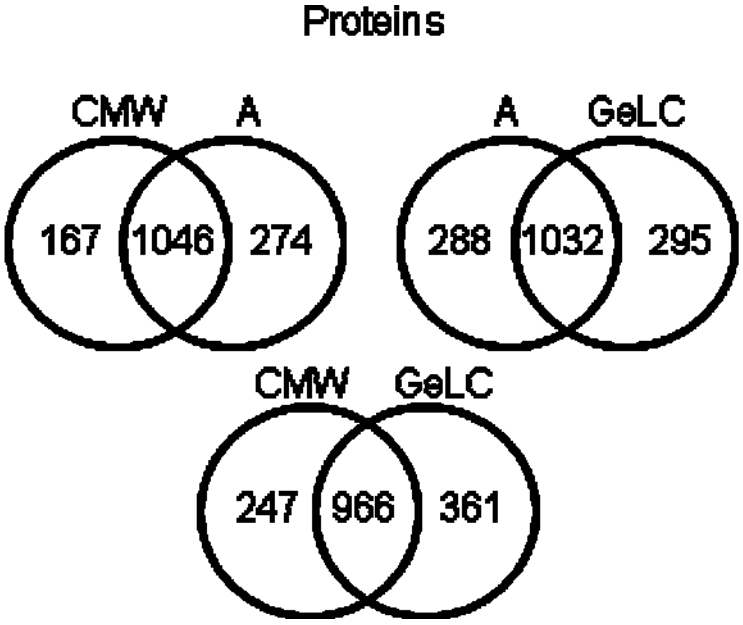


Figure 5.3: Venn diagrams comparing the number of proteins identified by GeLC and GELFrEE. Proteins fractionated by GELFrEE and precipitated by chloroform-methanol-water and acetone are indicated as (CMW) and (A) respectively.

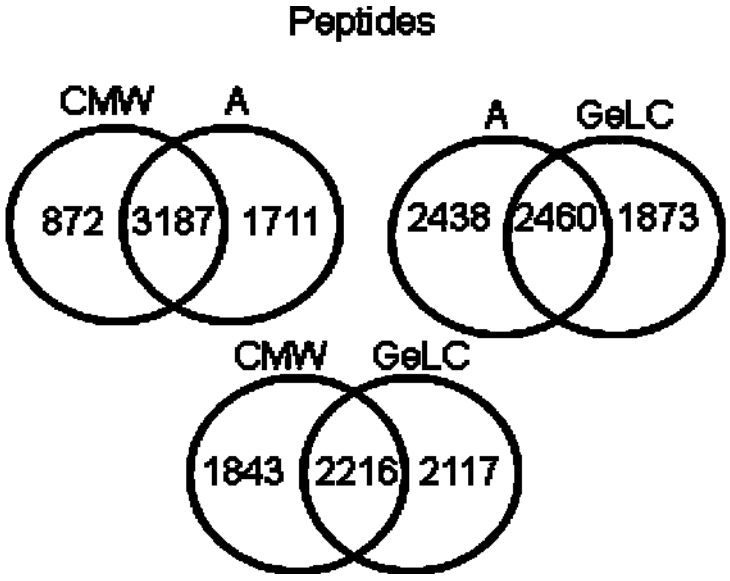


Figure 5.4: Venn diagram comparing the number of peptides identified by GeLC and GELFrEE. Peptides identified by chloroform-methanol-water (CMW) and acetone (A) precipitation following GELFrEE fractionation are indicated.

The most informative comparison is made between the GELFrEE and GeLC workflows. Remarkable agreement is observed between these independent workflows. The total number of identified proteins was virtually identical (1320 for GELFrEE/acetone *vs* 1327 for GeLC). It is worth noting that a total of 20 fractions were collected for the GeLC platform while only 17 were collected from the GELFrEE workflow. Thus, a slightly greater proportion of MS time was devoted to analysis of the GeLC slices. Also, while the GELFrEE system has a capacity for higher sample loading, a more valid comparison of the separation platforms was made by eliminating this variable. Thus, much of the collected GELFrEE fractions were sacrificed, and a smaller percentage of these fractions was injected for analysis compared to the GeLC fractions.

When comparing the identified proteins, approximately 80% were commonly detected between the GELFrEE/acetone and GeLC protocols. A closer examination of the masses of unique proteins identified *via* GeLC and GELFrEE is presented in Figure 5.5. The mass distribution profile for the unique proteins is indeed very similar between the two platforms, with a slight preference towards low-mass proteins in GELFrEE and high mass proteins in GeLC. Considering that the GELFrEE experiment can be optimized to collect fractions over a user-defined time interval, it can be concluded that there is no significant mass bias between GeLC and GELFrEE in terms of protein molecular weight.

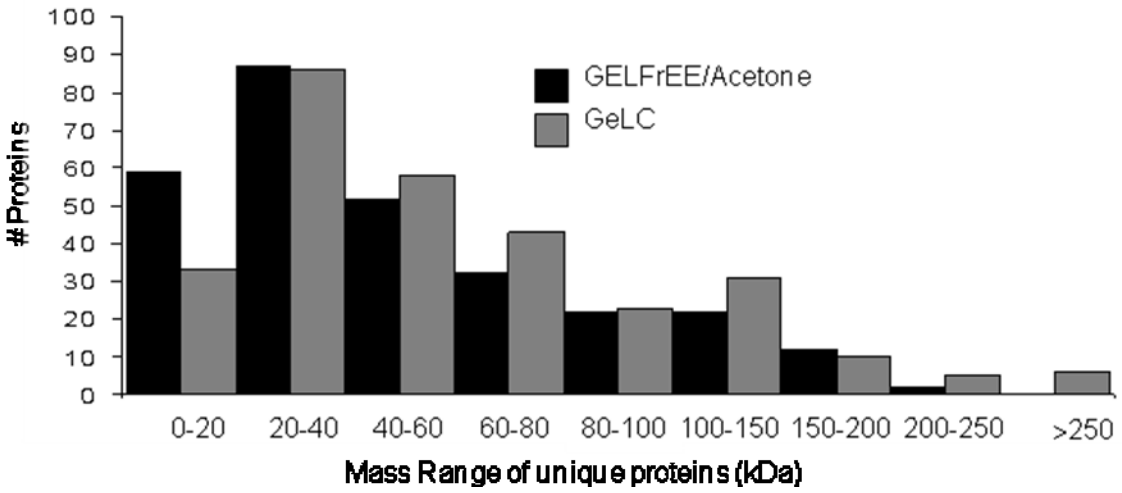


Figure 5.5: A mass distribution plot showing the unique proteins identified from each of the GeLC and GELFrEE platforms (common proteins not included).

Interestingly, although GeLC delivered a slightly higher number of identified proteins, the GELFrEE/acetone workflow yielded 565 more peptides than that of GeLC (Figure 5.4). That is, GELFrEE proteins were identified with slightly more peptides per protein than GeLC proteins. This is presumably due to differences in protein digestion (in-gel *vs* in solution). With the GeLC protocol, the necessity to extract peptides from the gel can be a source of sample loss, ultimately preventing MS detection of larger or more hydrophobic peptides, for example. With a solution digestion, essentially all peptides from a given protein have the potential to be detected.

To further examine the differences observed at the peptide level, the number of missed cleavages was assessed for each technique. It was found that the GeLC approach yielded a higher level of missed cleavages (19% of peptides containing one missed cleavage, 2% having two, and 0.09% with three or more). This compares to 11% of peptides in GELFrEE with one missed cleavage, 0.8% with two, and 0.09% with three or

more. Thus, the complementary nature of identified peptides, combined with the technical simplicity, increased sequence coverage and reduced missed cleavages of the GELFrEE workflow establishes this novel technique as a viable alternative, or complementary strategy to the bottom up, GeLC proteome experiment.

5.3.2 GELFrEE Versus GeLC Protein Resolution

Figure 5.6 shows the mass distribution of all the identified proteins in each fraction for both the GeLC and GELFrEE/acetone workflows. With each platform, the majority of proteins were found in their expected fraction. It is clear from the figure that the GeLC technique provides greater resolution, which would be expected as protein elution from a GELFrEE column into solution will partially compromise protein resolution compared to the proteins remaining 'locked' into a gel as they are in a GeLC approach.

The degree of protein overlap between adjacent fractions can be assessed more accurately in Figure 5.7. It is seen that 63% of the identified proteins were observed in a single GeLC fraction, as opposed to 44% observed for GELFrEE. This gap is slightly reduced when considering proteins found in either 1 or 2 fractions, wherein 81% were found in GeLC compared to 66% in GELFrEE.

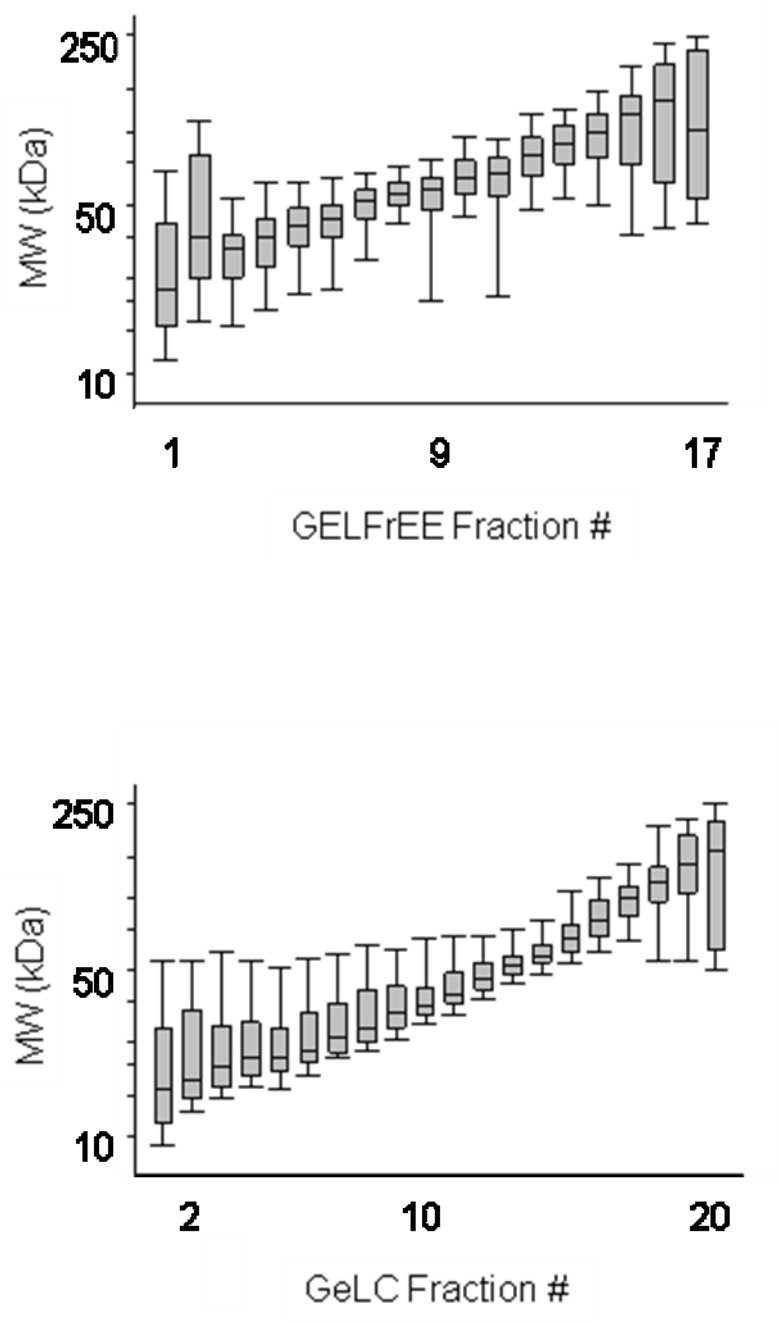


Figure 5.6: Mass distribution profiles of proteins identified by the GeLC platform (top) and the GELFrEE platform (bottom). The middle line in each box represents the median. The top and bottom whiskers represent the highest and lowest values in the set respectively. The top and bottom of the boxes represent the highest and lowest quartile values respectively.

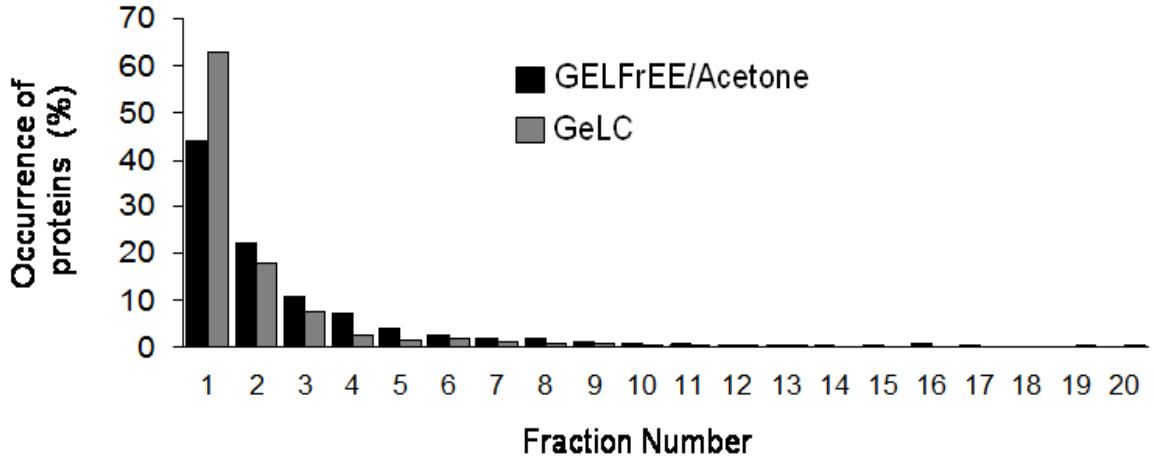


Figure 5.7: The percentage of proteins identified across multiple fractions for GELFrEE/acetone and GeLC separations.

Interestingly, proteins were observed to span multiple fractions in both platforms. For example, 5 proteins were seen across all 17 fractions employing the GELFrEE separation while the GeLC technique yielded 4 proteins observed in all 20 fractions (only two of these being common to both platforms). In all cases the proteins spanning multiple fractions present highly abundant species which are also likely being observed as partially degraded proteins in the multiple fractions.

A final observation relates to the relatively high spread of protein molecular weight observed in the upper mass GELFrEE fractions and GeLC slices. This observation can be explained by the presence of small proteins which are clearly detected in the high mass fractions. For example, glyceraldehyde-3-phosphate dehydrogenase (MW 36 kDa)

was observed in fraction 16 (*i.e.* the last fraction collected) for GELFrEE, while ubiquitin (MW 8.5 kDa) was observed in fraction 20 (the uppermost gel slice) for GeLC. Clearly, these proteins are not in the fractions predicted based on their masses. Despite denaturing the sample in heated SDS buffer, it is most probable that these smaller proteins remained associated with larger proteins, and are likely strongly associative binding partners. This observation serves as a caution in trying to incorporate the intrinsic information gained from separation into the data analysis platform. On the other hand, such an observation could become a useful screen for the detection of interacting protein partners.

Also assessed for ability to digest and clean proteins following GELFrEE fractionation was the filter aided sample preparation (FASP) protocol described by Mann *et al*. [112]. These FASP-cleaned fractions were analyzed in an identical fashion by nano LC-MS^2 to asses the number of proteins identified by this approach. However, relative to the organic solvent precipitation protocols, a considerably smaller number of proteins and peptides were identified. Table 5.1 summarizes the peptides and proteins detected *via* all three platforms carried out. It is suspected that the FASP protocol, which includes multiple steps involving spin column centrifugation, was not sufficiently optimized for the brand of cartridges or the centrifugation speed available in the Doucette laboratory. While the approach clearly shows strong potential for SDS cleanup, it is advised that each laboratory conducts its own assessment of SDS cleanup approaches. However, the results shown here clearly demonstrate a favourable strategy for SDS cleanup, both in terms of improved protein identification, and technical simplicity. Moreover, the FASP method is not an option for top down proteomic workflows and it can be much more

costly (compared to the simpler organic precipitation protocols) when considering multiple fractions.

Table 5.1: Proteins and peptides identified from GeLC and GELFrEE using chloroform-methanol-water (CMW), acetone and the spin column procedure (FASP) for sample preparation.

	GeLC	GELFrEE CMW	GELFrEE Acetone	GELFrEE FASP
Peptides	4333	4059	4898	1663
Proteins	1327	1213	1320	676

5.3.3 GELFrEE of Urine Proteins

It has just been shown that SDS-based protocols, namely GeLC and GELFrEE are two effective protein separation platforms. An advantage of these techniques is that the proteins can be separated at the intact level, ultimately enabling a top down MS analysis of the proteins. Currently, urine proteins are studied extensively *via* one or two dimensional SDS-PAGE. Although these platforms have proven to be useful for urine proteomics, new strategies for urine protein analysis could prove more effective. As shown above, GELFrEE of yeast proteins proved to be comparable to the GeLC platform, yet revealed additional benefits in terms of fewer missed cleavages and higher protein sequence coverage. Given these findings, a GELFrEE separation of urine proteins was conducted to determine the compatibility of this platform for urine protein analysis.

An important question to address is the initial preconcentration method for recovery of proteins from urine. An advantage to the use of acetone precipitation is that urinary sediment proteins are not excluded from the sample-they are simply combined with the precipitated proteins. This allows for complete characterization of the proteome. SPE extraction requires that all precipitated material be removed from the sample prior to column loading. Thus, by the SPE protocol, it is inevitable that the sediments are separated from the supernatant prior to protein extraction. The advantages and disadvantages of SPE and acetone extraction protocols as well as GELFrEE-MS^2 of urine proteins are the subject of this section.

To first gain an overall sense of the ability of urine proteins to be separated by GELFrEE, a urine sample was acetone precipitated and separated *via* GELFrEE (Figure 5.8). It is apparent from the figure that urine proteins are successfully separated using this intact separation platform. Also seen in the figure is that proteins from all mass ranges are visible, indicating that separation of the urinary proteins is unbiased over the entire mass range of the sample.

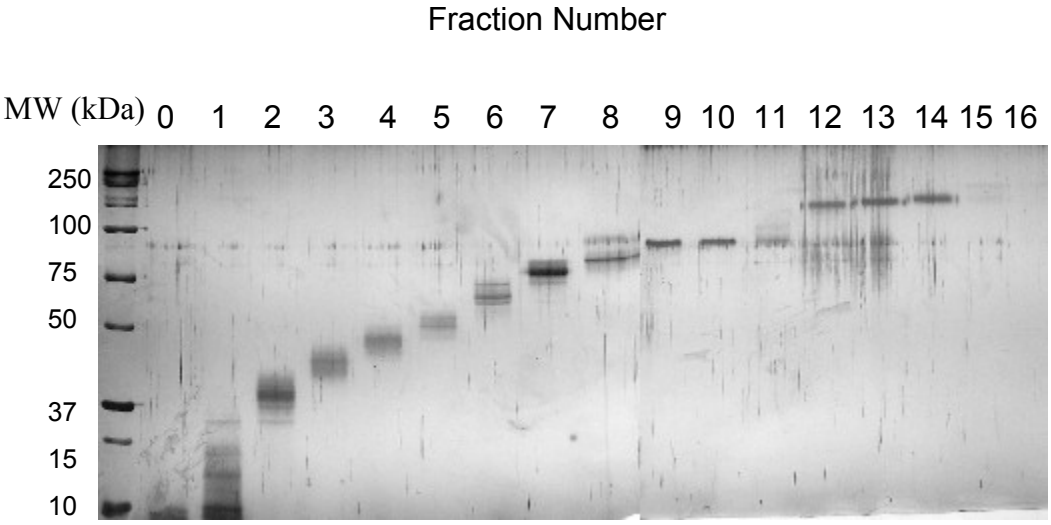


Figure 5.8: 1D-SDS-PAGE image resulting from analysis of GELFrEE fractions obtained from separation of urinary proteins. The proteins were isolated by acetone precipitation and include the sediment portion of the sample.

.

It was of particular interest to determine which protocol for protein extraction, namely acetone precipitation or SPE extraction, was most suited for protein extraction prior to GELFrEE fractionation. To investigate this, proteins were isolated by each of the two protocols, subjected to GELFrEE separation and the resulting fractions were visualized *via* SDS-PAGE. The same volume of urine (1.5 mL) was used for each extraction technique. Given that sediments were separated from the supernatant prior to the SPE protocol, this portion of the sample was fractionated in its own GELFrEE lane alongside the supernatant proteins. As can be seen from Figure 5.9, vortexing the samples aided tremendously in solubilizing sediment proteins for both the SPE and acetone samples, thus very little sediment proteins remained.

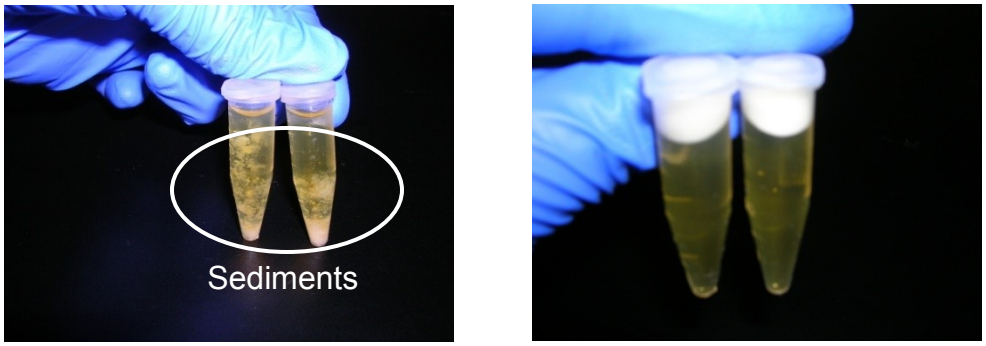


Figure 5.9: Images of the urine samples used for SPE and acetone protein extraction after a freeze-thaw cycle before vortexing (left) and after vortexing (right).

Following the addition of Laemmli sample buffer to the samples for GELFrEE analysis, it was noticed that some insoluble species remained in the acetone precipitated sample. Closer examination of the insoluble material led to the appearance of salt crystals. This is not surprising since it has been observed that acetone will precipitate some salts as well as protein (not shown). The SPE sample dissolved readily in the sample buffer.

Observation of the GELFrEE separation showed that proteins in the acetone lane exhibited lower mobility and did not focus relative to the SPE extract and sediment lanes (Figure 5.10). This indicated that resolution in the gel column of the acetone precipitated sample was much lower than that of the SPE sample. This is consistent with the fact that the acetone sample presumably contained more salts which interfered with the electrophoretic process.

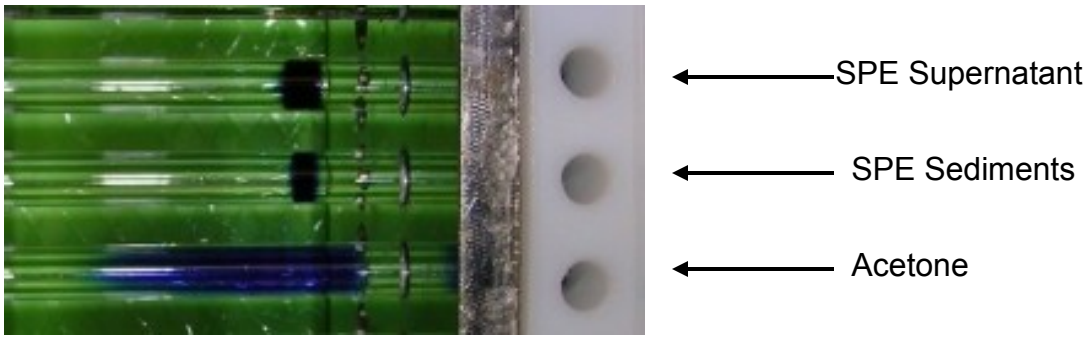


Figure 5.10: Image of the SPE extract of the supernatant, the remaining sediments and acetone extracted proteins from urine as they are fractionated by GELFrEE.

The fractions isolated from each lane were visualized by SDS-PAGE (Figure 5.11). As seen in the figure, the acetone precipitated proteins did not begin to elute from the GELFrEE column until fraction 6. Again, this is consistent with the fact that the sample was lagging behind during the electrophoresis compared to the SPE and sediment samples.

An initial conclusion to this observation would be that acetone precipitated proteins in GELFrEE is disadvantageous and so this preparation should be avoided in GELFrEE analysis. However, from examination of subsequent fractions, it is evident that an effective GELFrEE separation was still obtained and that distribution was simply shifted to a later time. Moreover, the concentration of acetone precipitated proteins, as determined by the intensity of the gel bands in the figure, are much higher than for SPE. Also seen in the image is that the sediment portion of the SPE sample is richer in protein

than the SPE supernatant sample. This is despite the fact that the sample was vigorously vortexed to resolubilize proteins from the freeze-thaw cycle. This is strong evidence that SPE extraction is only reliable if the sediment proteins are not discarded. Given that the acetone precipitated proteins essentially represents a combination of both the supernatant and sediment proteins, this preparation should be considered most representative of the entire urine proteome.

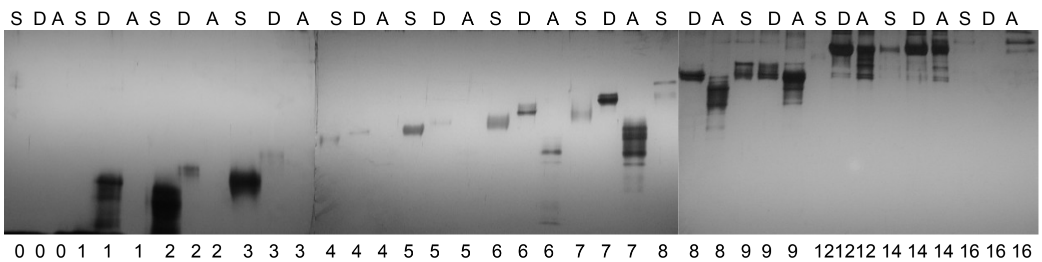


Figure 5.11: 1D-SDS-PAGE of urinary GELFrEE fractions in which the urine proteins were initially extracted using SPE and acetone protocols. From the top of the gel image, S represents the SPE supernatant, D represents the SPE sediment sample and A represents the proteins prepared by acetone. Indicated beneath each lane is the GELFrEE fraction number.

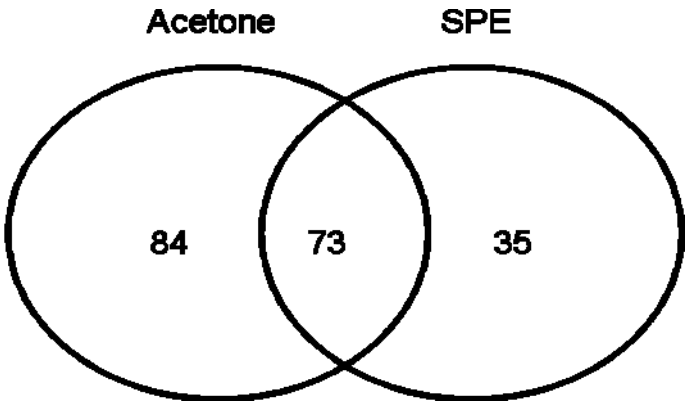
5.3.4 GELFrEE-MS Analysis of Acetone and SPE Extracted Proteins

For a full comparison, LC-MS^2 analysis of each GELFrEE fraction was conducted. To minimize bias between the two extraction protocols with respect to urinary sediments, a fresh urine sample was used such that very little sediments exist.

A quantitative assessment of protein concentration in the GELFrEE fractions was made. Following precipitation and digestion of the proteins from GELFrEE fractions, each sample was subjected to RPLC cleanup. From this cleanup, the total peak area of the eluting peptides was 2.6 times higher from the acetone precipitated samples than the SPE samples. This indicated that there were more peptides present in the acetone extracted samples compared to the SPE samples. This is consistent with the SDS-PAGE image shown in Figure 5.11.

MS analysis of SPE and acetone fractions were consistent with concentration of protein in the respective fractions. The number of identified proteins was 157 for acetone versus 108 for SPE (Figure 5.12). The number of peptides identified for the SPE (257) was also less than that of acetone (368). Given that each fraction was only analysed once by LC-MS^2, as expected, the list for both protein extraction protocols is relatively small. However, it was not the goal of this experiment to obtain the largest list possible; rather, this experiment was completed to investigate which protein extraction protocol (acetone or SPE) should be employed before GELFrEE-LC-MS^2. Further work will undoubtedly have to be explored to optimize downstream manipulations of the sample for urine proteome analysis.

Proteins



Peptides

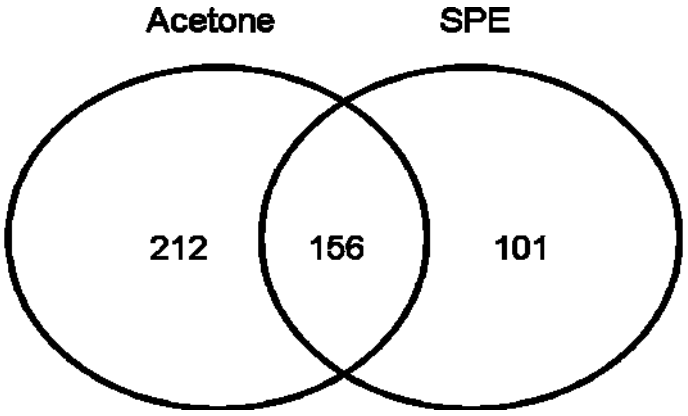


Figure 5.12 : Venn diagrams displaying the urinary proteins (top) and peptides (bottom) identified *via* acetone and SPE protocols.

Although the acetone precipitation resulted in a higher number of identified proteins, this difference is not drastically higher, indicating that either approach is acceptable depending on the goal of the experiment. For example, the SPE procedure may be favoured in instances where a protein assay is required, as it is much easier to dissolve the sample for a protein assay compared to trying to redissolve an acetone pellet in solvents compatible for the assay. On the other hand, acetone precipitation allows for the co-precipitation of sediment and supernatant proteins and thus the entire urine proteome can be processed with no concerns of protein loss due to discarding sediments. Upon combining SPE and acetone fractions, a total of 192 unique proteins were identified.

5.4 Conclusions

This chapter demonstrated that GELFrEE fractionation reveals comparable results to GeLC-MS, both in terms of the number and type of identified proteins for bottom up MS. Furthermore, the mass distribution and resolution of proteins was comparable with GeLC being slightly better at resolving proteins. An advantage to GELFrEE is that proteins are detected with higher sequence and a lower amount of missed cleavage sites compared to GeLC. Also, the potential for top down MS, or possibly other types of experiments with intact proteins, remains solely with GELFrEE as proteins must be digested prior to extraction from a gel slice in a GeLC experiment.

When comparing SPE extraction of proteins to acetone precipitation for GELFrEE-LC-MS^2, acetone gives higher protein yields and as a result yields a larger number of identified proteins. It is therefore recommended that acetone precipitation be

employed over SPE to extract urine proteins for GELFrEE. Moreover, if acetone precipitation is employed, sediments can be analyzed directly with the rest of the sample.

CHAPTER 6: INTACT ISOTOPE LABELLING OF URINE PROTEINS AND SEPARATION *VIA* GELFrEE

6.1 Introduction

As mentioned in the introduction to this thesis, isotope labelling for the purpose of protein quantitation is a very useful tool in proteomics. Of particular interest for this thesis is an intact protein labelling strategy. This is due to the fact that GELFrEE and GeLC, two intact separation platforms, require labelling at the intact level in order to avoid apparent differences in the sample resulting from the separation. Thus, if GELFrEE and GeLC separation platforms are to be employed in a quantitative study, labelling at the intact level is the only strategy that can be employed. As mentioned earlier in this thesis, intact protein analysis or "top down" proteomics has an incredible number of advantages and so this approach is favoured.

It was the purpose of this work to try and develop an intact isotope labelling strategy for the quantitation of urine proteins. Amine-specific dimethylation of lysine residues was particularly favourable due to the fact that the bottom up protocol is rapid and straight forward and thus should be easily adaptable to a top down strategy. Also, the reagents are not expensive compared to other commercially available labelling reagents. Another feature of this labelling strategy is that is has not been performed before and so if successful, it would enable another labelling approach for intact proteins. Currently, there are few published strategies for intact labelling of proteins. Hanash and co-workers employ a dual stable isotope coding strategy [121] and Schmidt *et al*. developed the novel isotope-coded protein label (ICPL) technique [122]. Due to the limited amount of

resources for intact protein labelling, another technique would be beneficial to the top down platform.

The formaldehyde labelling approach [31] has primarily been directed as a peptide-level modification and is currently only performed on peptides. Incomplete labelling has been observed by other researchers, namely Dr. Devanand Pinto (personal communication) when attempted at the intact protein level. Presumably, this is due to the fact that intact proteins often conceal labelling sites because of their native conformation and so not every residue capable of being labelled is exposed to do so.

In addition to this challenge, there is no current software available to interpret dimethylated labelled lysine residues from intact proteins. When conducted at the peptide level, the amino terminus is exposed as well as the lysine residues. However, for intact labelling, the amino terminus of the peptide will not be labelled since it is not exposed, and thus existing software will wrongfully interpret MS spectra (Figure 6.1). As a consequence, the program will yield no 'hits' as the masses will be shifted entirely from the expected mass.

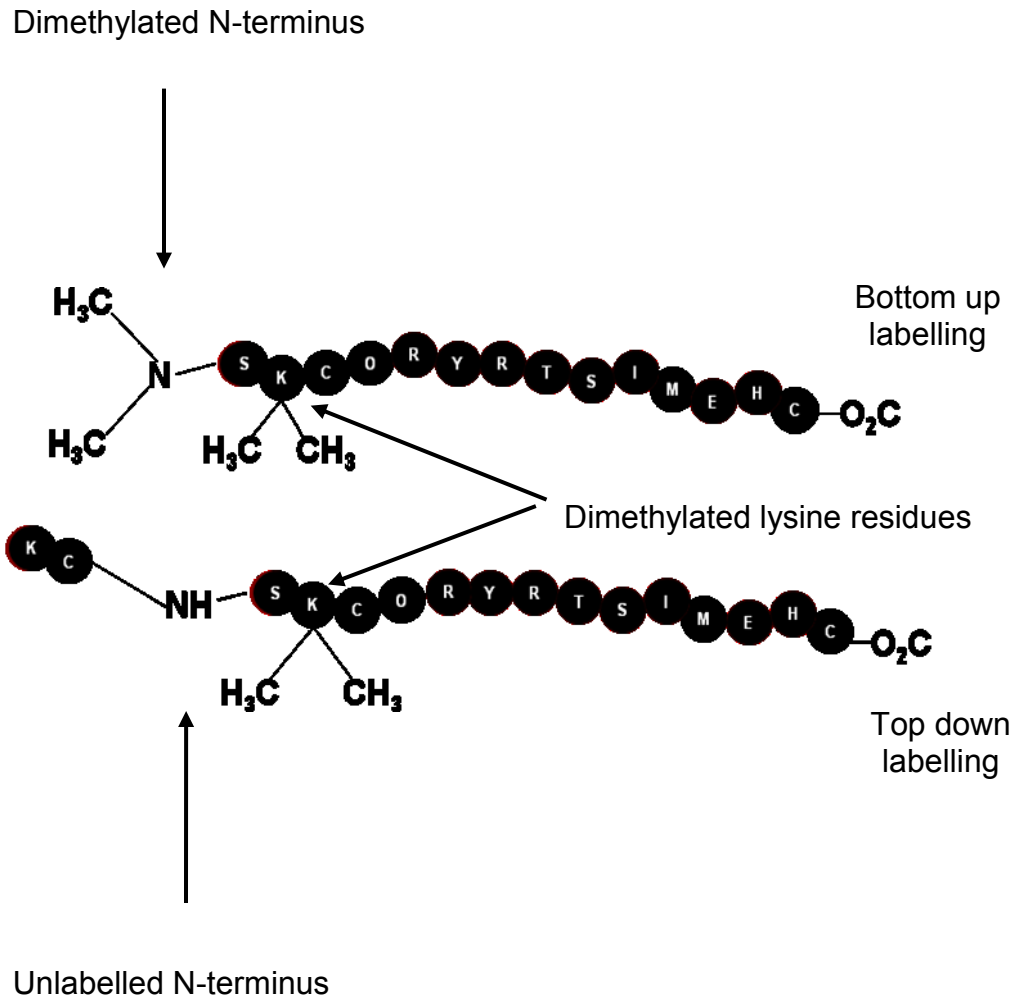


Figure 6.1: Diagram showing the difference in labelling with formaldehyde when performed at the bottom up level versus the top down.

Yet another concern with this strategy is that typical protein digestion is conducted with trypsin. For peptide-level dimethyl labelling, trypsin digestion occurs prior to labelling thus no digestive problem exists. However, if labelling occurs before

digestion, the labelled lysine residues would prevent tryptic digestion at this site. Thus, digestion would presumably only occur at arginine residues. To summarize, the following three complications have hindered the intact formaldehyde labelling of proteins;

1. Incomplete labeling due to unexposed lysine residues
2. Database searching is only available for bottom up strategy
3. Trypsin is not an ideal enzyme for digestion of the labelled protein

Each of these complications is addressed in the following sections.

6.2 Materials and Methods

6.2.1 Solvents and Reagents

Reagents for formaldehyde labelling (formaldehyde, D_2-formaldehyde, sodium cyanoborohydride, triethylammonium bicarbonate) were a donation from the Institute for Marine Biosciences in Halifax, N.S. All solvents came from Fisher (Ottawa, Canada). All proteins/enzymes were obtained from Sigma (Oakville, Canada). Protein assay reagents, dithiothreitol, iodoacetamide, urea and SDS were obtained from Bio-Rad (Mississauga, Canada).

6.2.2 Extraction of Urine Proteins

Three × 10 mL aliquots of human urine, obtained from a single healthy donor, were successively loaded onto a solid phase extraction column (Varian, Mississauga, Canada) extracted with 0.5 mL of 30% acetonitrile/0.1% trifluoroacetic acid (TFA) water followed by 0.5 mL of 50% acetonitrile/0.1% TFA water. The extracts from the original 30 mL of urine were pooled (3 mL total). The extract was assayed by Bio-Rad protein assay, and a 200 μg and 100 μg sample was aliquoted, dried to completely remove solvent in a speed vacuum, and subjected to intact protein labelling (section 6.2.3), with a goal of obtaining 2:1 ratio of 'light' to 'heavy' formaldehyde-labelled samples.

6.2.3 Isotope Labelling of Proteins

Dried protein samples were resuspended in 0.1 M triethylammonium bicarbonate at a concentration of 0.5 μg/μL. Prior to labelling, all samples were reduced with dithiothreitol and alkylated with iodoacetamide. Following reduction and alkylation, samples were spiked with either 8 M urea (urea samples) or 10% SDS (SDS samples) to give a final concentration of 4 M urea and 1% SDS respectively. All samples were incubated for 5 min at 35 ºC to facilitate denaturation of the proteins. The samples were reacted with either formaldehyde (200 μg sample) or deuterated (D_2) formaldehyde (100 μg sample) and incubated at room temperature for 10 min with gentle agitation. 6 M $NaCNBH_3$ was added to each sample (8.35 μL per 100 μg protein) and allowed to incubate overnight with gentle agitation. Samples were combined such that the ratio of 'light' to 'heavy' samples was 2:1. The pooled sample was precipitated overnight with 4 volumes cold acetone.

6.2.4 GELFrEE of Labelled Urine Proteins

The resulting protein pellet from acetone precipitation was dissolved in 50 µL of 1 × Laemmli gel buffer [37] and loaded onto a 12% T tube gel for GELFrEE analysis. Voltage was held at 240 V for the first 10 min where it was then increased to 300 V for the remainder of the run. When the dye front began to elute from the tube gels, fraction collection was started and were as follows; collection of dye front (fraction 0), 5 × 1 min, 5 × 2 min, 3 × 5 min and 3 × 15 min for a total of 17 fractions. Random fractions were selected (fractions 0, 1, 3 and 6) for precipitation with either chloroform-methanol-water or acetone followed by digestion and analysis by LC-MS^2.

6.2.5 1D-SDS-PAGE for Visualization of GELFrEE Fractions

10 µL of each labelled fraction from the urea and SDS fractions was mixed with 8 µL of 5 × gel sample buffer and 22 µL of water. 20 µL was loaded into each well of a 12% T SDS-PAGE slab gel and the gels were run at 120 V until the dye front reached the bottom of the gel. Gels were silver stained for visualization.

6.2.6 Digestion of Samples

Sample pellets from GELFrEE fractions were dissolved in 20 µL of 8 M urea with multiple pumps to aid solubilization. Samples were diluted with a sufficient amount of 50 mM ammonium bicarbonate to bring the pH to 8.0. Either 100 ng of trypsin or 100 ng of chymotrypsin was added to the sample and incubated overnight at 37 ºC. Digestion was then stopped by addition of 10 µL of 10% TFA.

6.2.7 Sample Cleanup by RPLC

Prior to LC-MS^2, samples were cleaned using a 1200 HPLC system (Agilent Technologies, Mississauga, Canada). The column used was a self packed 1 mm × 50 mm C18 column. Solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in acetonitrile. The gradient was as follows; 0% B for 5 min hold, jump to 70% B at 5 min, 5 % B at 10 min, held at 5% for 5 min, and ramped to 70% by 20 min. Collection was obtained over the time interval 18.5-28 min.

6.2.8 LC-MS^2 and Database Searching of Fractions

Peptide LC-MS^2 analysis was performed using a 1200 Agilent nano-HLPC system with autosampler coupled to a Thermo LTQ mass spectrometer with a nanospray ionization source. Chromatography was performed on a home packed 75 μm × 25 cm C18 spray tip (New Objectives) packed with 4 μm beads ('Jupiter' beads from Phenomenex). Column flow was maintained at 0.3 μL/min and the gradient was as follows; 8% B at 0 min, 28 %B at 85 min, 5% B at 85.01 min, 5% B at 90 min, 60 % B at 95 min and 5 % B at 95.01 min. Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. Mass spectrometric analysis was performed in a data dependent fashion using the triple play method as described in section 4.2.3. Database searching was employed using the MASCOT search engine. Parameters for the search were as follows; human database, 4 missed cleavages, fixed modification of carbamidomethyl C, variable modifications including oxidation of methionine, phosphorylated tyrosine and propionamide cystine. Mass tolerance for MS and MS^2 was set to 0.5, a probability of 0.05 or less and an ion cut off score of 40. For the searches

indicating intact labels, identical search criteria were used with the addition of indicating that proteins were labelled at the intact level. Quantitation was performed using the MASCOT Distiller program on all hits determined from the MASCOT search with ion score cut off removed. A correlation of 0.9 with a standard error of 0.08 or less was considered a reliable quantitation.

6.3 Results and Discussion

6.3.1 Qualitative Investigation of Labelling

It was presumed that denaturation of the intact protein prior to labelling would be sufficient to ensure that all lysine residues were exposed and hence labelled. To investigate this, an identical urine sample was aliquoted such that 200 µg and 100 µg were isotopically labelled with the normal and deuterated (D_2) formaldehyde respectively. These amounts were chosen such that successful labelling would produce a 2:1 ratio for light to heavy labels as observed in the mass spectrum.

Urea and SDS were chosen as denaturants for the labelling process. Each of these species was spiked into the sample prior to labelling, with gentle heating to assist denaturation. Following labelling, the samples were precipitated with acetone, though it was observed that the urea samples produced very little pellet. Additional acetone was added to induce precipitation, producing a small amount of pellet a though significantly less was seen than for the SDS samples. The pellets were suspended in gel buffer and subjected to separation by GELFrEE. SDS-PAGE analysis of the resulting GELFrEE fractionations revealed little protein in the samples containing urea as compared to the

SDS-spiked samples (Figure 6.2). It was suspected that the presence of a high concentration of urea prevents precipitation of the sample, and thus, given the incompatibility with an existing proteome workflow, these samples were not explored further in the labelling experiment.

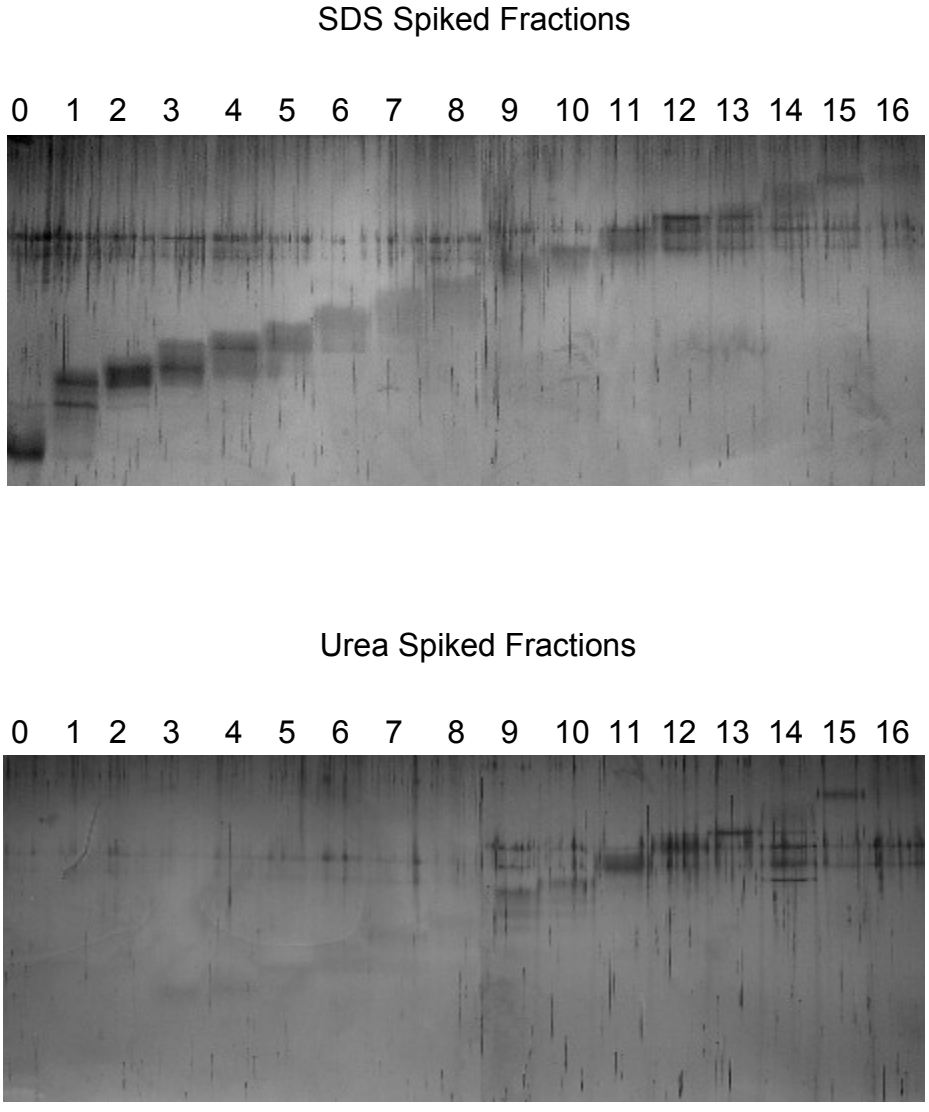


Figure 6.2: SDS-PAGE images showing the labelled samples spiked with SDS (top) and urea (bottom) following fractionation by GELFrEE.

Following fractionation by GELFrEE, the samples were acetone precipitated and prepared for digestion. Since it was presumed that all lysine residues were isotopically labelled, it was believed that trypsin would not be the optimal digestive enzyme. Due to labelled lysine residues, exclusive arginine cleavage would result in larger peptide fragments. Chymotrypsin was chosen as an alternative enzyme for digestion. Chymotrypsin cleaves at aromatic residues (tyrosine, tryptophan and phenylalanine) and so would not be hindered by dimethylated lysine residues. Trypsin was still used in a separate set of fractions to compare the relative benefits of each enzyme. Following digestion, a random selection of fractions was subjected to RPLC cleanup followed by LC-MS^2 to investigate whether the labelling process was successful.

Initial evidence of successful labelling was observed upon manual examination of the MS spectra. Both trypsin and chymotrypsin digested fractions appeared to have protein peaks with a spacing of 4 *m/z* units which would indicate that the peptides detected were possible labelled pairs. Figures 6.3 and 6.4 show the spectra with possible labelled pairs obtained from both trypsin and chymotrypsin digested proteins. Moreover, examination of these selected peptides revealed intensities which were as expected, where the light labelled sample was approximately double the heavy sample.

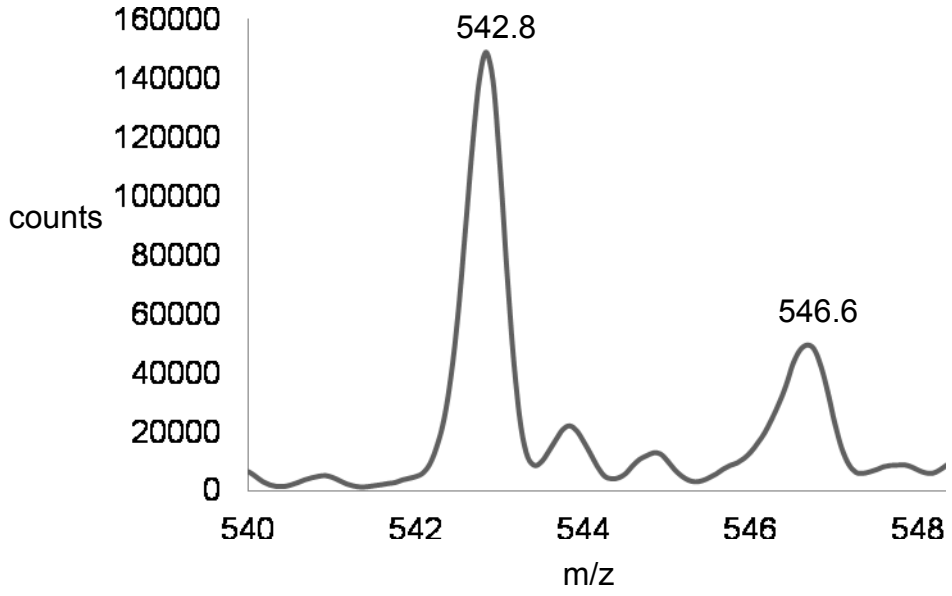


Figure 6.3: MS selected ion chromatogram from the MS analysis of peptides digested by chymotrypsin after intact formaldehyde labelling and GELFrEE fractionation. Light: heavy ratios appear to be close to 2:1.

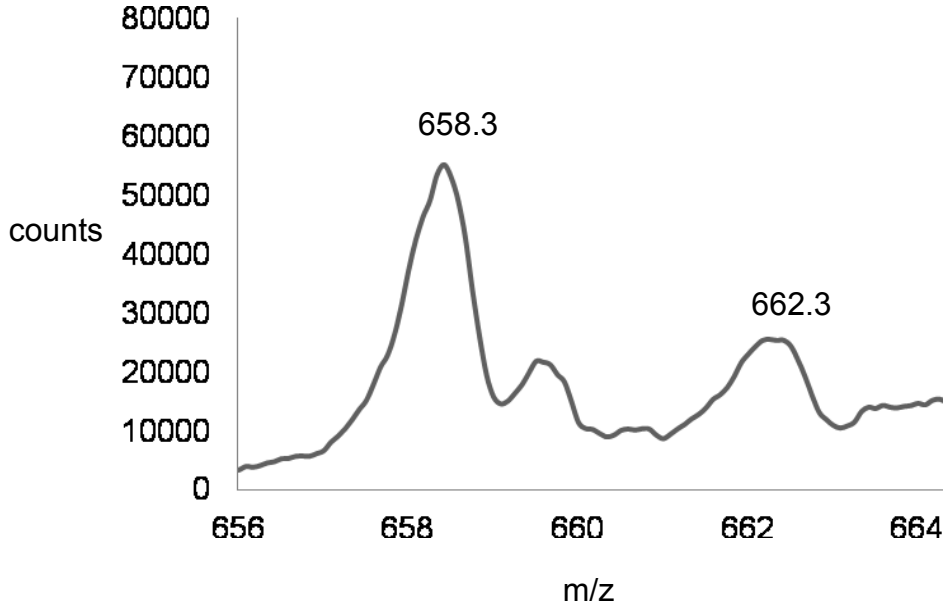


Figure 6.4: Selected ion chromatogram from the MS analysis of peptides digested by trypsin after intact formaldehyde labelling and GELFrEE fractionation. Light: heavy ratios appear to be close to 2:1.

For confirmation that the labelling was indeed successful, it was necessary to interpret the resulting MS spectra through spectral database searching. As mentioned previously, the current database for interpretation of dimethylated proteins assumes that the N-terminus is also labelled.

Through collaboration with personnel responsible for computer programming at the Institute for Marine BioSciences in Halifax NS, a new feature was programmed in MASCOT, which was termed "IDOG Intact". This feature allows MASCOT to search labelled lysine residues only, maintaining an unmodified amino terminus. To test the IDOG Intact feature, the labelled samples were searched and compared to the same data when searched using no labelling parameters. If labelling was successful, then database searching with no labelling parameters would result in the identification of peptides that do not contain lysine residues (*i.e.* unaffected by dimethylation). Unless labelling was incomplete, lysine-containing peptides would not be seen in this search, but would be through the IDOG Intact feature. This apparently redundant form of searching ensured that proteins were not identified simply due to a forced identification (false positive).

One of the most abundant urine proteins, alpha-1-microglobulin (AMBP) was chosen for the investigation. This protein was identified from multiple peptides, containing some with lysine and some without. Figure 6.5 displays the amino acid sequence of alpha-1-microglobulin and Table 6.1 illustrates the AMBP peptides that were identified using the conventional MASCOT search with no labels indicated. Table 6.2 shows the same sample as searched by MASCOT including the IDOG Intact feature. Searches were performed for both the trypsin and chymotrypsin digested samples.

Alpha-1-microglobulin amino acid sequence

MRSLGALLLLLSACLAVSAGPVPTPPDNIQVQENFNISRIYGKWYNLAIGSTCPW
LKKIMDRMTVSTLVLGEGATEAEISMTSTRWRKGVCEETSGAYEKTDTDGKFY
HKSKWNITMESYVVHTNYDEYAIFLTKKFSRHHGPTITAKLYGRAPQLRETLLQ
DFRVVAQGVGIPEDSIFTMADRGECVPGEQEPEPILIPRVRRAVLPQEEEGSG
GGQLVTEVTKKEDSCQLGYSAGPCMGMTSRYFYNGTSMACETFQYGGCMGN
GNNFVTEKECLQTCRTVAACNLPIVRPCRFIQLWAFDAVKGKCVLFPYGGCQG
NGNKFYSEKECREYCGVPGDGDEELLRFSN

Figure 6.5: The amino acid sequence of human alpha-1-microglobulin.

Table 6.1: The tryptic and chymotryptic peptides of alpha-1-microglobulin through MASCOT search engine without specifying any labels. Also indicated is the number of labels found. The modification indicated is on the amino acid residue underlined in the peptide sequence. Also indicated in bold is the MASCOT score of the protein.

UNMODIFIED PROTEIN MASCOT SEARCH	
Trypsin	**Chymotrypsin**
MASCOT SCORE: 308	**MASCOT SCORE: 511**
R.ETLLQDFR.V	Y.SAGPCMGMTSRY.F + Oxidation (M)
R.TVAACNLPIVR.G	Y.SAGPCMGMTSRY.F + 2 Oxidation (M)
R.TVAACNLPIVR.G + Propionamide (C)	Y.CGVPGDGDEELLRF.S
R.TVAACNLPIVRGPCR.A	Y.VVHTNYDEYAIFL.T
R.EYCGVPGDGDEELLR.F	F.RVVAQGVGIPEDSIF.T
R.EYCGVPGDGDEELLRFSN.-	Y.CGVPGDGDEELLRFSN.-
	L.VLGEGATEAEISMTSTRW.R
	L.VLGEGATEAEISMTSTRW.R + Oxidation (M)
	F.TMADRGECVPGEQEPEPIL.I + Oxidation (M)
	L.QTCRTVAACNLPIVRGPCRAF.I

Table 6.2: The tryptic and chymotryptic peptides of alpha-1-microglobulin through MASCOT search engine incorporating the IDOG Intact feature. Peptides not found in Table 6.1 are bolded. Also indicated is the number of labels found. Light labels are indicated by 0P while heavy labels are indicated by 4P. The modification indicated is on the amino acid residue underlined in the peptide sequence. Also indicated in bold is the MASCOT score of the protein.

DIMETHYL LABELS OF LYSINE MASCOT SEARCH (IDOG INTACT)	
Trypsin	**Chymotrypsin**
MASCOT Score: 310	**MASCOT Score: 574**
R.ETLLQDFR.V	Y.SAGPCMGMTSRY.F + Oxidation (M)
R.TVAACNLPIVR.G	Y.SAGPCMGMTSRY.F + 2 Oxidation (M)
R.TVAACNLPIVR.G + Propionamide (C)	Y.CGVPGDGDEELLRF.S
R.TVAACNLPIVRGPCR.A	Y.VVHTNYDEYAIFL.T
R.EYCGVPGDGDEELLR.F	F.RVVAQGVGIPEDSIF.T
R.EYCGVPGDGDEELLRFSN.-	**W.AFDAVKGKCVLFPY.G + 2 IDOG +0P**
	W.AFDAVKGKCVLFPY.G + 2 IDOG +4P
	Y.CGVPGDGDEELLRFSN.-
	L.VLGEGATEAEISMTSTRW.R
	L.VLGEGATEAEISMTSTRW.R + Oxidation (M)
	F.TMADRGECVPGEQEPEPIL.I + Oxidation (M)
	F.QYGGCMGNGNNFVTEKECL.Q + IDOG +0P
	L.QTCRTVAACNLPIVRGPCRAF.I

Several interesting conclusions can be made from the examination of these two tables. First, using trypsin as the digestive enzyme, only arginine terminated peptides are observed. If labelling was successful, then lysine residues would not be cleaved, thus the absence of such peptides is an indication of successful labelling.

Second, from Table 6.1, it can be seen that the chymotrypsin digested proteins yielded more peptides than a tryptic digest, resulting in a significant increase of the

MASCOT score (511 from chymotrypsin *vs* 308 from trypsin). Chymotrypsin is unaffected by lysine labelling. Moreover, upon examination of the tryptic peptides, it can be seen that none contain lysine residues since any lysine containing peptide (presumed to be labelled) would not be identified.

From Table 6.2, while the MASCOT search should now be able to correctly interpret MS spectra of labelled peptides, no additional peptides were identified from the tryptic digest. While this result was unexpected, examination of peptides from chymotryptic digestion did reveal several additional peptides. These new peptides correspond to lysine-containing segments of the protein, all of which have the dimethyl label incorporated.

To thoroughly examine the efficiency of the labelling process more peptides should be examined. Table 6.3 lists the peptides identified from tryptic and chymotryptic digests, of additional proteins, as determined *via* MASCOT searching with and without the IDOG Intact feature for protein labelling. Only the proteins that displayed differences between the two searches are listed.

Table 6.3: Labelled proteins digested with trypsin and their corresponding peptides identified *via* MASCOT search with and without considering dimethyl labelling.

UNMODIFIED SEARCH	DIMETHYL LABELS SEARCHED
Insulin-like growth factor-binding protein 7	
R.YPVCGSDGTTYPSGCQLR.A	R.YPVCGSDGTTYPSGCQLR.A
	R.KGKAGAAAGGPGVSGVCVCKSR.Y +3 IDOG
KRT13 Isoform 1 of Keratin, type I cytoskeletal 13	
R.LQSSSASYGGGFGGGSCQLGGGR.G	R.LQSSSASYGGGFGGGSCQLGGGR.G
R.TLQGLEIELQSQLSMK.A	
KRT9 Keratin, type I cytoskeletal 9	
R.SGGGGGGGLGSGGSIR.S	R.SGGGGGGGLGSGGSIR.S
R.GGSGGSHGGGSGFGGESGGSYGGGEEASGSGGGYGGGSGK.S	
KRT2 Keratin, type II cytoskeletal 2 epidermal	
R.GFSSGSAVVSGGSR.R	R.GFSSGSAVVSGGSR.R
K.FASFIDKVR.F	
SPINK1 Pancreatic secretory trypsin inhibitor	
Not found	R.QTSILIQKSGPC.- + IDOG
NEGR1 Neuronal growth regulator 1	
Not found	R.VSISTLNKR.D + IDOG
HSPG2 Basement membrane-specific heparan sulfate proteoglycan core protein	
R.YQLGSGEAR.L	R.YQLGSGEAR.L
R.SPGPNVAVNAK.G	
R.FSSGITGCVK.N	
R.RGSIQVDGEELVSGRSPGPNVAVNAK.G	
CCNL2 Isoform 1 of Cyclin-L2	
Not found	K.NTKRRLEGAK.K + 2 IDOG
CD300LG Isoform 1 of CMRF35-like molecule 9	
R.KGGILFSR.C	R.KGGILFSR.C + IDOG
Myosin-reactive immunoglobulin kappa chain variable region (Fragment)	
R.LLIYGASTR.A	Not found
WFDC2 Isoform 2 of WAP four-disulfide core domain protein 2	
R.DQCQVDSQCPGQMK.C	Not found
KNG1 Isoform LMW of Kininogen-1	
R.IGEIKEETTSHLR.S	Not found
BTN2A1 butyrophilin, subfamily 2, member A1 isoform 2 precursor	
R.TTFVSKDISR.G	R.TTFVSKDISR.G + IDOG

All differences observed between the two searches of tryptic peptides involve those which contain lysine residues. For example, Insulin-like growth factor-binding protein 7 had an additional peptide identified by the labelled search, which is consistent with proper labelling peptide. However, as seen with the WFDC2 Isoform 2 of WAP four-disulfide core domain protein 2, for example, an unmodified lysine-containing peptide was identified in the non-labelled search. Thus, it can be concluded from the tryptic peptides the labelling was not 100% efficient. The same trend was observed with the chymotrypsin peptides (Table 6.4). Considering less than 100% efficiency, it would be preferred that the MASCOT search considers dimethyl labelling as a variable modification. Thus, both labelled and unlabelled peptides could be identified in a single search.

Table 6.4: Labelled proteins digested with chymotrypsin and their corresponding peptides identified *via* MASCOT search with and without considering dimethyl labelling.

UNMODIFIED SEARCH	DIMETHYL LABELS SEARCHED (IDOG INTACT)
PTGDS Prostaglandin-H2 D-isomerase	
Y.SQGSKGPGEDF.R	not found
Y.SQGSKGPGEDFRMATLY.S + Oxidation (M)	
APOD Apolipoprotein D	
Y.WILATDYENY.A	Y.WILATDYENY.A
	F.HLGKCPNPPVQENFDVNKY.L + 2 IDOG
	F.HLGKCPNPPVQENFDVNKY.L + 2 IDOG
CTRB1 Chymotrypsinogen B	
W.QVSLQDKTGF.H	not found
ATP9B Isoform 1 of Probable phospholipid-transporting ATPase IIB	
Y.QGGILMY.G + Oxidation (M); Phospho (Y)	not found
ITIH4 Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4	
F.SSHVGGTLGQF.Y	F.SSHVGGTLGQF.Y
	L.LLLSDPDKVTIGL.L + IDOG
LMAN2 Vesicular integral-membrane protein VIP36	
F.LDTYPNDETTERVFPY.I	not found
AGRN Agrin	
Not found	W.LGGLPELPVGPALPKAY.G + IDOG

6.3.2 A Quantitative Investigation of Labelled Peptides

As mentioned above, the proteins were labelled in a 2:1 ratio of light to heavy samples, thus the intensity ratios observed in MS should also be 2:1. Before one determines peak ratios, it must first be decided which peaks are considered 'good' hits, based on the correlation feature of the MASCOT distiller program. This feature considers the isotope pattern of the experimental spectra *vs* that of a theoretical isotope pattern. A

correlation of at least 0.90 was required for reliable spectra. Figure 6.6 shows the spectral quality for a peptide that had a correlation of 0.91. For comparison, a correlation of 0.7 displayed poor agreement between the experimental and theoretical patterns (Figure 6.6, bottom).

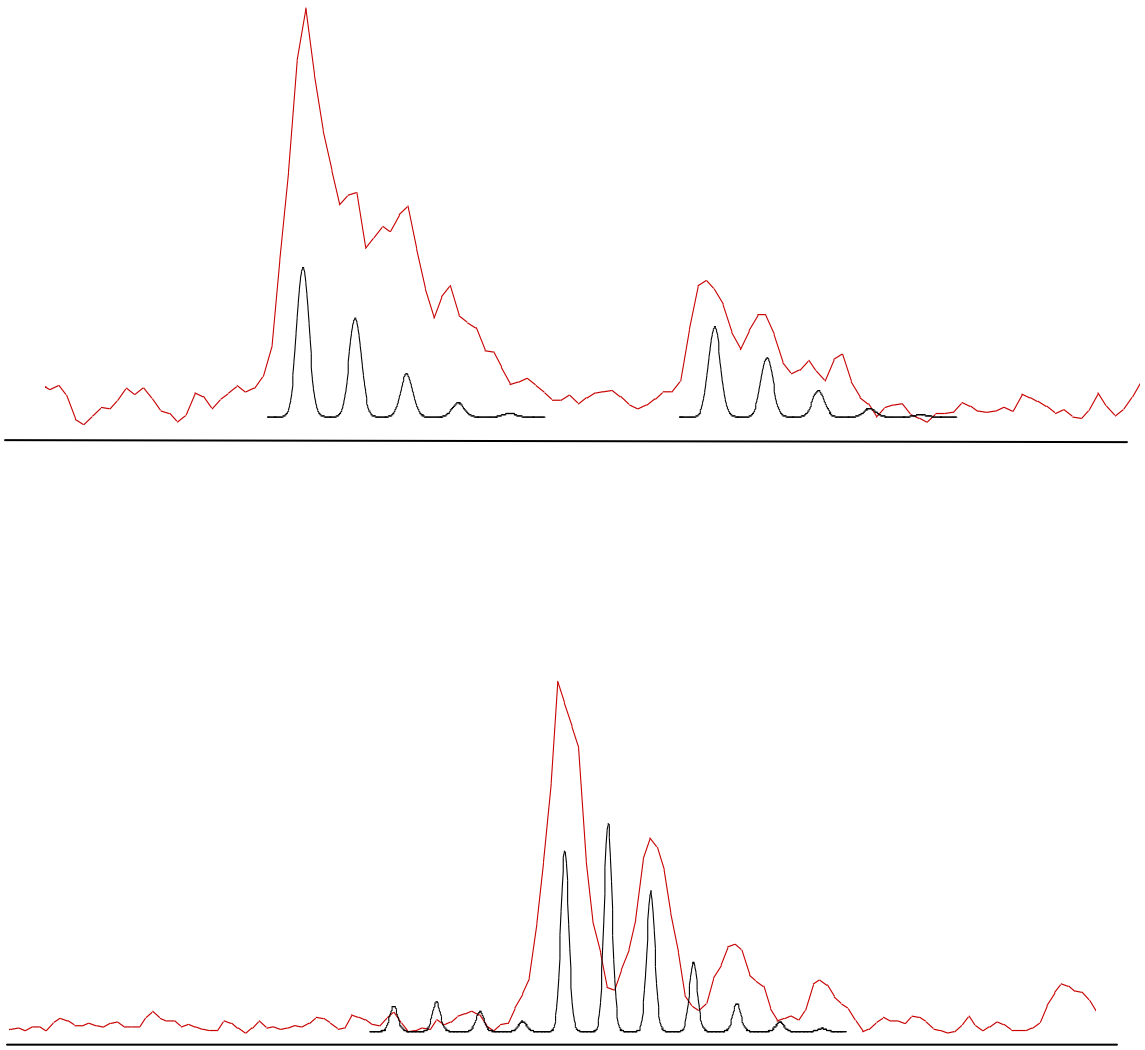


Figure 6.6: Correlation spectra of two labelled peptides with a correlation value of correlation: 0.91 (top) and 0.7 (bottom). The theoretically generated spectrum is indicated in black while the experimental spectrum is indicated in red.

In addition to the correlation value of the spectra, the standard error of the ratio calculation was also considered. The standard error is based on a least squares fit to the component intensities from the scans in the extracted ion chromatogram and is a good measure of the reliability of the correlation value. A standard error less than 0.08 was set as the critical value. In these set of experiments, for *initial* filtering of spectra, ratios would only be considered if the peptide peaks possessed a correlation of at least 0.90, and a standard error of 0.08 or lower.

Table 6.5 shows the ratio of the heavy: light samples determined from proteins that met all filtering criteria. Because chymotrypsin was the better digestive enzyme, quantitation was performed only on these peptides. The peptide score cut off value of 40 (set during the MASCOT search) was removed as a filter to enable more peptides to be examined. This is justified by the added filtering *via* correlation and standard errors. As seen in Table 6.5, MASCOT protein scores remain high. It is also noted that only 4 of the 17 GELFrEE fractions were analyzed, thus only a small list of proteins met the required quantitation criteria.

Table 6.5: Human urinary proteins quantified from intact labelling. A 2:1 ratio of light to heavy is expected.

Protein Name	MASCOT Score	Light/heavy
26 kDa protein	151	2.2
Alpha-1 microglobulin	1021	1.6
Apolipoprotein D	173	1.7
Hemopexin	58	1.2
Ig kappa chain C region	208	1.9
IGK @ protein	181	2.6
Isoform BI-1(V1) of voltage-dependent P/Q-type calcium channel subunit alpha-1A	45	1.8

As seen in the table, the expected 2:1 ratio was approximately met (average ratio 1.9 +/- 0.4 from 7 proteins). While such results are obtained from a limited number of proteins, the table demonstrates the applicability of the intact labelling strategy to the quantitation of urinary proteins. The recommended procedure for such a platform is outlined below.

Recommended Procedure for Isotope Labelling of Intact proteins

1. Isolate proteins from control and test samples.

2. Prepare to label 100 – 200 μg of protein.

3. Aliquot the proper amount of protein for test and control samples, dry to completeness.

4. Resuspend protein in 0.1 M triethylammonium bicarbonate such that the concentration of protein in each sample (test and control) is 0.5 μg/ μL. Pump solution through pipette multiple times to ensure proper suspension.

5. Vortex for 20 seconds.

6. Centrifuge at 13,000 rpm for 1 min.

7. Reduction: Add a stock DTT solution to a final concentration of 5 mM. Pump pipette multiple times while adding DTT. Incubate at 56 ºC for 30 min.

8. Alkylate: Add a stock iodoacetamide solution to a final concentration of 12.5 mM. Pump pipette multiple times while adding. Incubate at room temperature for 1 h in the dark.

9. Denature the proteins: Add a stock SDS solution so that the final concentration is 1% of the solution. Pump pipette multiple times while adding. Vortex the sample for 30 seconds, centrifuge for 1 min at 13,000 rpm and heat at 35 ºC for 5 min.

10. Labelling: Add 14.3 µL of 6.98 M deuterated formaldehyde (heavy label) and the same amount of 6.98 M formaldehyde (light label) for a 200 µg sample. Adjust if only labelling 100 µg. Pump pipette multiple times to ensure proper mixing.

11. Vortex for 30 seconds.

12. Centrifuge at 13,000 rpm for 1 min.

13. Incubate for 10 min at room temperature.

14. Add 16.7 µL of 6 M $NaCNBH_3$ for a 200 µg sample. Adjust if only labelling 100 µg. **This compound is EXTREMELY TOXIC, perform all actions in a properly vented fume hood.** Pump pipette multiple times to ensure proper mixing.

15. Combine samples in a 1:1 ratio.

16. Precipitate with cold acetone overnight at -20 ºC.

17. Centrifuge precipitates at 13,000 rpm for 15 min, decant and dry pellets.

18. Dissolve protein pellets in gel loading buffer so that protein is a concentration (estimated) of 2 µg/µL. Load 50 µL (equivalent of 100 µg) onto a GELFrEE tube gel and perform GELFrEE.

6.4 Conclusions

A number of conclusions can be drawn from the data shown above. First, SDS is an effective denaturant for proteins in a sample prior to intact labelling. Urea is not favourable as it is difficult to precipitate the samples following labelling, and thus is not generally compatible with the quantitative proteome workflow incorporating GELFrEE separation. Both trypsin and chymotrypsin are able to produce peptides from intact labelled proteins, however, more labels and higher protein scores can be observed from the chymotrypsin samples. An indication that the labelling is less than 100% efficient was provided by the fact that unlabelled lysine-containing peptides were identified from the MASCOT searches. Further work will be necessary to determine the extent of incomplete labelling and the influence on quantitation. However, the data presented in this chapter provides strong evidence that intact labelling of proteins with formaldehyde, followed by separation of the proteins by GELFrEE, is a very promising new platform for the quantitation of proteins.

CHAPTER 7: INTACT LABELLING AND GELFrEE FRACTIONATION OF RAT URINE FOR BIOMARKER DISCOVERY

7.1 Introduction

A variety of kidney disorders were described in Chapter 2 of this thesis. Of particular interest to this project was ureteropelvic junction kidney obstruction (UPJO), a disorder in which the passage of urine from the kidney to the bladder is inhibited. In order to conduct an investigation as to whether there exists a urinary protein biomarker (or series of biomarkers) for this disorder, rats were surgically altered to provide a model for the obstruction disorder. This was performed by inserting a pin in the renal pelvis which essentially acted as the obstruction. The rat urine was collected over the next few weeks. In addition to the rats who received the operation, urine samples from a corresponding set of control rats were collected. The intact protein labelling, and GELFrEE separation workflow was chosen for exploratory analysis on the presence of protein biomarkers.

7.2 Materials and Methods

7.2.1 Collection of Control and Test Urine Samples

Rat urine samples were obtained from Dr. Dawn MacLellan's lab at the Izaak Walton Killam Health Centre in Halifax, Nova Scotia. A rodent model for urinary tract obstruction was employed, wherein rats were surgically altered to have total blockage of the urinary tract. Urine from healthy and completely obstructed rats were collected over a three week time course study, combining the sample with protease inhibitor in a 10:1 ratio; samples were kept at -80ºC until ready for analysis.

7.2.2 Preparation of Urine Samples for Intact Labelling and GELFrEE

One set of healthy (control) and of obstructed (test) urine samples were chosen for analysis. These samples were taken from the third week of the experiment. The samples were thawed and 2 mL from each set was precipitated with 8 mL of cold acetone overnight at -20 ºC. Following precipitation the samples were centrifuged at 3716 × g for 15 min. The supernatants were decanted and the pellets allowed to air dry. 400 µL of 4 M urea was used to solubilize each protein pellet, using multiple pumps with the pipette to aid in resolubilization. Remaining sediments were centrifuged for 6 min at 13,000 rpm to remove the insoluble materials. The soluble portion of each sample was subjected to a Bio-Rad protein assay and 75 µg from each sample was taken for labelling. Labelling was performed exactly as described in Chapter 6 using the recommended procedure. Samples were combined in a 1:1 ratio, combining test with control, and control with control. The control samples were labelled with the heavy label and the obstructed samples were labelled with the light label.

7.2.3 GELFrEE of Labelled Samples

The labelled, pooled proteins were separated by GELFrEE as described in section 6.2.4. Collected fractions were acetone precipitated overnight at -20ºC with 400 µL of cold acetone. Pellets were centrifuged at 13,000 rpm for 15 min and two washes were employed each using 400 µL acetone.

7.2.4 Digestion of GELFrEE Samples

Acetone-precipitated proteins were digested as described in section 6.2.6 with the exception that 100 ng of chymotrypsin was used as the digestive enzyme.

7.2.5 MS of Proteins and Database Searching

MS analysis was conducted as described in section 6.2.8. Database searching was conducted with the MASCOT search engine with the IPI rat database. Parameters for the search were as follows; 'IDOG Intact' specified, 4 missed cleavages, chymotrypsin as the digestive enzyme, ion cut off score of 40, probability of 0.05 or less, MS and MS^2 tolerance of 0.5. Quantitation was performed using the MASCOT Distiller program and was set to quantify all hits. Quantitation ratios were only considered if the correlation value of spectra was at least 0.9 and the standard error was less than 0.08.

7.3 Results and Discussion

To validate the quantitative workflow, an assessment of the heavy *vs* light labelled control samples was performed. Following quantitation of all MASCOT protein hits using MASCOT Distiller, a list of 825 candidate proteins was generated with a protein score of 40 or higher. Upon examination of all proteins meeting the criteria for reliable quantitation, 13 of these proteins remained (Table 7.1).

Table 7.1: List of control –control rat urinary proteins meeting all requirements for reliable quantitation.

Protein name	MASCOT Accession #	MASCOT Score	H:L Ratio
Alpha-2-HS-glycoprotein	IPI00327469	206	0.48
Alpha-2-microglobulin	IPI00201616	322	0.87
C4-2 complement component 4, gene 2	IPI00422037	57	0.82
Cadherin-1	IPI00206662	301	1.1
carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	IPI00365582	91	0.43
Glandular kallikrein-7, submandibular/renal	IPI00201592	203	1.5
kallikrein 1-related peptidase b3	IPI00231193	131	1.6
LOC500183 protein	IPI00568389	289	1.0
Major urinary protein	IPI00191711	254	0.50
Serum amyloid P-component	IPI00327745	121	0.98
similar to androgen binding protein zeta	IPI00388042	104	2.9
Urinary protein 3	IPI00205971	286	0.72
Uromodulin	IPI00198021	1455	0.98
Average of all ratios			**1.1**
Standard deviation of all ratios			**0.65**

Examination of the table shows that most of the ratios of heavy to light samples are approximately 1, which is the expected ratio for these pooled control samples. However, up to a 3 fold variation in protein concentration was observed. This nonetheless provides evidence that the labelling of the proteins as well as the quantitation parameters are reliable. Examination of the correlation spectra of alpha-2-microglobulin (Figure 7.1)

reinforces the fact that the quantitation parameters are acceptable. Not only does this protein meet all cut-off requirements for reliable quantitation, but its spectrum is very impressive.

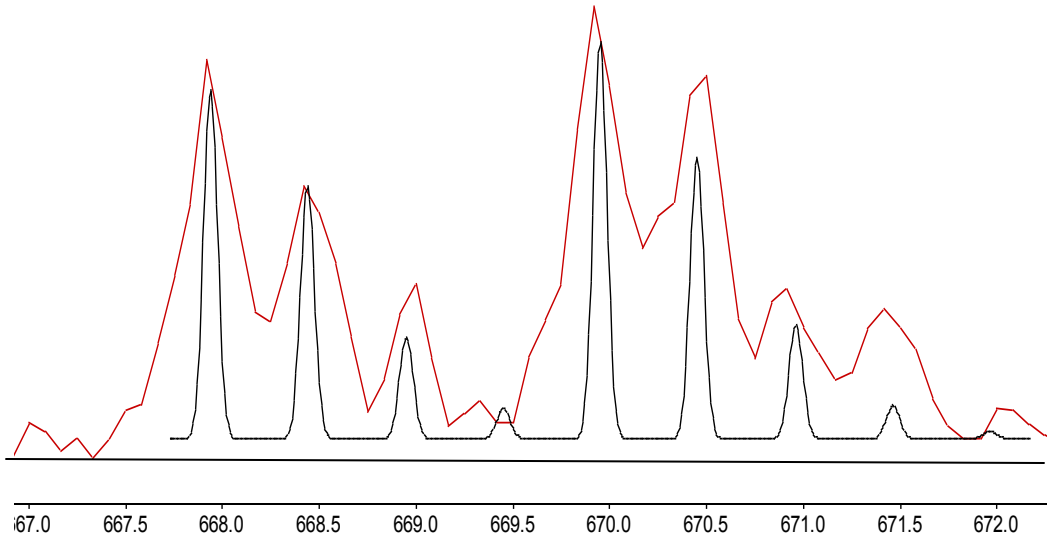


Figure 7.1: Correlation spectrum for alpha-2-microglobulin, observed at a ratio of 0.87:1 for heavy to light samples. This protein from the control samples met all requirements for reliable quantitation.

The test *vs* control samples (*i.e.* obstructed vs healthy rats) were examined using the same criteria as for the controls. A list of 712 candidate proteins was obtained. After filtering, only 5 proteins could be reliably quantified (Table 7.2).

Table 7.2: Quantitation of proteins from the obstructed *vs* unobstructed rat samples. Control samples were given the heavy label and obstructed samples were given the light label.

Protein Name	MASCOT Accession #	MASCOT score	H:L
Cadherin-1	IPI00206662	504	1.1
kallikrein 1-related peptidase b3	IPI00231193	385	0.95
olfactory receptor Olr476	IPI00362879	53	57
Serine protease inhibitor A3K	IPI00200593	543	1.1
Utrophin 392 kDa protein	IPI00212030	86	0.079

Examination of Table 7.2 shows two proteins from this list which can be designated as candidate biomarker proteins. One of these proteins, utrophin 392 kDa protein showed a 12 fold increase in the obstructed rat model. Figure 7.2 reveals the correlation spectrum for this protein, which also indicated that the quantitation ratio was reliable.

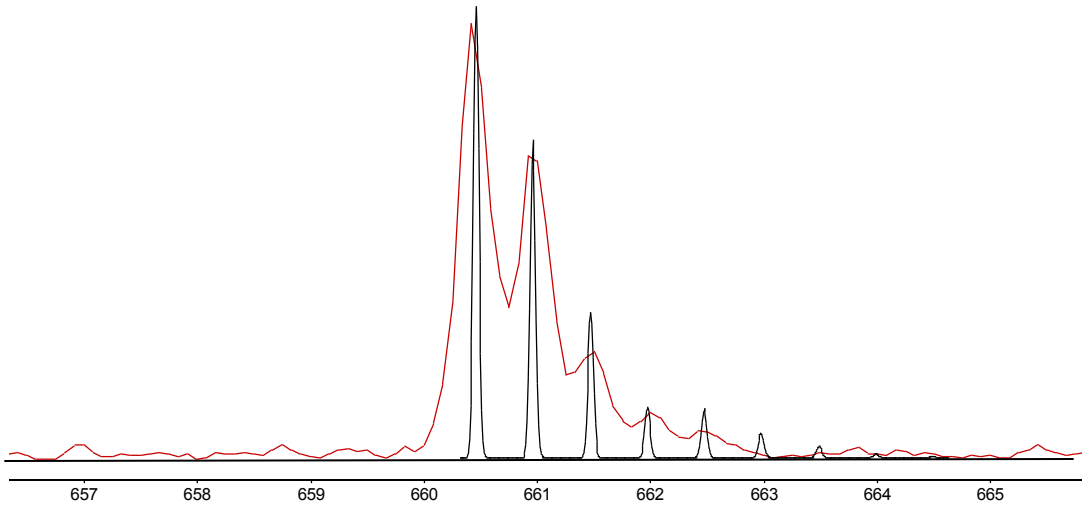


Figure 7.2: Correlation spectrum for utrophin, a protein that was 12 fold lower in the control sample compared to the obstructed sample. The correlation value was determined to be 0.963 and the standard error was determined to be 0.07.

When surveying the literature for properties of utrophin proteins, it was interesting to discover that the utrophin proteins belong to a family of proteins known as the dystrophin glycoprotein complex (DGP) [123]. Specifically, the DGP is a multimeric protein complex that is composed of either dystrophin or its autosomal homologue, utrophin [123]. Upon further investigation, it was discovered that DGP is expressed in the central and peripheral nervous system as well as tissues with a secretory function or that form barriers between two or more functional components in a biological system. Such tissues include the blood brain barrier, choroid plexus and the kidney [124]. Other

researchers have been able to show *via* high-resolution immunofluorescence imaging and Western blotting that DGP proteins exist in the kidney nephrons of mice [125,126]. Recalling from introductory chapters, the nephron is the main filtering unit of the kidney. Upon obstruction of a kidney, the nephron's filtering capacity would be hindered and as a result, proteins regularly filtered from the nephron to the bladder (*i.e.* into urine) such as utrophin may be altered, resulting in a change in expression of this protein compared to in urine from a healthy kidney.

Even more interesting is the growing body of evidence suggesting that the DGP protein complex plays an important role in the kidney. For example, Raats *et al.* [125] provide evidence that utrophin is involved in linking the podocyte cytoskeleton to the glomerular basement membrane, indicating a structural importance for this protein in the kidney. Moreover, immunolocalization studies performed on a mouse kidney revealed that utrophin was found in the glomeruli and collecting ducts of the kidney [126]. All of the scientific evidence indicating the role of utrophin in the kidney as well as the 12 fold factor increase discovered in the obstructed samples provides support that this protein may be a candidate urinary biomarker for kidney obstruction and should be investigated further.

Perhaps the most interesting protein in Table 7.2 is the olfactory receptor protein, which was 57 fold higher in the control than in the obstructed samples (Figure 7.4). Olfactory receptors are proteins that are chemosensors for chemical odorants [127]. Interestingly, the mammalian species contains over 1,000 olfactory receptor proteins,

each of which is able to signal one specific protein (a G protein) upon detection of an odorant molecule [127].

Intuitively, one would not think that an olfactory receptor protein would have any function with the kidney. However, an extremely interesting study conducted by Pluznick *et al*. [127] sought to answer this very question. Because olfactory receptor proteins responded to external chemical odorant molecules, signalling the G protein pathway for odorant detection, it was hypothesized by these researchers that perhaps the olfactory receptor proteins in kidney epithelial cells were able to respond to chemical changes in the chemical system internally. The researchers hypothesized further that the kidney was an ideal organ to study in order to answer this question since it is the kidney that must adjust glomerular filtration, readsorption of ions and proteins as well as many other functions in response to its environment. The researchers believed that olfactory receptors in the kidney were being used as chemosensors in order to detect changes in its internal environment and adjust itself accordingly to avoid complications such as stone formation or improper glomerular filtration.

It was concluded that the olfactory proteins were located in the distal nephron and macula densa within mouse and rat kidney. It was further implicated that "olfactory machinery may play a physiologically critical role in regulating fundamental aspects of renal function" [127]. This is particularly interesting for this thesis, as it was determined that the olfactory receptor protein was 57 fold higher in the control sample. It is very possible that this protein was down-regulated in the obstructed samples due to the fact that the kidney essentially stops functioning, creating a 'null' chemical environment with

no need for the olfactory receptor protein. In any case, it is very interesting that this protein was essentially not detected in the obstructed sample (Figure 7.3).

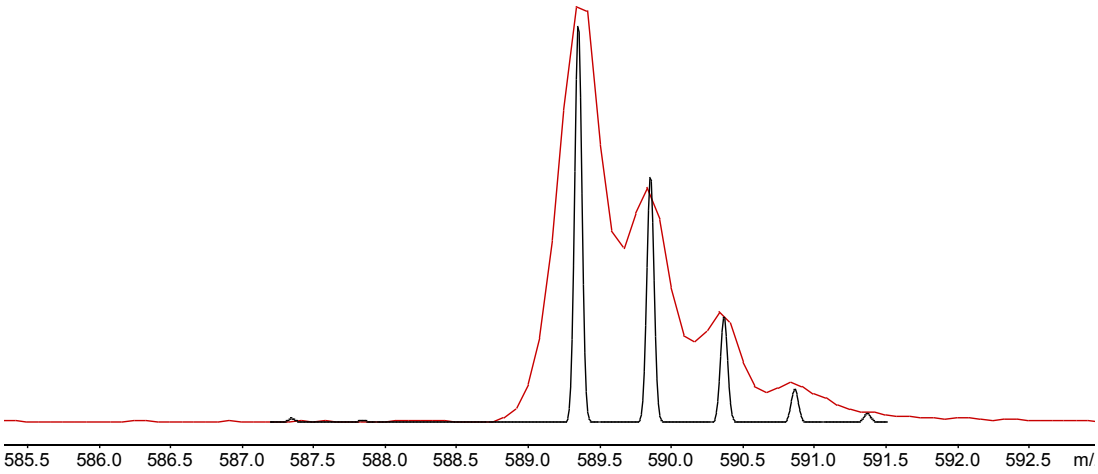


Figure 7.3: The correlation spectra for Olfactory receptor protein, a protein from rat urine identified to be 57 fold higher in the control sample. The correlation for this spectrum was determined to be 0.97 and the standard error was found to be 0.04.

7.4 Conclusions

Intact labelling followed by GELFrEE separation of rat urinary proteins isolated from healthy and obstructed kidneys were analyzed by MS^2. A comparison of the healthy sample to itself revealed quantitative protein ratios of heavy to light labelled proteins of approximately 1. Upon comparison of the healthy samples to the obstructed samples, it was discovered that two proteins, namely utrophin 392 kDa protein and olfactory receptor protein may be biomarker candidates for kidney obstruction. Moreover, from examination of the literature, each protein was scientifically shown to have significance with the structure and function of the kidney [125-129]. This reinforces the possibility that these two proteins could play a significant role in kidney obstruction disorders.

CHAPTER 8: CONCLUSIONS

The work presented in this thesis focussed on techniques for the proper analysis of urinary proteins. It was shown in chapter 3 that urinary sediments are rich in protein and that some of these proteins are unique to the sediments while some are common to both supernatant and sediment phases. This chapter showed *via* MS analysis of sediment and supernatant proteins that in order to investigate the urinary proteome with minimal bias, the sediments must be analyzed and considered an important component of the urinary proteome.

With the goal of an intact protein separation in mind for the analysis of urinary proteins, a solution-based separation platform, specifically GELFrEE, was proposed. However, this being a technique which incorporates SDS to impart separation, it was realized that the MS tolerance towards SDS needed to be established. GELFrEE fractions are known to contain up to 2% SDS and so it was crucial to establish at what level SDS needed to be removed prior to MS analysis. Chapter 4 presented the MS tolerance to SDS as well as effective methods for reducing SDS in GELFrEE samples below the threshold level of 0.01% SDS.

As GELFrEE is a relatively new SDS-based intact separation protocol, it was important to investigate how this new technique compared to the leading proteome profiling technique, which corresponds to another SDS-based separation occurring in gels. The GeLC experiment for separation detection of yeast proteins was compared to that of a GELFrEE workflow for comprehensive proteome characterization. While it is acknowledged that the existing MS instrumentation in this laboratory does not permit

complete proteome coverage, these two techniques were nonetheless shown to be of similar level of performance in terms of total number of identified proteins. However, each method of sample preparation has its advantages, noting in particular that GELFrEE permits top down MS, and provided a higher number of peptide hits than the GeLC experiment. Urinary proteins could also be subjected to the GELFrEE workflow, noting that initial protein extraction has considerable effect on the separation. Acetone precipitation yields more proteins than SPE extraction and is preferred due to the fact that urinary sediments could be precipitated directly with the supernatant proteins.

Chapter 6 showed a novel strategy for the intact dimethylation of lysine residues for the purpose of protein quantitation. This intact labelling strategy was intended to be compatible with proteome prefractionation platforms, including GELFrEE separation. The technique proved to be successful given that labelling was conducted for SDS-denatured proteins and by employing chymotrypsin as the digestive enzyme. Also presented in this chapter was a new MASCOT search program that was utilized to interpret the dimethylation of intact proteins.

Chapter 7 applied the newly developed intact labelling and separation strategy to rat urine obtained from healthy rats and those whom had kidney obstruction. It was revealed that two proteins were candidate biomarkers for this disorder. Utrophin 392 kDa protein was found to be 12 fold decreased in the control compared to the obstructed sample. In addition, the olfactory receptor protein was a factor of 56 higher in the control sample compared to the obstructed sample. A survey of the literature indicated that both

of these proteins are associated with the kidney and so it is very possible that they both may be protein biomarkers for kidney obstruction disorders.

8.1 Future Work

Much of the future work surrounding urinary protein biomarker discovery involves the intact labelling strategy. It still needs to be determined as to the percent of proteins that are not effectively labelled. Future experiments for this determination would involve high resolution TOF experiments on labelled standards and urine samples to see ratios of labelled to unlabelled peaks.

In addition, the MASCOT program for searching data from intact labelled proteins needs to be improved so that it is able to search for labelled and unlabelled proteins concurrently. This should be able to be achieved with more manipulation of the current intact search, using a series of protein standards and urine proteomes as a test. In addition, the candidate biomarker proteins discovered in this thesis need to be examined in further detail, choosing more time points from the rat urine samples, and extending to a larger population. The results indicated in this thesis are preliminary, as replicate experiments on urine from healthy and obstructed kidneys will need to be analyzed to reveal whether these proteins consistently implicate themselves as biomarkers for this disorder.

REFERENCES

1. Nature Biotechnology Editorial. A Cast of Thousands. *Nat. Biotech.* **2003**, *21*, 213.

2. Sanger, F.E.; Thompson, O.P.; Kitai, K. The Amide Groups of Insulin. *Biochem. J.* **1955**, *59*, 509-518.

3. Edman, P.; Henschen, A. method for Determination of the Amino Acid Sequence in Peptides. *Acta. Chem. Scand.* **1950**, *4*, 283-293.

4. 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.

5. Hillencamp, F.; Karas, M.; Beavis, R. C.; Chait, B.T. Matrix-Assisted Laser Desorption Ionization Mass Spectrometry of Biopolymers. *Anal. Chem.* **1991**, *63*, A1193-A1202.

6. 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.

7. Kebarle, P.; Verkerk, U.H. Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. *Mass Spetrom. Rev*. **2009**, *28*, 898-917.

8. Karas, M.; Kruger, R. Ion Formation in MALDI: The Cluster Ionization Mechanism. *Chem. Rev.* **2003**, *103*, 427-439.

9. Kriwacki, R.; Siuzdak, G. Protein Structure Characterization with Mass Spectrometry. *Spectroscopy, an International Journal* **2004**, *18*, 37-47.

10. Ashraf, M. Inducing Drought Tolerance in Plants. *Biotechnol. Adv.* **2010**, *28*, 169-183.

11. Pappin, D.J.; Hojrup, P.; Bleasby, A.J. Rapid Identification of Proteins by Peptide-Mass Fingerprinting. *Curr. Biol.* **1993**, *3*, 327-332.

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

13. Mann, M.; Hojrup, P.; Roepstorff, P. Use of Mass Spectrometric Molecular Weight Information to Identify Proteins in Sequence Databases. *Biol. Mass Spectrom.* **1993**, *22*, 338-345.

14. James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. Protein Identification by Mass Profile Fingerprinting. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58-64.

15. Yates, J.R.; Speicher, S.; Griffin, P.R.; Hunkapiller, T. Peptide Mass Maps: A Highly Informative Approach to Protein Identification. *Anal. Biochem.* **1993**, *214*, 397-408.

16. Clauser, K.R.; Baker, P.; Burlingame, A.L. Role of Accurate Mass Measurement (+/- 10 ppm) in Protein Identification Strategies Employing MS or MS-MS and Database Searching. *Anal. Chem.* **1999**, *71*, 2871-2882.

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

18. Lee, J.E.; Kellie, J.F.; Tran, J.C.; Tipton J.D.; Catherman, A.D.; Thomas, H.M.; Ahlf, D.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.

19. Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72*, 1156-1162.

20. Marshall, A.; Hendrickson, C.L.; Jackson, G.S. Fourier Transform Ion Cyclotron Mass Spectrometry: A Primer. *Mass Spectrom. Rev*. **1998**, *17*, 1-35.

21. Boyne, M.T.; Garcia, B.A.; Li, M.X.; Zamdborg, L.; Wenger, C.D.; Babai, S.; Kelleher, N.L. Tandem Mass Spectrometry with Ultrahigh Mass Accuracy Clarifies Peptide Identification by database Retrieval. *J. Proteome Res.* **2009**, *8*, 374-379.

22. Gerber, S.A.; Rush, J.; Stemman, O.; Kirschnre M.W.; Gygi, S.P. Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS. *Proc. Natl. Acad. Sci.*. **2003**, *100*, 6940-6945.

23. Diadaroglu, M. Chemical determination of Free-Radical Induced damage to DNA. *Free Radical Biol. Med.* **1991**, *10*, 225-242.

24. Kuhn, E.; Wu, J.; Karl, J.; Liao, H.; Zolg, W.; Guild, B. Quantification of C-reactive Protein in the Serum of Patients With Rheumatoid Arthritis Using Multiple Reaction Monitoring Mass Spectrometry and C-13-Labelled Peptide Standards. *Proteomics* **2004**, *4*, 1175-1186.

25. Ong, S.E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D.B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labelling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression proteomics. *Mol. Cell Proteomics* **2004**, *4*, 1175-1186.

26. Yao, X.D.; Freas, A., Ramirez, J.; Demirev, P.A.; Fenselau, C. Proteolytic O-18 Labeling for Comparative Proteomics: Model Studies With Two Serotypes of Adenovirus. *Anal. Chem.* **2001**, *73*, 2836-2842.

27. Gygi, S.P.; Rist, B.; Gerber, S.A.; Turecek, F.; Gelb, M.H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotech.* **1999**, *17*, 994-999.

28. Ross, P.L.; Huang, Y.N.; Marchese, J.N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D.J. Mutiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-Reactive isobaric Tagging Reagents. *Mol. Cell Proteomics* **2004**, *3*, 1154-1169.

29. Elliott, M.H.; Smith, D.S.; parker, C.E.; Borchers, C. Current Trends in Quantitative Proteomics. *J. Mass Spectrom.* **2009**, *44*, 1637-1660.

30. Boersema, P.J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A.J.R. Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. *Nat. Protoc.* **2009**, *4*, 484-494.

31. Melanson, J.; Avery, S.L.; Pinto, D. High-Coverage Quantitative Proteomics Using Amine-Specific Isotope Labeling. *Proteomics* **2006**, *6*, 4466-4474.

32. Srivastava, G.; Singh, K.; Tiwari, M.N.; Singh, M.P. Proteomics in Parkinson's Disease: Current Trends, Translational Snags and Future Possibilities. *Expert Rev. Proteomics* **2010**, *7*, 127-139.

33. Lee, V.M.Y.; Balin, B.J.; Otvos, L.; Trojanowski, J.Q. A68-A Major Subunit of Paired Helical Filaments and Derivatized Forms of Normal-Tau. *Science* **1991**, *251*, 675-678.

34. Hutchens, T.W.; Yip, T.T. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 576-580.

35. Tucholska, M.; Bowden, P.; Jacks, K.; Zhu, P.H.; Furesz, S.; Dumbrovsky, M.; Marshall, J. Human Serum Proteins Fractionated by Preparative Partition Chromatography Prior to LC-ESI-MS/MS. *J. Proteome Res.* **2009**, *8*, 1143-1155.

36. 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.

37. Laemmli, U.K. Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. *Nature* **1970**, *227*, 680-&.

38. Wiese, S.; Reidegeld, K.A.; Meyer, H.E.; Warscheid, B. Protein Labeling by iTRAQ: A New Tool for Quantitative Mass Spectrometry in Proteome Research. *Proteomics* **2007**, *7*, 340-350.

39. Taylor, G.K.; Kim, Y.B.; Forbes, A.J.; Meng, F.; McCarthy, R.; Kelleher, N.L. Web and Database Software for Identification of Intact Proteins Using "Top-Down" Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 4081-4086.

40. Julian, B.A.; Suzuki, H.; Suzuki, Y.; Tomino, Y.; Spasovski, G.; Novak, J. Sources of Urinary Proteins and Their Analysis by Urinary proteomics for the Detection of Biomarkers of Disease. *Proteomics Clin. Appl.* **2009**, *3*, 1029-1043.

41. Image adapted from *http://www.myvisiontest.com/newsarchive.php?id=735*

42. Grasso, M.; Rajamahanty, S.; Gitlin, J. Ureteropelvic Junction Obstruction. Article published online at http://emedicine.medscape.com/article/450785-overview.

43. Decramer, S.; Wittke, S.; Mischak, H.; Zurbig, P.; Walden, M.; Bouissou, F.; Bascands, J.L.; Scanstra, J. Predicting the Clinical Outcome of Congenital Unilateral Ureteropelvic Junction Obstruction in Newborn by Urinary Proteome Analysis. *Nat. Med.* **2006**, *12*, 398-400.

44. Medical Term Dictionary Available online. http://www.beautynationpl.com/en/medical_terms.asp?pageID=b_dictionary

45. Krishna, J.; Shah, Z.; Merchant, M.; Klein, J.; Gozal, D. Urinary Protein Expression Patterns in Children With Sleep-Disordered Breathing: Preliminary Findings. *Sleep Medicine* **2006**, *7*, 221-227.

46. Tantipaiboonwong, P.; Sinchaikul. S.; Sriyam, S.; Phutrakul, S.; Tein-Chen, S.T. Different Techniques for Urinary Protein Analysis of Normal and Lung Cancer Patients. *Proteomics* **2005**, *5*, 1140-1149.

47. Adachi, J.; Kumar, C.; Zhang, Y.; Olsen, J.; Mann, M. The Human Urinary Proteome Contains More Than 1500 Proteins, Including a Large Proportion of Membrane Proteins. *Genome Biology* **2006**, *7*, R80-R80.16.

48. Pisitkun, T.; Johnstone, R.; Knepper, M. Discovery of Urinary Biomarkers. *Mol. Cell. Prot*. **2006**, *5*, 1760-1771.

49. Khan, A.; Packer, N. Simple Urinary Sample Preparation for Proteomic Analysis. *J. Prot. Res*. **2006**, *5*, 2824-2838.

50. Magistroni, R.; Ligabue, G.; Lupo, V.; Furci, L.; Leonelli, M.; Manganelli, L.; Masellis, M.; Gatti, V.; Cavazzini, F.; Tizzanini, W.; Albertazzi. A. Proteomic Analysis of Urine From Proteinuric Patients Shows a Proteolytic Activity Directed Against Albumin. *Nephrol. Dial. Transpl.* **2009**, *5*, 1672-1681.

51. Castagna, A.; Cecconi, D.; Sennels, L.; Rappsilber, J.; Guerrier, L.; Fortis, F.; Boschetti, E.; Lomas, L.; Righetti, P.G. Exploring the Hidden Human Urinary Proteome via Ligand Library Beads. *J. Prot. Res.* **2005**, *4*, 1917-1930.

52. Terracciano, R.; Pasqula, L.; Casadonte, F.; Frasca, S.; Preiano, M.; Falcone, D.; Savino, R. Derivatized Mesoporous Silica Beads for MALDI-TOF MS Profiling of Human Plasma and Urine. *Bioconjugate Chem*. **2009**, *5*, 913-923.

53. Kumar, Y.; Uppulluri, N.R.V.; Babu, K.; Phadke, K.; Kumar, P.; Ballal, S.; Tatu, U. Proteomics of Renal Disorders: Urinary Proteome Analysis by Two-Dimensional Gel Electrophoresis and MALDI-TOF Mass Spectrometry. *Curr. Sci.* **2002**, *82*, 655-663.

54. Meier, K.; Herrmann, A.; Knueppel, S.; Hillmann, M.; Koester, P.; Danne, T.; Haller, H.; Fliser, D.; Mischak, H. Identification of Urinary Protein pattern in Type 1 Diabetic adolescents With Early Diabetic Nephropathy by a Novel Combined Proteome Analysis. *J. Diabetes Complicat.* **2005**, *19*, 223-234.

55. 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.

56. Sharma, K.; Lee, S.; Han, S.; Lee, S.; Francos, B.; McCue, P.; Wassell, R.; Shaw, M.A.; RamachandraRao, S. P. Two-Dimensional Fluorescence Difference Gel Electrophoresis Analysis of the Urine Proteome in Human Diabetic Nephropathy. *Proteomics* **2005**, *5*, 2648-2655.

57. Kentsis, A.; Monigatti, F.; Dorff, K.; Campagne, F.; Bachur, R.; Steem, H. Urine Proteomics for Profiling of human Disease Using High Accuracy Mass Spectrometry. *Prot. Clin. Appl.* **2009**, *3*, 1052-1061.

58. O'Connor, J.F.; Ellish, N.; Kakuma, T.; Schlatterer, J.; Kovalevskaya, G. Differential Urinary gonadotropin profiles in early Pregnancy and Early Pregnancy Loss. *Prenatal Diagn.* **1998**, *18*, 1232-1240.

59. Tasian, G.E.; Aaronson, D.S.; McAninch, J.W. Evaluation of Renal Function After Major Renal Injury: Correltion With the American Association for the Surgery of Trauma Injury Scale. *J. Urol.* **2010**, *183*, 196-200.

60. Ramaswami, U.; Najafian, B.; Schieppati, A.; Mauer, M.; Bichet, D, G. Assesment of Renal Pathology and Dysfunction in Children With Fabry Disease. *Clin. J. Am. Soc. Nephrol.* **2010**, *5*, 365-370.

61. Fujimoto, T.; Takechi, S.; Machida, M.; Isu, N.; Imamura, W.; Katinoki, S.; Kanda, K.; Kobayashi, T.; Nomura, A. Excretions of Urinary Albumin and Various Proteins Increase in Hypertension. *Blood Pressure* **2008**, *17*, 270-273.

62. Sleat, D.E.; Zheng, H.; Lobel, P. The Human Urine Mannose 6-Phosphate Glycoproteome. *Biochem. Biophys, Acta*. **2007**, *1774*, 368-372.

63. Jenkins, D.; Woolf, A. S. Uroplakins: New Molecular Players in the Biology of Urinary Tract Malformations. *Kidney Int.* **2007**, *71*, 195-200.

64. Park, M.R.; Wang, E.H.; Chan, D. C.; Cha, J.H.; Lee, K.H.; Yang, C.W.; Kang, C.S.; Choi, Y.J. Establishment of a 2-D Human Urinary Proteomic Map in IgA Nephropathy. *Proteomics* **2006**, *6*, 1066-1076.

65. Hoi-Yee Ngai, H.; Sit, W.H.; Jiang, P.P.; Xu, R.J.; Man-Fan Wan, J.; Thongboonkerd, V. Serial Changes in Urinary Proteome Profile of Membranous Nephropathy: Implications for Pathophysiology and Biomarker Discovery. *J. Prot. Res.* **2006**, *5*, 3038-3047.

66. Loong, C.C.; Hsieh, H.G.; Lui, W.Y.; Chen, A.; Lin, C.Y. Evidence for the Early Involvement of Interleukin 17 in Human and Experimental Renal Allograft Rejection. *J. Pathol.* **2002**, *197*, 322-332.

67. Rehman, A.R.; Azzouzi, J.W.F.; Catto, S.; Allen, S.; Cross, S.; Feeley, K.; Meuth, M.; Hamdy, F.C. Proteomic Analysis Of Voided Urine After Prostatic Massage From Patients With Prostate Cancer: A Pilot Study. *J. Urol.* **2004**, *64*, 1238-1243.

68. Trof, R.J.; Di Maggio, F.; Leemreis, J.; Groeneveld, A.B.J. Biomarkers of Acute Renal Injury and Renal Failure. *Shock* **2006**, *26*, 245-253.

69. Zhou, H.; Pisitkun, T.; Aponte, A.; Yuen, P.S.T.; Hoffert, J.D.; Yasuda, H.; Hu, X.; Chawla, L.; Shen, R.F. Exosomal Fetuin-A Identified by Proteomics: A Novel Urinary Biomarker for Detecting Acute Kidney Injury. *Kidney Int.* **2006**, *70*, 1847-1857.

70. Ryuichiro, K.; Kiyohide, S.; Shozo, O.; Atsushi, T.; Seiichi, O. Follow up Study of Renal Function in Children With Reflux Nephropathy After Resolution of Vesicoureteral Reflux. *J. Urol.* **1997**, *157*, 975-979.

71. Furness, P.D.; Maizels, M.; Han, S.W.; Cohn, R.A.; Cheng, E.Y. Elevated Bladder urine Concentration of Transforming Growth Factor-beta 1 Correlates With Upper Urinary Tract Obstruction In Children. *J. Urol.* **1999**, *162*, 3-11.

72. Kavukcu, S.; Soylu, A.; Turkmen, M. The Clinical Value of Urinary N-Acetyl-beta-D-Glucosaminidase Levels in Childhood Age Group. *Acta Medica Okayama* **2002**, *56*, 7-11.

73. Jun, W.; Ryuichiro, K.; Hiroshi, S.; Kiyohide, S.; Sadayoshi, I.; Seiichi, O. Clinical Significance Of Urinary Interleukin-6 In Children With Reflux Nephropathy. *J. Urol.* **2001**, *165*, 210-214.

74. Roy, R.; Wewer, U.M.; Zurakowski, D.; Pories, S.; Moses, M.A. ADAM 12 Cleaves Extracellular Matrix Proteins and Correlates with Cancer Status and Stage. *J. Biol. Chem.* **2004**, *279*, 51323-52330.

75. Raffi, H.S.; Bates, J.M.; Laszik, Z.; Kumar, S. Tamm-Horsfall Protein Acts as a General Host-Defense Factor Against Bacterial Cystitis. *Am. J. Nephrol.* **2005**, *25*, 570-578.

76. Serafini-Cessi, F.; Malagolini, N.; Cavallone, D. Tamm-Horsfall Glycoprotein: Biology and Clinical Relevance. *Am. J. Kidney Dis.* **2003**, *42*, 658-676.

77. Kroot, J.J.C.; Hendriks, J.C.M.; Laarakkers, C.M.M.; Klaver, S.M.; Kemna, E.H. J.M.; Tjalsma, H.; Swinkles, D. W. Preanalytical Imprecision, Between-Subject Variability, and Daily Variations in Serum and Urine Hepcidin: Implications for Clinical Studies. *Anal. Biochem.* **2009**, *389*, 124-129.

78. HUKPP Homepage, http://www.hupo.org/research/urine/

79. Yamamoto, T.; langham, R.G.; Ronco, P.; Knepper, M.A.; Thongboonkered, V. Towards Standard Protocols and Guidelines for Urine Proteomics: A Report on the Human Kidney and urine Proteome Project (HKUPP) Symposium and Workshop. *Proteomics* **2008**, *8*, 2156-2159.

80. Traun, A.Z.; Wells, M.P.; Aivado, M.; Libermann, T.A.; Ramoni, M.F.; Schachter, A.D. SELDI-TOF MS of Quadruplicate Urine and Serum Samples to Evaluate Changes Related to Storage Conditions. *Proteomics* **2006**, *6*, 1676-1680.

81. Zhou, H.; Yuen, P.S.T.; Pisitkun, T.; Gonzales, P.A.; Yasuda, H.; Dear, J.W.; Gross, P., Knepper, M.A.; Star, R.A. Collection, Storage, Preservation, and Normalization of Human Urinary Exosomes for Biomarker Discovery. *Kidney Int.* **2006**, *69*, 1471-1476.

82. Anderson, N.G.; Anderson, N.L.; Tollaksen, S.L.; Hahn, H.; Giere, F.; Edwards, *Anal. Biochem.* **1979**, *95*, 48-61.

83. Smith, G.; Barratt, D.; Rowlinson, R.; Nickson, J.; Tonge, R. Development of a High-Throughput Method for Preparing Human Urine for Two-Dimensional Gel Electrophoresis. *Proteomics* **2005**, *5*, 2315-2318.

84. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacryamide Gels. *Anal. Chem*. **1996**, *68*, 850-858.

85. Tran, J.C.; Doucette, A.A. Rapid and Effective Focusing in a Carrier Ampholyte Solution Isoelectric Focusing System: A Proteome Prefractionation Tool. *J. Proteome Res.* **2008**, *7*, 1761-1766.

86. Marshall, T.; Williams, K.M. Electrophoresis Indicates Proteins Loss on Centrifugation of Urine. *Clin. Chem.* **1986**, *32*, 2105-2106.

87. Fogazzi, G.B.; Garigali, G. The Clinical Art and Science of Urine Microscopy. *Curr. Opin. Nephrol. Hypertens.* **2003,** *12*, 625-632.

88. Spahr, C.S.; Davis, M.T.; McGinly, M.D.; Robinson, J.H.; Bures, E.J.; Beierle, J.; Mort, J.; Courchesene, P.L.; Chen, K.; Whal, R.C.; Tu, W.; Luethy, R.; Patterson, S.D. Towards Defining the Urinary Proteome Using Liquid Chromatography-Tandem Mass Spectrometry 1. Profiling and Unfractionated Tryptic Digest. *Proteomics* **2001**, *1*, 93-107.

89. Pieper, R.; Gatlin, C.L.; McGrath, A.M.; Makusky, A.J.; Mondal, M.; Seonarain, M.; Field, E.; Schatz, C.R.; Estock, M.A.; Ahmed, N.; Anderson, N.G.; Steiner, S. Characterization of the Human Urinary Proteome: A method for High Resolution Display of Urinary Proteins on Two-Dimensional Electrophoresis Gels With a Yield of Nearly 1400 Distinct Protein Spots. *Proteomics* **2004**, *4*, 1159-1174.

90. Wasilewska, A.; Zoch Zwierz, W.; Tomaszewska, B.; Zelazowwska, B. Relationship of Serum Interleukin-7 Concentration and the Coagulation State in Children with Nephrotic Syndrome. *Pediatr. Int.* **2005**, *47*, 424-429.

91. Matheroo-Ball, B.; Ratcliff, L.; Lancashire, L.; Ugurel, S.; Miles, S.K.; Weston, D.J.; Rees, R.; Schadendorf, D.; Ball, G.; Creaser, C.S. Diagnostic Biomarkers Differentiating metastatic melanoma Patients From Healthy Controls Identified by an Integrated MALDI-TOF Mass Spectrometry/Bioinformatic Approach. *Proteomics Clin. Appl.* **2007**, *1*, 605-620.

92. Khositseth, S.; Kanitsap, N.; Warnnissorn, N.; Thonboonkerd, V. IgA Nephrology Associated with Hodgkin's Disease in Children: A Case Report Literature Review and Urinary Proteome Analysis. *Pediatr. Nephrol.* **2007**, *22*, 541-546.

93. Kim, S.J.; Kim, S.; Park, J.; Lee, H.K.; Park, K.S.; Yu, H.G.; Kim, Y. Differential Expression of Vitreous Proteins in Proliferative Diabetic Retinopahy. *Curr. Eye Res.* **2006**, *31*, 231-240.

94. Gangadharan, B.; Antrobus, R.; Dwek, R.A.; Zitzmann, N. Novel Serum Biomarker Candidates for Liver Fibrosis in Hepatitus C Patients. *Clin. Chem.* **2007**, *53*, 1792-1799.

95. Irmak, S.; Tilki, D.; Heukeshoven, J.; Oliveira-Ferrer, L.; Freidrich,; Huland, H.; Ergun, S. Stage-Dependant Increase of Orosomucoid and Zinc-alpha-2-glycoprotein in Urinary Bladder Cancer. *Proteomics* **2005**, *5*, 4296-4304.

96. Scopes, R.K. Protein Purification: Principles and Practice 2nd Edition. *Springer-Verlag*, New York, **1987**, pp. 41-71.

97. Corriere, J.N. Essentials of Urology, *Churchill Livingstone* New York, **1986**, pp. 33.

98. Beavis, R.C.; Chait, B.T. Rapid, Sensitive Analysis of Protein Mixtures by Mass Spectrometry. *Fresenius J. Anal. Chem.,* **1992**, *343*, 25-25.

99. Ikonomou, M.G.; Blades, A.T.; Kebarle, P. Investigations of the Electrospray Interface for Liquid Chromatography Mass Spectrometry. *Anal. Chem.*, **1990**, *62*, 957-967.

100. Rundlett, K.L.; Armstrong, D.W. Mechanism of Signal Suppression by Anionic Surfactants in Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry. *Anal. Chem.*, **1996**, *68*, 3493- 3497.

101. Asanuma, M.; Fukuzawa, S.; Matsuda, T.; Hirota, H. The Effect of Sodium Dodecyl Sulfate and Anion Exchange Silica Gel on Matrix Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Proteins. *Rapid Commun. Mass Spectrom.,* **2009**, *23*, 1647-1653.

102. Breaux, G.A.; Green-Church, K.B.; France, A.; Limbach, P.A. Surfactant-Aided Matrix Assisted Laser Desorption/Ionization Mass Spectrometry of Hydrophobic and Hydrophilic Peptides. *Anal. Chem.*, **2000**, *72*, 1169-1174.

103. Tummala, R. Limbach, P.A. Effect of Sodium Dodecyl Sulfate Micelles on Peptide Mass Fingerprinting by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom*. **2004,** *18*, 2031-2035.

104**.** Chevallet, M.; Santoni, V.; Poinasm A. ; Rouquie, D.; Fuchs, A.; Kieffer, S.; Rossignol, M.; Lunardi, J.; Garin, J.; Rabilloud, T. New Zwitterionic Detergents Improve the Analysis of Membrane Proteins by Two-Dimensional Electrophoresis. *Electrophoresis* **1998**, *19*, 1901-1909.

105. Jaeger, D.A.; Jamrozik, J.; Golich, T.G.; Clennan, M.W.; Mohebalian, J. Preparation and Characterization of Glycerol-Based Cleavable Surfactants and Derived Vesicles. *J. Am. Chem. Soc.* **1989**, *111*, 3001-3006.

106. Anderson, 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.

107. Hudgin, R.L.; Ashwell, G. Studies on role of Glycosyltransferases in Hepatic Binding of Asialoglycoproteins. *J. Biol. Chem*. **1974**, *249*, 7369-7372.

108. Wessel, D.; Flügge, 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.

109. Stearne, P.A.; Vandriel, I.R.; Grego, B.; Simpson, R.J.; Goding, J.W. The Murine Plasma Cell Antigen PC-1 Purification and Partial Amino Acid Sequence. *J. Immunol.* **1985**,*134*, 443-448.

110. Puchades, M.; Westman, A.; Blennow, K.; Davidsson, P. Removal of Sodium Dodecyl Sulfate from Protein Samples Prior to Matrix Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 344-349.

111. Zhang, N.; Chen, R.; Young, N.; Wishart, D.; Winter, P.; Weiner, J.H.; Li, L. Comparison of SDS- and Methanol-Assisted Protein Solubilization and Digestion Methods for Escherichia Coli Membrane Proteome Analysis by 2-D LC-MS/MS *Proteomics* **2007**, *7*, 484-493.

112. Wisniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6*, 359-363.

113. Arand, M.; Friedberg, T.; Oesch, F. Colorimetric Quantitation of Trace Amounts of Sodium Lauryl Sulfate in the Presence of Nucleic-Acids and Proteins. *Anal. Biochem.* **1992**, *207*, 73-75.

114. Tran, J.C.; Doucette, A.A. Gel Eluted Liquid Fractionation Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Determination. *Anal. Chem*. **2008**, *80*, 1568-1573.

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

116. Lee, J.E.; Kellie, J.F.; Tran, J.C., Tipton, J.D.; Catherman, A.D.; Thomas, H.M., Ahlf, D.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.

117. Schirle, M.; Heurtier, M.A.; Kuster, B. Profiling Core Proteomes of Human Cell Lines by One Dimensional PAGE and Liquid Chromatography-Tandem Mass Spectrometry. *Mol. Cell. Proteomics* **2003**, *2*, 1297-1305.

118. de Godoy, L.M.F.; Olsen, J.V.; Cox, J.; Nielsen, M.L.; Hubner, N.C.; Frohlich, F.; Walther, T.C.; Mann, M. Comprehensive Mass Spectrometry Based Proteome Quantification of Haploid Versus Diploid Yeast. *Nature* **2008**, *455*, 1251-U60.

119. Lemaire, M.; Deschamps, S.; Moller, J.V.; Lecafer, J.P.; Rossier, J. Electrospray-Ionization Mass-Spectroscopy on Hydrophobic Peptides Electroeluted from Sodium Dodecyl-Sulfate Polyacrylamide-Gel Electrophoresis Application to the Topology of the Sarcoplasmic-Reticulum Ca^{2+} ATPase. *Anal. Biochem.* **1993**, *214*, 50-57.

120. Shevchenko, A.H; Havlis, J.; Olsen, J.V.; Mann, M. In-gel Digestion for Mass Spectrometric Characterization of Proteins. *Nat. Protoc*. **2006**, *1*, 2856-2860.

121. Wang, H.; Wong, C.H.; Chin, A.; Kennedy, J.; Zhang, Q.; Hanash, S. Quantitative Serum Proteomics Using Dual Stable Isotope Coding and Nano LC-MS/MS. *J. Prot. Res.* **2009**, *8*, 5412-5422.

122. Schmidt, A.; Kellerman, J.; Lottspeich, F. A Novel Strategy for Quantitative Proteomics Using Isotope-Coded Protein Labels. *Proteomics* **2005**, *5*, 4-15.

123. Haenggi, T.; Schaub, M.C.; Fritschy, J.M. Molecular Heterogeneity of the Dystrophin-Associated Protein Complex in the Mouse Kidney Nephron: Differential Alterations in the Absence of Utrophin and Dystrophin. *Cell. Tissue Res.* **2005**, *319*, 299-313.

124. Haenggi, T.; Fritschy, J.M. Role of Dystrophin and Utrophin for Assembly and Function of the Dystrophin Glycoprotein Complex in Non-Muscle Tissue. *Cell. Mol. Life Sciences* **2006**, *63*, 1614-1631.

125. Raats, C.J.I.; van den Born, J.; Bakker, M.A.H.; Oppers-Walgreen, B.; Pisa, B.J.M.; Dijkman, H.B.P.M.; Assmannan, K.J.M.; Berden, J.H.M. Expression of Agrin, Dystroglycan, and Utrophin in Normal Renal Tissue and in Experimental Glomerulopathies. *Am. J. Pathol.* **2000**, *156*, 1749-1765.

126. Loh, N.Y.; Newey, S.E.; Davies, K.E.; Blank, D.J. Assembly of Multiple Dystrobrevin-Containing Complexes in the Kidney. *J. Cell Science* **2000**, *113*, 2715-2724.

127. Pluznick, J.L.; Zou, D.J.; Zhang, Z.; Ya,. Q.; Rodrigez-Gil, D.J.; Eisner, C.; Wells, E.; Greer, C.A.; Wang, T.; Firestein, S.; Schnermann, J.; Caplan, M.J. Functional Expression of the Olfactory Signaling System in the Kidney. *Proc. Natl. Acad. Sci. U S A.* **2008**, *106*, 2059-2064.