

ABSTRACT

Title of dissertation: **PROTEOME ANALYSIS OF
FORMALIN-FIXED AND PARAFFIN-
EMBEDDED TISSUE**

Tong Guo, Doctor of Philosophy, 2008

Dissertation directed by: **Professor Cheng S. Lee
Department of Chemistry and Biochemistry**

Because of the long history of the use of formalin as the standard fixative for tissue processing in histopathology, these archival formalin-fixed and paraffin-embedded (FFPE) tissues present invaluable resources for conducting retrospective disease investigations. However, the high degree of covalently cross-linked proteins in FFPE tissues hinders efficient extraction of proteins from tissue sections and prevents subsequent proteomics efforts from opening the door to a veritable treasure trove of information sequestered in archival tissues.

To this end, a protein extraction methodology has been optimized and demonstrated to achieve effective protein extraction together with combined technological development for enabling comprehensive and comparative proteome studies across archival FFPE tissue collections. An effective discovery-based proteome platform combining capillary isoelectric focusing (CIEF)-based

multidimensional separation system with electrospray ionization-mass spectrometry (ESI-MS) has been developed to enable ultrasensitive analysis of minute protein amounts extracted from targeted cells in tissue specimens in this thesis.

Based on our initial success in analyzing protein profiles within microdissected FFPE tissues, this project further demonstrates the ability to achieve high confidence and comparative proteomic analysis using tissue blocks stored for as many as 28 years. Vacuolar proton translocating ATPase 116 kDa subunit isoform a3, one of the unique proteins expressed in the ASPS, is further validated by immunohistochemistry (IHC). Although IHC is highly sensitive and provides the subcellular resolution, MS-based proteome profiling enables global identification and quantification of thousands of proteins without a prior knowledge of individual proteins being analyzed or the need of validated antibodies.

PROTEOME ANALYSIS OF
FORMALIN-FIXED AND PARAFFIN-EMBEDDED TISSUE

By

Tong Guo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Professor Cheng S. Lee, Chair
Professor Donald Lad DeVoe
Professor Douglas S. English
Professor Catherine C. Fenselau
Professor Amy S. Mullin

© Copyright by

Tong Guo

2008

DEDICATION

I dedicate this dissertation to my loving parents, who have supported me through the hard times and have encouraged me to persevere, no matter what the obstacle.

I can never thank you enough.

TABLE OF CONTENTS

Dedication.....	ii
Table of Contents.....	iii
List of Figures.....	v
List of Tables.....	ix
Chapter One: Capillary Separations Enabling Tissue Proteomics-Based Biomarker Discovery.....	1
1.1 Biomarker Discovery via Tissue Proteomics.....	1
1.2 Two-Dimensional Gel Electrophoresis-Based Tissue Proteomics.....	3
1.3 Tissue Protein Profiling Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry.....	8
1.4 Capillary Separations for Tissue Proteomics.....	9
1.4.1 Multidimensional Liquid Chromatography Systems.....	9
1.4.2 CIEF-Based Multidimensional Separations.....	10
1.5 Introduction to the Thesis.....	24
1.6 Acknowledgement.....	28
Chapter Two: Characterization of the Human Salivary Proteome by Capillary Isoelectric Focusing/Nano-reversed Phase Liquid Chromatography Coupled with ESI-Tandem MS.....	29
2.1 Introduction.....	29
2.2 Experimental Section.....	32
2.2.1 Materials and Reagents.....	32
2.2.2 Sample Collection and Preparation.....	32
2.2.3 Integrated CIEF/Nano-RPLC Multidimensional Separations.....	33
2.2.4 Data Analysis.....	34
2.3 Results and Discussion.....	35
2.3.1 Analysis of Tryptic Peptides Prepared from Whole Saliva.....	36
2.3.2 Detection of Small Proteins and Antimicrobial Proteins in Whole Saliva.....	45
2.3.3 Bacterial Proteins Identified in Whole Saliva.....	45
2.4 Conclusion.....	59
2.5 Acknowledgement.....	60
Chapter Three: Proteome Analysis of Microdissected Formalin-Fixed and Paraffin-Embedded Tissue Specimens.....	61
3.1 Introduction.....	61
3.2 Experimental Section.....	64
3.2.1 Clinical Materials.....	64
3.2.2 Materials and Reagents.....	65

3.2.3	Tissue Microdissection and Protein Sample Preparation.....	65
3.2.4	Integrated CIEF/Nano-RPLC Multidimensional Peptide Separations.....	67
3.2.5	Data Analysis.....	68
3.3	Results and Discussion	69
3.4	Conclusion	84
3.5	Acknowledgement	87
Chapter Four:	Evaluation of Archival Time on Shotgun Proteomics of Formalin-Fixed and Paraffin-Embedded Tissues.....	88
4.1	Introduction.....	88
4.2	Experimental Section.....	90
4.2.1	Materials and Reagents.....	90
4.2.2	Tissue Sample Preparation.....	90
4.2.3	Transient CITP/Capillary Zone Electrophoresis (CZE)- Based Tissue Proteome Analysis.....	91
4.2.4	MS Data Analysis.....	93
4.2.5	Validation of Proteins Retrieved and Identified from FFPE Tissues Using IHC.....	94
4.3	Results and Discussion	95
4.4	Conclusion	113
4.5	Acknowledgement	114
Chapter Five:	Conclusion.....	115
	List of Abbreviations.....	119
	References.....	121

LIST OF FIGURES

Figure 1-1	Optical microscope photograph of stained GBM tissue section	5
Figure 1-2	The 2-D PAGE protein expression pattern of GBM procured by the laser-free microdissection technique.....	7
Figure 1-3	Schematic of on-line integration of CIEF with nano-RPLC as a concentrating and multidimensional protein/peptide separation platform. Solid and dashed lines represent the flow paths for the loading of CIEF fractions and the injection of fractions into a nano-RPLC column, respectively.....	11
Figure 1-4	Base peak chromatograms of a representative CIEF/nano-RPLC multidimensional separation of tryptic peptide digest prepared from microdissection -procured GBM tissue sample. Each number represents the sequence of CIEF fractions further analyzed by nano-RPLC from basic to acidic pHs.....	14
Figure 1-5	The overlap in the proteins identified from three CIEF-nano-RPLC-ESI-MS/MS runs using a single GBM tissue sample.....	16
Figure 1-6	SDS-PAGE of proteins extracted from the FFPE tissue sections (Lanes 1, 2, and 4) and comparable fresh tissue (Lane 3) of renal carcinoma. Lane 1: heat-induced AR at pH 9. Lane 2: heat-induced AR at pH 7. Lane 4: AR without heating.	20
Figure 1-7	Base peak chromatograms of CIEF/nano-RPLC multidimensional separation of tryptic peptides obtained from the FFPE human renal carcinoma tissue using the heat-induced AR technique. Each number represents the sequence of CIEF fractions further analyzed by nano-RPLC from acid to basic pHs.....	21
Figure 1-8	Diagram showing the gene ontology-predicted subcellular localization of proteins found from the FFPE human renal carcinoma tissue using the heat-induced AR technique.....	23
Figure 2-1	Venn diagrams comparing the proteome results obtained from this study (A) versus those achieved by Xie and co-workers ¹⁸ (B).	37

Figure 2-2	Overlaid plots containing the CIEF-UV trace monitored at 280 nm, the number of distinct peptides identified in each of the 29 CIEF fractions, and the distribution of the peptide's mean pI values in each fraction.....	38
Figure 2-3	Plots of the false positive rate and the numbers of distinct peptide and protein identifications versus the E-value obtained from the search of the peak list files against a decoyed SwissProt human database using OMSSA22.....	40
Figure 2-4	Tandem MS spectra of distinct peptide hits contributing to the identification of representative small proteins.....	53
Figure 3-1	Comparison of protein profiles obtained from microdissected fresh frozen (Lane 1) and FFPE (Lane 2) GBM tissue specimens of the same patient using SDS-PAGE.	71
Figure 3-2	Overlaid plots containing the CIEF-UV trace monitored at 280 nm, the number of distinct peptides identified in each of the CIEF fractions, and the distribution of the peptide's mean pI values over the entire CIEF separation.....	73
Figure 3-3	Plots of the false positive rates and the numbers of total peptide, distinct peptide, and distinct protein identifications versus the E-value obtained from the search of the peak list files against a decoyed SwissProt human database using OMSSA.....	74
Figure 3-4	The overlap in the proteins identified from repeated analyses using a single FFPE GBM tissue sample.....	77
Figure 3-5	Distribution of PSLT-predicted subcellular localization of proteins identified from the microdissection-procured FFPE GBM tissue specimen.....	78
Figure 3-6	Peptide coverage of representative transmembrane proteins such as (A) tenasin and (B) basigin, and tandem MS spectra of unique peptides leading to their identifications.....	79

Figure 3-7	The overlap in the proteins identified from microdissected fresh frozen (the soluble and pellet fractions) and FFPE GBM tissue specimens of the same patient using combined CIEF/nano-RPLC separations coupled with nano-ESI-LTQ-MS/MS.....	83
Figure 3-8	The overlap in the proteins identified from microdissected fresh frozen (the soluble and pellet fractions) and FFPE GBM tissue specimens predicted to contain at least one or more transmembrane domains using TMHMM (www.cbs.dtu.dk/services/TMHMM-2.0/).....	85
Figure 4-1	H&E staining of (A) uterine leiomyoma and (B) vaginal ASPS.....	97
Figure 4-2	The Venn diagram comparing proteins identified from the sarcoma (small cycle) and leiomyomas (large cycle).....	100
Figure 4-3	(A) Tandem MS spectrum of unique peptide LGALQQLQQSQELQEVLGETER leading to the identification of VPP3. IHC staining of VPP3 on (B) ASPS and (C) leiomyoma tissue sections.....	101
Figure 4-4	(A) Pearson correlation plot between 2002 and 1990 leiomyoma tumor groups. (B) Pearson correlation plot between 1980 sarcoma and 1990 leiomyoma tumor groups.....	104
Figure 4-5	The hierarchical cluster analysis of all ten mesenchymal tumors.....	106
Figure 4-6	Distribution of protein expression of three leiomyoma markers, including actin, desmin, and progesterone receptor, over the archival years from 1990 to 2002.....	107
Figure 4-7	Representative IHC staining of three leiomyoma markers, including (A) desmin, (B) actin, and (C) progesterone receptor, on the same tissue blocks employed for MS-based proteome profiling.....	108

Figure 4-8	The k-means clustering of all leiomyoma proteins over the archival years from 1990 to 2002 on the x-axis. Proteins in each of ten clusters are varied in abundance as determined by the number of spectral counts on the y-axis.....	110
Figure 4-9	Performing ANOVA among proteomic results obtained from leiomyomas collected in 1990, 1997, and 2002. (A) Two-way comparison between 1990 and 2002 leiomyoma tumor groups. (B) Two-way comparison with each company containing a roughly equal mix of cases from each archival date.....	112

LIST OF TABLES

Table 2-1	Peptides and Proteins Identified at a False Positive Rate of 1%.....	42
Table 2-2	Proteins Identified from Fully Tryptic Peptides at a False positive Rate of	43
Table 2-3	List of Small Proteins (< 20 kDa) Identified in Human Saliva Studies.....	46
Table 2-4	Bacterial Species Identified by Two or More Fully Tryptic Peptide Sequences (False Positive Rate of 0.1%) Which Are Unique to the NCBI Non-Redundant Protein Database.....	58
Table 4-1	Summary of Proteomic Results Obtained from Mesenchymal Tumor Tissues	99

CHAPTER ONE

CAPILLARY SEPARATIONS ENABLING TISSUE PROTEOMICS-BASED BIOMARKER DISCOVERY

Reproduced with permission from Tong Guo, Cheng S. Lee, Weijie Wang, Don L. DeVoe, and Brian M. Balgley, *Electrophoresis* (2006), 27, 3523-3532.

Copyright 2006 Wiley-VCH.

1.1 BIOMARKER DISCOVERY VIA TISSUE PROTEOMICS

A biological marker, biomarker, is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacological responses to a therapeutic intervention”[1]. Biomarkers can be categorized into three major classes including diagnostic, prognostic, and predictive markers. Even though routine cancer diagnosis is largely based on microscopical assessments of morphologic alternations of cells and tissues[2], diagnostic markers can be used to aid histopathological tumor classification which is critical for optimal treatment choices. Prognostic markers provide information regarding the malignant potential of tumors. Predictive markers are employed to choose between different alternative treatments modalities. For example, validated predictive factors for breast cancer patients include the assessment of estrogen-receptor, progesterone-receptor, and human epidermal growth factor

receptor 2[3]. Assay of HER-2/neu is now mandatory in deciding which patients with metastatic breast cancer should receive treatment with the monoclonal antibody, Herceptin[4].

Predictions of cancer behavior and likely drug response have been confounded by the great complexity of the human genome and, very often, the cellular heterogeneity of tumors. While analyses of DNA and RNA expression profiles through techniques including cDNA microarrays, comparative genomic hybridization, loss of heterozygosity, and single nucleotide polymorphism (SNP) analysis are important in identifying genetic abnormalities and uncovering the molecular dysfunctions existing in tumor cells, the presence of SNPs, changes in DNA copy numbers, or altered RNA levels may have little or no effect on the events actually happening at the protein level. For example, although 62 genes or more are possibly associated with the onset, progression, and/or severity of breast cancer[5], the specific roles played by the majority of these genes are yet to be clearly elucidated at the protein level, and only a small number have been clinically validated or associated with clinical phenotype. Besides the lack of understanding of the information encoded in the DNA sequence to identify the complete set of open-reading frames, protein diversity is further complicated and multiplied through alternative splicing and chemical/enzymatic modifications.

Thus, there is an urgent need for the development of technologies that allow the monitoring of protein expression and processing in tumor tissues resulting from development, physiology, and disease state. The ability to monitor the presence or absence of particular proteins, an increase or decrease in protein expression, changes in protein

post-translational modifications (PTMs), or combination of these variations in protein profiles will provide a more accurate snapshot of the molecular basis of a cancer lesion due to a systematic understanding of the multiple components of complex cellular networks driving differentiation and proliferation. For example, the transformation of a normal cell into a cancer cell requires multiple genetic modifications or changes, such as alterations in the cell cycle specifically modulating protein modifications or mechanisms of DNA repair. The identification of gene products either distinctive to cancers, or playing a possible role related to their development, will help define the sequence of molecular events that lead to cancer. As the efforts in tissue proteomics-based biomarker discovery progress, a profound impact will be made directly on disease related research ranging from the ability to enhance early detection and diagnosis of disease to the identification of biologically relevant targets for drug development and screening.

1.2 TWO-DIMENSIONAL GEL ELECTROPHORESIS-BASED TISSUE PROTEOMICS

Determinations of the levels or activities of specific analytes in tissue biopsies, used as biomarkers to detect, diagnose or evaluate prognosis of a patient, are typically performed biochemically or by IHC[6, 7]. If the analyte is measured biochemically, a tissue specimen consisting of a heterogeneous cell population is homogenized and the final concentration of the analyte from the diseased cells is “diluted” by the contribution of other proteins released from non-diseased cells (e.g. normal epithelium and stromal cells). Therefore, an under-estimate of the analyte concentration is likely to be observed, complicating the appropriate cut-off value between disease and normal states. While some of these types of tumor markers in tissue biopsies have served the clinical management of

cancer patients well, e.g. estrogen receptors in the selection of tamoxifen-responsive breast cancer[7], many questions of analyte expression in cancer remain. As reviewed by Cole et al.[8], collection and processing of human tissue biopsies primarily serve a clinical purpose (diagnosis, staging) with little emphasis on sampling and preservation for sophisticated genomic and proteomic analyses. Complicating these issues is the obvious problem of cell heterogeneity in the tissue section, which often results in misleading or confusing molecular findings[8, 9].

Still, biologically relevant proteomics data can only be generated if the samples investigated consist of homogeneous cell populations, in which no unwanted cells of different types and/or development stages obscure the results. One of the main problems with the analysis of tissue proteomics is the heterogeneous nature of the sample. As shown in (Fig. 1-1), many different cell types are typically present in tissue biopsies, and in the case of diseased tissue small numbers of abnormal cells may lie within or adjacent to unaffected areas. Thus, several tissue microdissection technologies, including laser capture microdissection (LCM) [10, 11], laser microdissection [12], and laser-free microdissection[13, 14], have been developed to provide a rapid and straightforward method for procuring homogeneous subpopulations of cells or structures for biochemical and molecular biological analyses. It should be noted that the use of laser for tissue preparation in LCM [10, 11] and laser microdissection [12] may subject samples to potential processing-induced changes at two different stages: first during the staining of tissue sections to enable selection of the relevant cell types and secondly during the dissection itself.

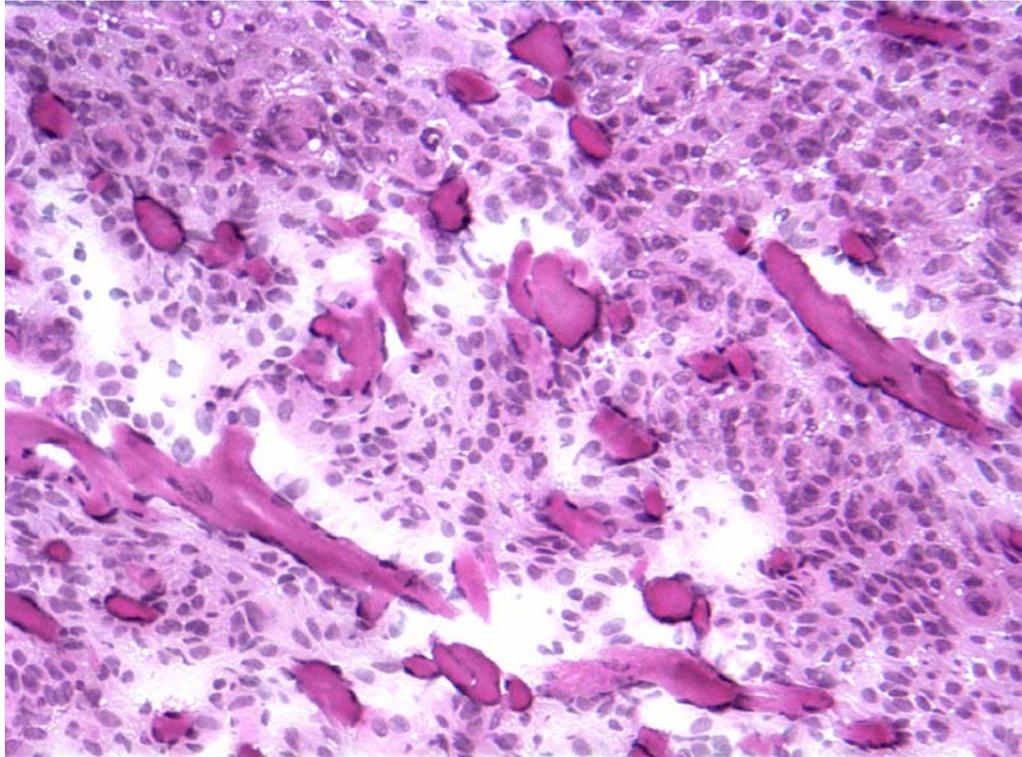


Figure 1-1 Optical microscope photograph of stained GBM tissue section[15]

These changes further impact the level of protein recovery as the result of sample alterations and the quality of subsequent proteome studies.

Current protein analysis technologies are still largely based upon two-dimensional polyacrylamide gel electrophoresis (2-D PAGE), which is central to much of what is now described as “proteomics”. This is due to its ability to provide detailed views of thousands of proteins expressed by an organism or cell type. However, in the absence of protein amplification techniques, proteomic analysis of microdissection-procured specimens is severely constrained by sample amounts ranging from 10^3 - 10^5 cells, corresponding to a total protein content of 0.1-10 μg [11]. While limited 2-D PAGE analyses of microdissection-derived tissue samples have been attempted[16-22], these studies involve significant manual effort and time to extract sufficient amounts of protein for obtaining gel patterns of good visual quality (Fig. 1-2), while providing little proteome information beyond a relatively small number of high abundance protein identifications. In addition, the 2-D PAGE-MS approach itself suffers from low throughput and poor reproducibility, and remains lacking in the proteome coverage, dynamic range, and sensitivity.



Figure 1-2 The 2-D PAGE protein expression pattern of GBM procured by the laser-free microdissection technique[13].

1.3 TISSUE PROTEIN PROFILING USING MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY

Cells isolated and captured by tissue microdissection techniques have been directly analyzed using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)[23-25]. The m/z signals or peaks obtained from MALDI-MS are correlated with protein distribution within a specific region of the tissue sample and can be used to construct ion density maps or specific molecular images for virtually every signal detected in the analysis. Similarly, surface-enhanced laser desorption/ionization-mass spectrometry (SELDI-MS) has been reported as a relatively simple, rapid, and sensitive protein biomarker analysis tool with potential clinical utility[26-31]. Although SELDI-MS spectral peaks do not provide protein identifications, the resulting fingerprint patterns have been explored for diagnostic applications including the early detection of several types of cancers. However, reproducibility of sample preparation and mass spectra among different laboratories has been problematic for both methodological and biological reasons, and little physical or biological meaning can be attributed to the spectral peaks which are clustered only in the low m/z region[29].

While SELDI has been shown to produce observable MS profiles with as few as 25-50 cells, the transition of protein pattern produced by SEIDI to protein identity is generally quite difficult. The efforts typically involve tracking the specific peak of interest through a number of chromatography and gel purification steps to ensure that the peptides measured in the proteolytic digest are mainly derived from the protein of interest[32]. The inherent large-scale and the difficulty of this protein identification process not only account for both the small number of proteins identified and the tendency toward the

identification of highly abundant proteins, but also call into question the practicality of overall SELDI approach when working with limited clinical samples in general, and microdissected tissue specimens in particular.

1.4 CAPILLARY SEPARATIONS FOR TISSUE PROTEOMICS

1.4.1 Multidimensional Liquid Chromatography Systems

Yates and co-workers[33]demonstrated the use of shotgun identification of protein mixtures by proteolytic digestion followed with multidimensional liquid chromatography and tandem MS to separate and fragment the resulting peptides. Total peak capacity improvements in multidimensional chromatography platforms have increased the number of detected peptides and proteins identified due to better use of the MS dynamic range and reduced discrimination during ionization. To increase the proteome coverage, particularly for the identification of low abundance proteins, these peptide-based proteome technologies often require large quantities of enzymatically/chemically cleaved peptides, ranging from a few mgs[34, 35]to several hundred μ gs[33, 36, 37], and are generally incompatible with protein levels extracted from microdissection-procured tissue samples.

Thus, the reported tissue proteomic studies employing multidimensional liquid chromatography separations [38-41] are mainly based on the analysis of entire tissue sections instead of subpopulations of microdissection-derived tumor cells. The minimal quantity of available sample from targeted tumor cells has restricted analysis to the use of only a single chromatography separation prior to tandem MS analysis in recent studies [42-45] and limited the ability for mining deeper into the tissue proteome. Since the sizes of human tissue biopsies are becoming significantly smaller due to early detection and

diagnosis, a more effective discovery-based proteome technology is critically needed to enable sensitive studies of protein profiles within tissue specimens procured by microdissection techniques[10-14].

1.4.2 CIEF-Based Multidimensional Separations

An effective discovery-based proteome platform[15, 46, 47], which combines CIEF with nano-reversed phase liquid chromatography (nano-RPLC) (Fig. 1-3), has recently been developed to enable ultrasensitive analysis of minute protein amounts extracted from small cell populations and limited tissue samples. As demonstrated in our previous studies[15, 46, 47], the key to establishing ultrasensitive proteome analysis is attributed to high analyte concentrations in small peak volumes as the result of electrokinetic focusing/stacking and high resolving multidimensional separation techniques, thereby enhancing the dynamic range and detection sensitivity of MS measurements. Furthermore, a key feature of CIEF-based proteome technology is the elimination of analyte loss and dilution in an integrated platform while achieving comprehensive and ultrasensitive quantification of protein expression profiles within microdissection-procured tissue specimens.

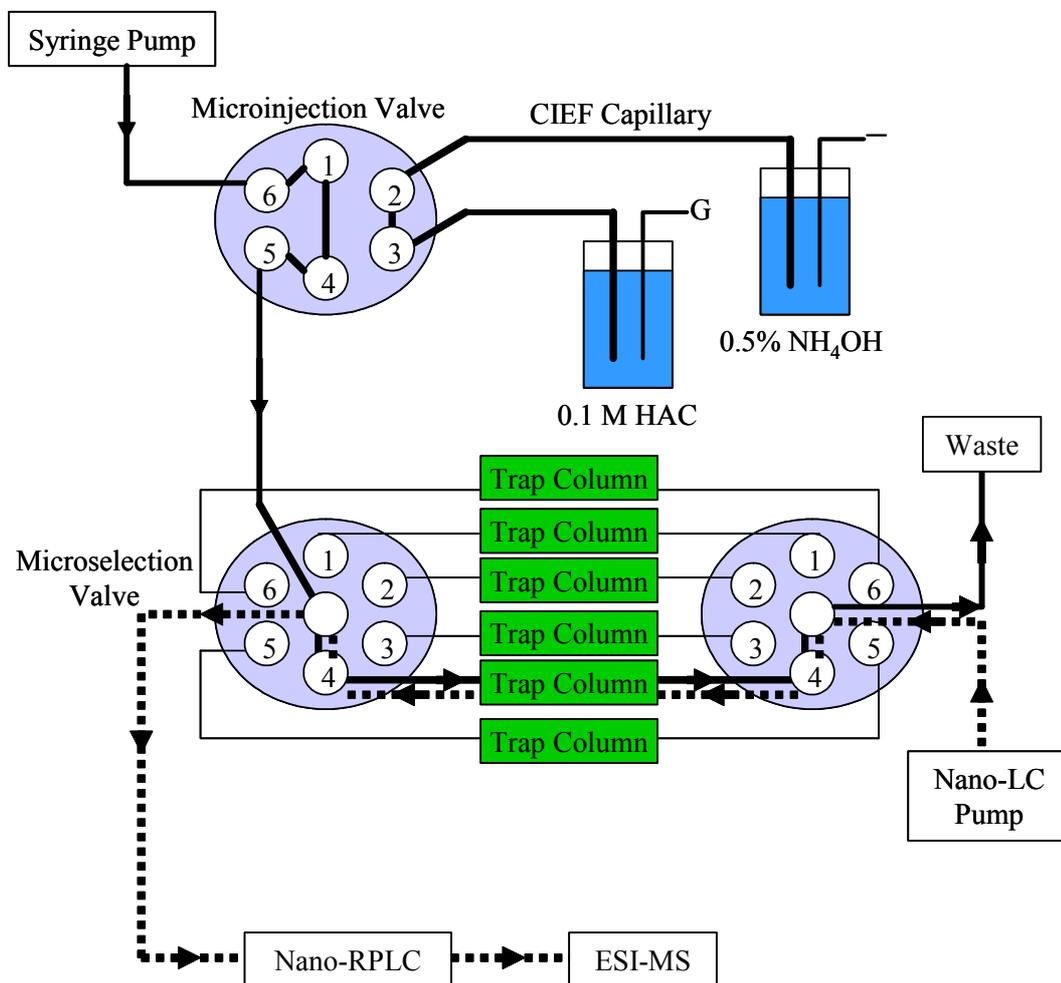


Figure 1-3 Schematic of on-line integration of CIEF with nano-RPLC as a concentrating and multidimensional protein/peptide separation platform. Solid and dashed lines represent the flow paths for the loading of CIEF fractions and the injection of fractions into a nano-RPLC column, respectively[48]

Compared with the coupling of strong cation exchange chromatography with RPLC in a multidimensional liquid chromatography system[33-37], the percentage of identified peptides present in more than one CIEF fraction is significantly less than that obtained from ion exchange chromatography fractions. A high degree of peptide overlap in the first dimension unnecessarily burdens the subsequent separation and greatly reduces the overall peak capacity in a multidimensional separation system. The presence of high abundance peptides in multiple fractions negatively impacts the selection of low abundance peptides for tandem MS identification. Additionally, the poor resolution of low abundance peptides adversely affects their final concentrations in the eluting peaks prior to MS detection.

Besides the ultrahigh resolving power of CIEF-based multidimensional separations, initial proteome studies [46, 47] have supported the ability to perform comprehensive analysis of yeast cell lysates while requiring the amount of yeast peptides or proteins which are two to three orders of magnitude less than those utilized in a multidimensional liquid chromatography system[33-37, 49-51]. Such an increase in proteome sensitivity resulting from isoelectric focusing concentration and separation of peptides and proteins is further supported by the study of Cargile and co-workers using immobilized pH gradient gels[52]. By comparing with immobilized pH gradient gels[52, 53], the significant advantages of CIEF include the ability to use higher electric fields resulting in rapid and ultrahigh resolution separations, improved reproducibility in fractions resulting from the use of gel free separation medium, and the amenability to automation and seamless coupling with nano-RPLC in an integrated platform while

avoiding potential sample loss, analyte dilution, and any post-gel peptide extraction/concentration procedures.

The solubility of concentrated proteins and peptides at their pIs may become a problem for the CIEF separation. It should be noted that a whole family of non-detergent solubilizers, including glycols, sugars, and common zwitterions such as taurine, has been reported to greatly improve protein solubility in the proximity of their pIs[54-56], particularly for the analysis of membrane proteins using 2-D PAGE. These additives may be employed for completely preventing or largely alleviating the solubility issue, which may be encountered during the CIEF separation.

1.4.2.1 Fresh frozen tissue biopsies

Combined CIEF/nano-RPLC separations have been employed for the analysis of protein profiles within small and selected tumor cell populations obtained from the microdissection of a glioblastoma multiforme (GBM) tissue specimen[15]. By using an electrospray ionization (ESI)-quadrupole time-of-flight (qTOF) mass spectrometer for the analysis of peptides eluted from nano-RPLC (Fig. 1-4), a total of 6,866 fully tryptic peptides have been detected, leading to the identification of 1,820 distinct proteins with each corresponding to a unique human International Protein Index database entry. These identifications are generated from a total of 18,843 tandem mass spectra from 3 runs of a single GBM tissue sample, with each run consuming only 10 μ g of total protein. These identifications are based on high mass accuracy and high confidence hits to fully tryptic peptides.

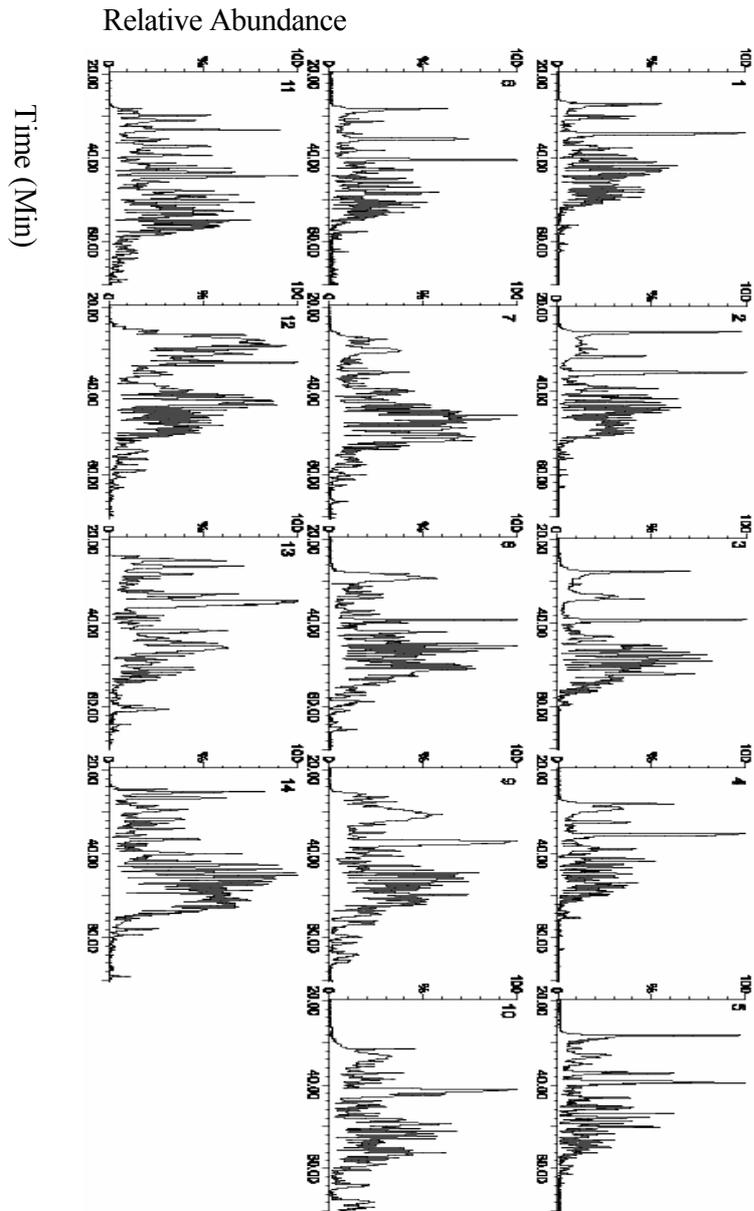


Figure 1-4 Base peak chromatograms of a representative CIEF/nano-RPLC multidimensional separation of tryptic peptide digest prepared from microdissection-procured GBM tissue sample. Each number represents the sequence of CIEF fractions further analyzed by nano-RPLC from basic to acidic pHs[15].

Furthermore, a variety of protein classes, including structural proteins, metabolic enzymes, receptors, and proteins involved in transcription, translation, and post-translational modification of gene products, are identified from the microdissection-procured GBM tissue samples.

The intrinsic high resolution of CIEF allows the number of fractions sampled in the first separation dimension to be significantly increased for enhancing the overall peak capacity and dynamic range critically needed for performing comprehensive proteome studies. More CIEF fractions can be obtained by simply increasing the number of microselection valves and trap columns (Fig. 1-3), or by employing microselection valves with a higher number of ports. For the proteome analysis of a single GBM tissue sample, the number of distinct proteins identified from a single experiment of 28 CIEF fractions (Fig. 1-5) is comparable or slightly higher than the cumulative data set achieved using two runs of 14 CIEF fractions which consume twice as much sample. For the proteome analysis of limited tissue samples, the ability to greatly increase the overall peak capacity for mining deeper into the proteome can therefore be realized by simply increasing the number of CIEF fractions without an accompanying increase in sample consumption. By using computer controlled microinjection and microselection valves, the CIEF-based multidimensional separations are further amenable to automated and high throughput analyses.

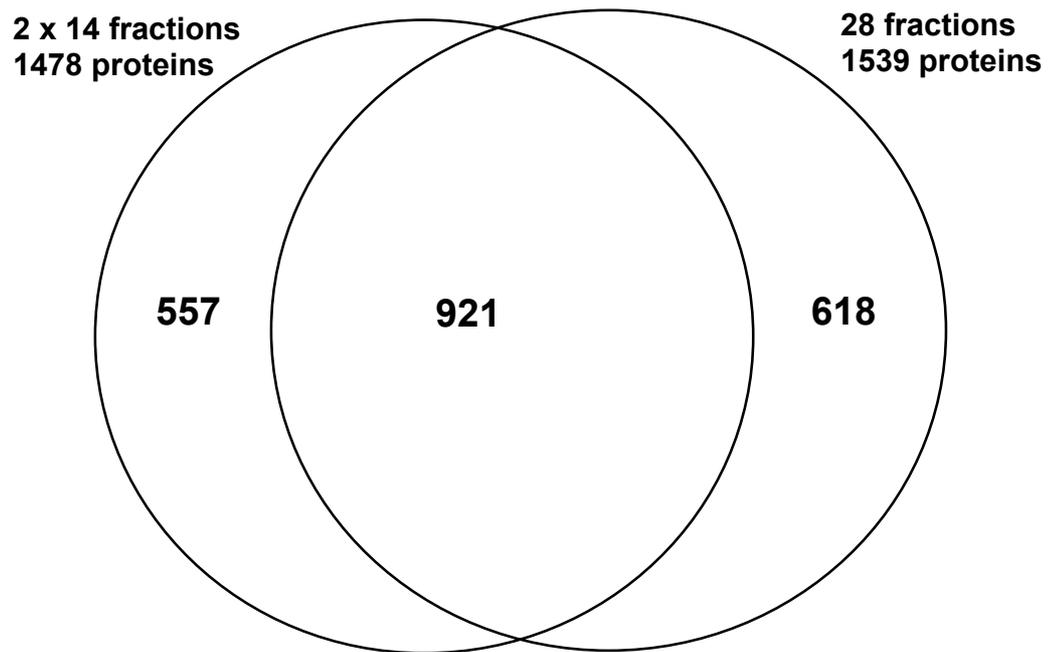


Figure 1-5 The overlap in the proteins identified from three CIEF-nano-RPLC-ESI-MS/MS runs using a single GBM tissue sample [15]

For the identification of peptides eluted from separations, tandem MS instruments such as a qTOF mass spectrometer obtain peptide MS/MS spectra in a data-dependent way for the determination of peptide sequences. However, the data-dependent scanning cannot acquire spectra fast enough to identify all the peptides as they enter the MS instrument, so many peptides and subsequent protein identifications are missed. The recently introduced rapid-scanning linear ion-trap instruments can acquire tandem mass spectra up to fifteen times faster than that using a qTOF mass spectrometer. By using a linear ion-trap mass spectrometer such as LTQ, each single proteome analysis of 10,000 enriched tumor cells yields over 10,000 distinct and high confidence peptide hits with a false positive rate of less than 1%[57]. These peptide hits are further filtered by pI constraints and lead to the identification of more than 2,500 nonreductant proteins with greater than 80% protein overlapping among repeated runs of the same GBM tissue specimens.

For peptide and protein identifications, raw LTQ data are first converted to peak list files. The Open Mass Spectrometry Search Algorithm (OMSSA) developed at the National Center for Biotechnology Information [58] is used to search the peak list files against a decoyed SwissProt human database. This database is constructed by reversing all 12,484 real sequences and appending them to the end of the sequence library. Searches are performed by setting up the following parameters: fully tryptic, 1.5 Da precursor ion mass tolerance, 0.8 Da fragment ion mass tolerance, 1 missed cleavage, alkylated Cys as a fixed modification and variable modifications of acetylated N-terminus and Lys, phosphorylated Ser, Thr, and Tyr, and oxidated Met. False positive rates are then determined using the method of Elias and co-workers[59].

Briefly, false positive rates are calculated by multiplying the number of false positive identifications (hits to the reversed sequences scoring below a given threshold) by 2 and dividing by the number of total identifications. Peptides identified below threshold, and also occurring as matches to the forward sequences, are not counted as false positives or true identifications. A curve is then generated by plotting E-value versus false positive rate [60] and an E-value threshold corresponding to a 1% false positive rate is used as a cutoff[57].

Besides comparing normal versus GBM tissues, we have further investigated the differences in protein expression profiles within short-term (< 9 months) and long-term (> 24 months) GBM survivors using a set of fresh frozen biopsies (n = 6 for both short- and long-term survivors) in triplicate for each specimen. These quantitative and comparative proteome studies using non-labeling approaches [61-65] have demonstrated our abilities to not only achieve high correlation within similar phenotypes, but also mine biomarkers across distinct phenotypes. A total of 25 differentially expressed proteins, as indicated by a t-test score of < 0.001, can potentially serve as prognostic biomarkers and are being validated using additional tissue specimens through multiple reaction monitoring, IHC, and Western blots.

1.4.2.2 FFPE tissues

Because of the long history of the use of formalin as the standard fixative for tissue processing in histopathology, there are a large number of archival FFPE tissue banks worldwide. However, the high degree of covalently cross-linked proteins in FFPE tissues hinders efficient extraction of proteins from tissue sections and prevents subsequent immunohistochemical. In this context, IHC as applied to the demonstration of

antigens (primarily proteins) in tissue sections represents the first form of proteomics, wherein the ability to identify proteins is greatly enhanced by a simple and effective antigen retrieval (AR) technology. In the AR method, boiling the FFPE tissue sections in water or buffer solution dramatically reduces the detection thresholds of IHC staining (increasing sensitivity) for a wide range of antibodies [66-68]. This AR technique has revolutionized the practice of diagnostic IHC to the extent that the relevant published literatures of IHC are divided into “pre-AR” and “post-AR” eras[69]. In addition to the AR technique, a Liquid Tissue™ kit has recently been developed and commercialized by Expression Pathology for protein extraction from FFPE tissues[45, 70].

Besides fresh frozen tissue-based proteome studies, combined CIEF/nano-RPLC separations equipped with ESI-linear ion-trap MS have recently been employed to examine protein profiles extracted from an archival FFPE tissue of human renal carcinoma using the AR technique[71]. Sections of FFPE tissue are boiled in a retrieval solution of 20 mM Tris containing 2% sodium dodecyl sulfate (SDS) for 20 min, followed by incubation at 60 °C for 2 hr. As a control experiment, the same FFPE tissue sample is also processed for AR without heating. Furthermore, fresh tissue taken from the same case of renal carcinoma is processed for protein extraction. Initial evaluation of the quality of proteins extracted from FFPE and fresh tissues is performed using SDS-PAGE. Clearly, protein patterns extracted from FFPE sections using the heated AR approach are comparable with that obtained from fresh tissue (Fig. 1-6).

By using combined CIEF/nano-RPLC separations, tryptic peptides obtained from proteolytic digestion of protein extracts using the heated AR approach at pH 7 are systematically resolved by their differences in pI and hydrophobicity (Fig. 1-7).

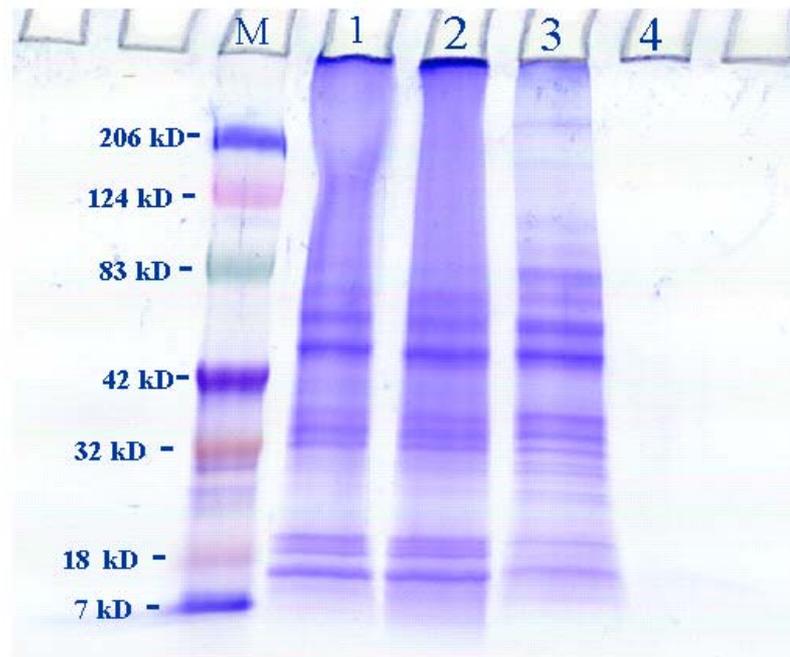


Figure 1-6 SDS-PAGE of proteins extracted from the FFPE tissue sections (Lanes 1, 2, and 4) and comparable fresh tissue (Lane 3) of renal carcinoma. Lane 1: heat-induced AR at pH 9. Lane 2: heat-induced AR at pH 7. Lane 4: AR without heating[71].

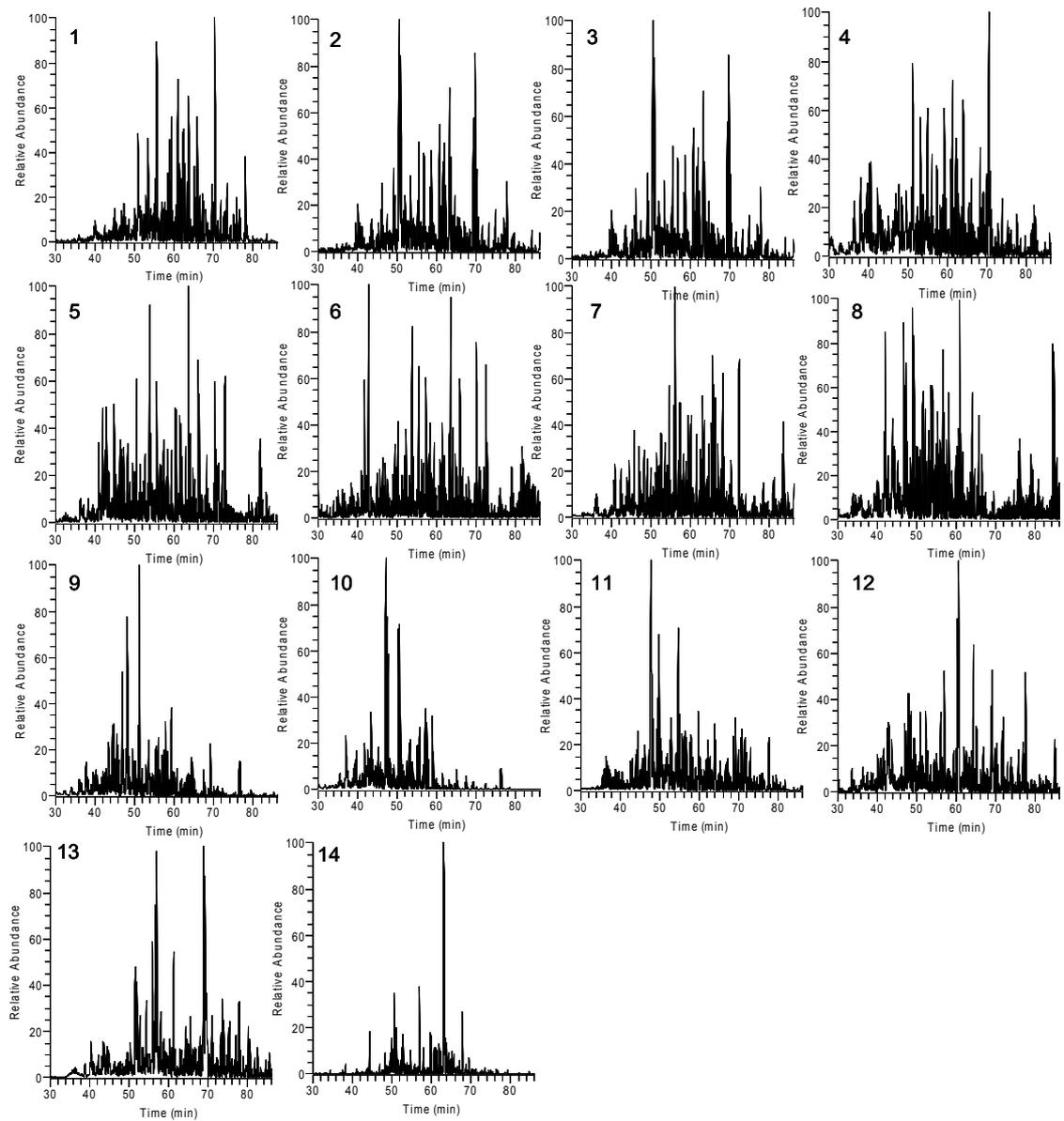


Figure 1-7 Base peak chromatograms of CIEF/nano-RPLC multidimensional separation of tryptic peptides obtained from the FFPE human renal carcinoma tissue using the heat-induced AR technique. Each number represents the sequence of CIEF fractions further analyzed by nano-RPLC from acid to basic pHs[71].

A total of 12,855 fully tryptic peptides are identified from the FFPE tissue section treated with the heat-induced AR approach, leading to the identification of 3,263 distinct proteins. By comparing with the reported FFPE tissue-based proteome studies [45, 72, 73], our results present the most comprehensive analysis of human tissue proteomics including the identification of low abundance membrane proteins (Fig. 1-8). Again, the application of only single-dimension chromatography separation in these recent studies [45, 72, 73] has significantly limited the dynamic range and detection sensitivity of MS measurements, and greatly impacted their ability to mine deeper into the tissue proteome.

Proteome results of FFPE tissue are further compared with the analysis performed on a fresh frozen tissue procured from the same case of renal carcinoma. More than 70% of proteins are identified in both tissue sections and demonstrate excellent correlation. Despite the potentially negative impact imparted from the FFPE process on protein extraction, a similar and significant amount of proteomic information can be retrieved from FFPE samples as compared to fresh frozen tissue specimens.

On the other hand, a significant discrepancy in the quality of proteome results is found among protein extracts obtained using the heat-induced and room temperature AR approaches. As also demonstrated by the IHC studies [66-68], a critical element in the mechanism of the AR technique may be based on heat-induced modification of the three-dimensional structure of “formalinized” protein, restoring the condition of a formalin-modified protein structure back towards its original structure.

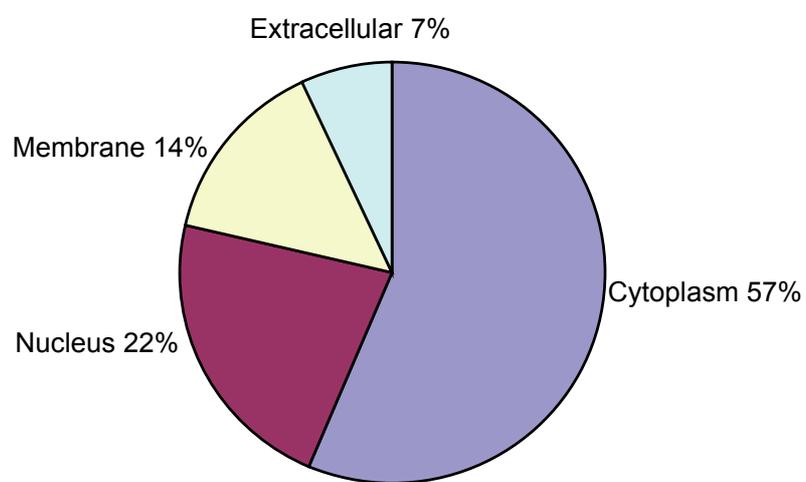


Figure 1-8 Diagram showing the gene ontology-predicted subcellular localization of proteins found from the FFPE human renal carcinoma tissue using the heat-induced AR technique[71].

In other words, the mechanism of AR technique appears to involve a re-naturation of the structure of fixed proteins through a series of conformational changes, including the possible breaking (hydrolysis) of formalin-induced cross-linkages, the entire process being driven by thermal energy from the heat source[67]. Still, high-temperature heating treatment may also induce negative results such as additional protein modifications and therefore requires further studies and comparisons with low-temperature heating treatments and combined retrieval protocols utilizing both heat and enzyme digestion [74-76].

1.5 INTRODUCTION TO THE THESIS

In human disease research, where knowledge of disease outcome is critical for the evaluation of the significance of phenotypic or genotypic profiles, as well as response to therapy and outcome, it may take five, ten or more years to gain a relatively complete picture of the pathophysiology of a disease. The ability to analyze well-characterized archival cases is highly desirable. In addition, because the capacity to store large numbers of catalogued samples under optimal conditions is limited by cost, space and personnel limitations, among others, the development of technologies to analyze traditional pathological specimens, such as FFPE is an important priority. The experiments described in this thesis are designed to develop new methodologies to identify new protein biomarkers associated with diseased tissue.

Still, biologically relevant proteomics data can only be generated if the tissue samples investigated consist of homogeneous cell populations, in which no unwanted cells of different types and/or development stages obscure the results. Thus, laser capture microdissection technology has been developed to provide a rapid and straightforward

method for procuring homogenous subpopulations of cells for biochemical and molecular biological analyses. Laser microdissection involves the use of a cap coated with a thermolabile ethylene vinyl acetate film that is placed in contact with a stained tissue section visualized by an inverted microscope. A focused laser beam is then employed to produce localized melting of the film over selected cells such that the underlying tissue becomes fused to the cap and is selectively removed when the cap is lifted.

However, in the absence of protein amplification techniques, proteomic analysis of microdissection-procured specimens is severely constrained by sample amounts ranging from 10^3 - 10^5 cells, corresponding to a total protein content of 0.1-10 μg [11]. Current proteomic techniques, including two-dimension polyacrylamide gel electrophoresis and shotgun-based multidimensional liquid chromatography separations, require substantially larger cellular samples which are generally incompatible with protein extract levels obtained from microdissected samples.

Development of the capability to enable large-scale proteome studies, analogous to comprehensive gene expression analysis, will clearly have far-reaching impacts on protein biomarker investigations of human diseases such as cancer through interrogation of the archived FFPE tissue collections. This project therefore aims to develop and optimize on-line combination of Capillary Electrophoresis based separation platforms with nano-RPLC in an integrated platform for the comprehensive analysis of archival FFPE tissue specimens.

Chapter two demonstrates the unique proteome capabilities of the CIEF-based multidimensional separation platform coupled with ESI-tandem MS using human salivary secretory components as a model proteome system. Compared with strong cation

exchange chromatography utilized as the first separation dimension in multidimensional protein identification technology[33, 77], CIEF not only contributed to high resolving power together with the enrichment of low abundance proteins/peptides, but also provided seamless coupling with nano-RPLC while avoiding any sample loss and analyte dilution in an integrated platform.

Chapter three discusses the CIEF-based proteome platform for the analysis of targeted tumor cells procured from microdissected fresh frozen and FFPE tissue specimens. Targeted proteomics research, based on the enrichment of disease-relevant proteins from isolated cell populations selected from high quality tissue specimens, offers great potential for the identification of diagnostic, prognostic, and predictive biological markers for use in the clinical setting and during preclinical testing and clinical trials, as well as for the discovery and validation of new protein drug targets. FFPE tissue collections, with attached clinical and outcome information, are invaluable resources for conducting retrospective protein biomarker investigations and performing translational studies of cancer and other diseases.

By combining the antigen retrieval (AR) technique with the CIEF/Nano-RPLC separations, a total of 14,478 distinct peptides were identified from a single FFPE GBM tissue specimen. The result not only led to the identification of 2,733 non-redundant SwissProt protein entries, but also presented the largest catalog of proteins from a single microdissected FFPE tissue specimen reported to date. Protein expressions measured from FFPE tissue were also compared with those detected from fresh frozen specimen of the same, matched patient. 83% of identified FFPE tissue proteins overlapped with those obtained from the pellet fraction of fresh frozen tissue of the same patient. This large

degree of protein overlapping was attributed to the application of detergent-based protein extraction in both the cell pellet preparation protocol and the AR technique. This represents an important step forward in bringing proteomic profiling of archival FFPE tissues into reality.

Chapter four evaluated the effects of length of storage period on archival tissue proteome analysis across ten archived uterine mesenchymal tumor tissue blocks, including nine uterine leiomyomas dating from 1990 to 2002 and a single case of ASPS from 1980 by using the capillary isotachopheresis-based proteome technology. Statistical measures, including the Pearson correlation coefficient and the hierarchical cluster analysis, were employed to evaluate the possibility of an archival effect on individual proteins or groups of proteins within nine leiomyomas. Low abundance proteins may be more susceptible to the long term storage as these proteins are more difficult to be retrieved and extracted as the tissue block ages in paraffin.

Despite using tissue blocks stored for as many as 28 years, high confidence and comparative proteome analysis between the leiomyomas and the sarcoma was achieved. Though sharing over 1,800 common proteins in a core set, a total of 80 proteins unique to the sarcoma were identified distinguishing the ASPS from the leiomyomas. Vacuolar proton translocating ATPase 116 kDa subunit isoform a3, one of the unique proteins expressed in the ASPS, was further validated by IHC.

Following the first medical school course in histology at Edinburgh in 1842, the diagnosis of cancer has come to be based upon the microscopic appearances of tissues (histopathology = morphologic phenotype). However, the last several decades have seen an accumulation of data describing the characteristics of cancer cells at a

protein and nucleic acid level (molecular phenotype). Thus, advanced proteome technologies employed and demonstrated in this study not only allow the rigorous evaluation of the quality and the reproducibility of proteins extracted from FFPE tissues for further optimization of AR methodology, but also provide significant opportunities in the pursuit of biomarker discovery using archived FFPE tissue collections. Additionally, the combination of histological criteria with advanced molecular analysis techniques will permit pathologists and biomedical investigators to incorporate alterations at the molecular level into the histological diagnosis of cancer, advancing the rapidly growing field of “molecular morphology”.

The results of comparative tissue proteome studies in combination with cancer pathology and biology are expected to provide significant details at the global level of the molecular mechanisms associated with cancer. Identification of differentially expressed proteins that are characteristic of a clearly defined disease state paves the way for defining the molecular and biochemical pathways by which normal cells progress to cancerous states in addition to nurturing discovery of biological markers and therapeutic targets for cancer. The greatest expectations for targeted proteomics researched using enriched nonmalignant or malignant cells from high quality specimens reside in the identification of diagnostic, prognostic, and predictive biological markers in the clinical setting, as well as the discovery and validation of new protein targets in the biopharmaceutical industry.

1.6 ACKNOWLEDGEMENT

We thank the National Cancer Institute (CA103086 and CA107988) and the National Center for Research Resources (RR021239 and RR021862) for supporting portions of our research activities reviewed in this article.

CHAPTER TWO

CHARACTERIZATION OF THE HUMAN SALIVARY PROTEOME BY CAPILLARY ISOELECTRIC FOCUSING/NANO-REVERSED PHASE LIQUID CHROMATOGRAPHY COUPLED WITH ESI-TANDEM MS

Reproduced with permission from Tong Guo, Paul A. Rudnick, Weijie Wang,
Cheng S. Lee, Don L. DeVoe, and Brian M. Balgley, *J Prot. Res.* (2006), 5, 1469-
1478. Copyright 2006 American Chemical Society

2.1 INTRODUCTION

The components contributing to the composition of whole saliva include salivary gland secretions, blood and other fluids such as bronchial and nasal secretions, and cells from the lining of the oral cavity and microflora. Saliva, and in particular the protein component of saliva, is involved in numerous functions including antimicrobial activities, digestion, enamel remineralization, and lubrication. Several diseases such as alcoholic cirrhosis, cardiovascular diseases, cystic fibrosis, diabetes mellitus, and Sjogren's syndrome are known to directly or indirectly affect the functions of salivary glands[78-81]. In turn, changes in salivary composition correlate with disease susceptibility and/or progression. Human saliva is therefore a potential source of novel diagnostic markers and therapeutic targets[78, 82-85].

There has been growing interest in identifying “salivary biomarkers” as a means of monitoring general health and for the early diagnosis of diseases including bacterial infection, human immunodeficiency virus (HIV), and oral cancer[78, 82-85]. In comparison with other body fluids such as serum or urine, several further advantages exist in analyzing saliva including straightforward sample collection, sufficient quantities for analysis, and lower costs of storage and shipping. Thus, there is a critical need to create a periodic table of the salivary proteome that will be essential for the elucidation of disease pathogenesis and the evaluation of the influence of medications on the structure, composition, and secretion of salivary constituents. The salivary proteome will greatly complement the development and validation of the salivary-based diagnostic technologies by providing a catalog of salivary proteins.

Similar to other body fluids such as serum, the large variation of protein relative abundances together with the complexity of the saliva proteome greatly and continuously challenges the development of many aspects of bioanalytical technologies from sample processing to separation and MS detection. Current protein analysis technologies are still largely based on 2-D PAGE, which has undeniably assumed a major role and is central to much of what is now described as “proteomics”. A number of research groups[86-91] have employed the 2-D PAGE-MS technique for the analysis of whole saliva and parotid gland secretions and have identified some 200 proteins, likely representing a relatively small portion of the salivary proteome. Moreover, the application of 2-D PAGE-MS to saliva samples suffers from low throughput, poor reproducibility, and reduced sensitivity.

Total peak capacity improvements in multidimensional chromatography platforms[33, 77] have increased the number of detected peptides and proteins identified due to better use of the MS dynamic range and reduced discrimination during ionization. Based on single or multidimensional liquid chromatography, a total of 102 and 266 proteins were identified from human whole saliva by Wilmarth[92] and Hu[91], respectively. Griffin and co-workers[93] have recently combined free flow electrophoresis with reversed phase liquid chromatography (RPLC) for achieving multidimensional peptide separations and identified 437 salivary proteins. The overlap within their catalogs[91-93] was relatively small and included most of the common proteins present in high abundance and associated with housekeeping functions.

As demonstrated in our previous studies for the proteome analysis of yeast cell lysates[46, 47]and microdissected tissue specimens[15], combined CIEF/nano-RPLC separations have greatly enhanced the dynamic range and detection sensitivity of MS measurements due to high analyte concentrations in small peak volumes as the result of ultrahigh resolving power and electrokinetic focusing of low abundance proteins and peptides. CIEF, an isoelectric point (pI)-based separation, not only exhibits greater resolving power than that achieved in strong cation exchange chromatography, but also is completely orthogonal to the second dimension RPLC which resolves peptides based on their differences in hydrophobicity. Thus, CIEF/nano-RPLC separations coupled with ESI-tandem MS are employed in this study to address the challenges of protein complexity and protein relative abundance inherent in human saliva proteome. A total of 5,338 distinct peptides are identified

with 99% confidence from which are inferred the presence of 1,381 non-redundant proteins, which is the largest catalog of proteins from a single saliva sample to date.

2.2 EXPERIMENTAL SECTION

2.2.1 Materials and Reagents.

Fused-silica capillaries (50 μm i.d./375 μm o.d. and 100 μm i.d./375 μm o.d.) were acquired from Polymicro Technologies (Phoenix, AZ). Acetic acid, ammonium hydroxide, dithiothreitol (DTT), iodoacetamide (IAM), and protease inhibitors were obtained from Sigma (St. Louis, MO). Acetonitrile, hydroxypropyl cellulose (HPC, average MW 100,000), tris(hydroxymethyl)aminomethane (Tris), and urea were purchased from Fisher Scientific (Pittsburgh, PA). Pharmalyte 3-10 was acquired from Amersham Pharmacia Biotech (Uppsala, Sweden). Sequencing grade trypsin was obtained from Promega (Madison, WI). All solutions were prepared using water purified by a Nanopure II system (Dubuque, IA) and further filtered with a 0.22 μm membrane (Millipore, Billerica, MA).

2.2.2 Sample Collection and Preparation.

Whole unstimulated saliva was collected from a healthy male volunteer. 1 mL of saliva was placed in a tube containing a mixture of protease inhibitors (1 μg aprotinin, 1 μg pepstatin A, and 1 μg leupeptin) and centrifuged at 20,000 g for 30 min. The supernatant was collected and placed in a dialysis cup (Pierce, Rockford, IL) and dialyzed overnight at 4 $^{\circ}\text{C}$ against 100 mM Tris at pH 8.2. The protein concentration of the dialyzed sample was determined to be 2.3 mg/mL using the BCA assay (Pierce). Urea and DTT were added to the sample with final concentrations of 8

M and 1 mg/mL, respectively, and incubated at 37 °C for 2 hr under nitrogen. IAM was added to a concentration of 2 mg/mL and kept at room temperature for 1 hr in the dark. Trypsin was added at a 1:20 (w/w) enzyme to substrate ratio and incubated overnight at 37 °C. The protein digest was desalted using a reversed phase trap column (Michrom Bioresources, Auburn, CA) and lyophilized to dryness using a SpeedVac (ThermoSavant, San Jose, CA), and then stored at -80 °C.

2.2.3 Integrated CIEF/Nano-RPLC Multidimensional Separations.

On-line coupling of CIEF with nano-RPLC as a multidimensional peptide and protein separation platform has been developed and described in detail in previous work [15, 46, 47] and was employed for systematically resolving peptide digests based on their differences in pI and hydrophobicity. Briefly, an 84-cm long CIEF capillary (100 μm i.d./365 μm o.d.) coated with hydroxypropyl cellulose was initially filled with a solution containing 2% pharmalyte 3-10 and 2.3 mg/mL tryptic peptides. Peptides were focused by applying electric field strength of 300 V/cm and using solutions of 0.1 M acetic acid and 0.5% ammonium hydroxide as the anolyte and the catholyte, respectively

Focused peptides were sequentially fractionated by hydrodynamically loading into individual trap columns (3 cm x 200 μm i.d. x 365 μm o.d.) packed with 5- μm porous C18 reversed phase particles. Each peptide fraction was subsequently analyzed by nano-RPLC equipped with an Ultimate dual-quaternary pump (Dionex, Sunnyvale, CA) and a dual nano-flow splitter connected to two pulled-tip fused-silica capillaries (50 μm i.d. x 365 μm o.d.). These two 15-cm long capillaries were packed with 3- μm Zorbax Stable Bond (Agilent, Palo Alto, CA) C18 particles.

Nano-RPLC separations were performed in parallel in which a dual-quaternary pump delivered two identical 2-hr organic solvent gradients with an offset of 1 hr. Peptides were eluted at a flow rate of 200 nL/min using a 5-45% linear acetonitrile gradient over 100 min with the remaining 20 min for column regeneration and equilibration. The peptide eluants were monitored using a linear ion-trap mass spectrometer (LTQ, ThermoFinnigan, San Jose, CA) operated in a data dependent mode. Full scans were collected from 400 - 1400 m/z and 5 data dependent MS/MS scans were collected with dynamic exclusion set to 30 sec. A moving stage housing two nano-RPLC columns was employed to provide electrical contacts for applying ESI voltages, and most importantly to position the columns in-line with the ESI inlet at each chromatography separation and data acquisition cycle.

2.2.4 Data Analysis.

Raw LTQ data were converted to peak list files by `msn_extract.exe` (ThermoFinnigan). OMSSA [58] was used to search the peak list files against a decoyed SwissProt human database. This database was constructed by reversing all 12,484 real sequences and appending them to the end of the sequence library. Searches were performed by setting up the following parameters: fully tryptic, 1.5 Da precursor ion mass tolerance, 0.8 Da fragment ion mass tolerance, 1 missed cleavage, alkylated Cys as a fixed modification and variable modifications of acetylated N-terminus and Lys, phosphorylated Ser, Thr, and Tyr, and oxidated Met. Follow-up searches were also performed using semi-tryptic and no enzyme specificities without variable modifications. Searches were run in parallel on a 12 node, 24 CPU Linux cluster (Linux Networx, Bluffdale, UT).

False positive rates were determined using the method of Elias and co-workers[59]. Briefly, false positive rates were calculated by multiplying the number of false positive identifications (hits to the reversed sequences scoring below a given threshold) by 2 and dividing by the number of total identifications. Peptides identified below threshold, and also occurring as matches to the forward sequences, were not counted as false positives or true identifications. A curve was then generated by plotting E-value versus false positive rate where the definition of E-value is the expected number of random hits from a search library to a given spectrum that the random hits have an equal or better score than the hit[60]. An E-value threshold corresponding to a 1% false positive rate was used as a cutoff in this analysis.

Our experience has been that OMSSA displays exceptional specificity with known protein mixtures and in complex proteome samples (data not shown). However, spectral quality, search space, and input parameters influence any scoring metric and therefore necessitate the application of secondary validation strategies which anyone can employ using any search set, algorithm, and scoring metric. In our hands, a decoyed database search represents one such commonly accepted option. After generation of search data, the result files were parsed and loaded into a custom MySQL database for visualization and reporting using in-house software.

2.3 RESULTS AND DISCUSSION

Salivary components are important determinants of oral health and essential in regulating the oral ecology. Alterations in the composition of the oral flora and/or immune dysfunction are often linked to common oral diseases[78-85]. For example, differences in proteolytic processing of basic proline rich proteins (PRPs) from the

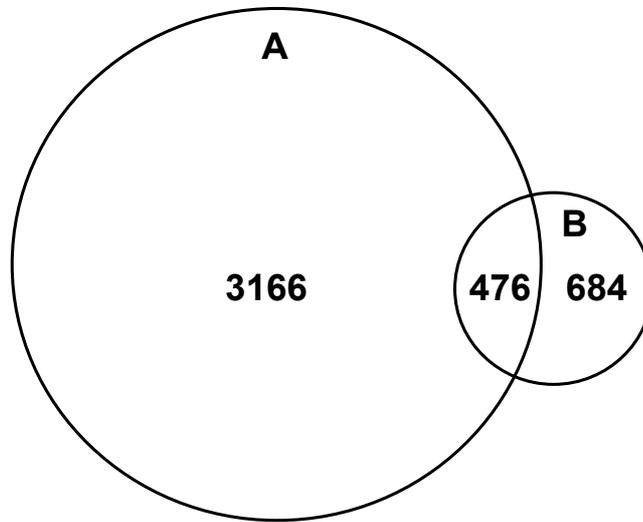
human parotid gland have been associated with caries[94]. Chen and co-workers have used MALDI-MS to compare saliva samples obtained from oral cancer patients with those from healthy control subjects[95]. Studying the expression of specific salivary proteins has demonstrated that the family of defensins exhibit antibiotic, antifungal, and antiviral properties[96-100].

2.3.1 Analysis of Tryptic Peptides Prepared from Whole Saliva.

In this study, protein digests obtained from a single human saliva sample were analyzed by combined CIEF/nano-RPLC separations. In all, 29 peptide fractions were sampled from the first dimension CIEF separation with a total peptide loading of only 15.6 μg (capillary volume of 6.8 μL x peptide concentration of 2.3 mg/mL). Each peptide fraction was further resolved by the second dimension nano-RPLC separation and the eluants were detected using ESI-tandem MS. Initial search of the resulting MS data led to the identification of 3,642 fully tryptic peptides (at a 1% false positive rate), covering 1,165 distinct SwissProt protein entries. As shown in Fig. 2-1, the search results were further compared with the most recent human saliva proteome reported by Xie and co-workers[93]. More than 50% of their cataloged saliva proteins were also identified in this study.

In addition to a CIEF-UV trace, the number of distinct peptides and the distribution of the peptide's mean pI values in each of the 29 CIEF fractions are further summarized in Fig. 2-2. The CIEF separation was able to cover a wide pH range of 4.0-9.8 using a 2% ampholyte 3-10 solution.

Distinct Peptides



Distinct Proteins

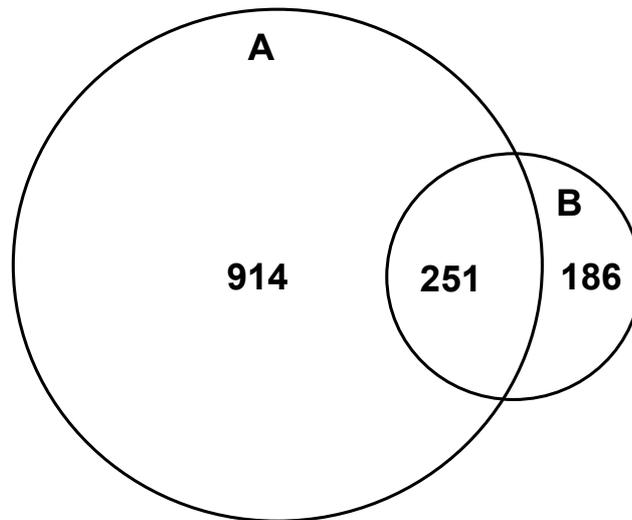


Figure 2-1 Venn diagrams comparing the proteome results obtained from this study (A) versus those achieved by Xie and co-workers[93](B).

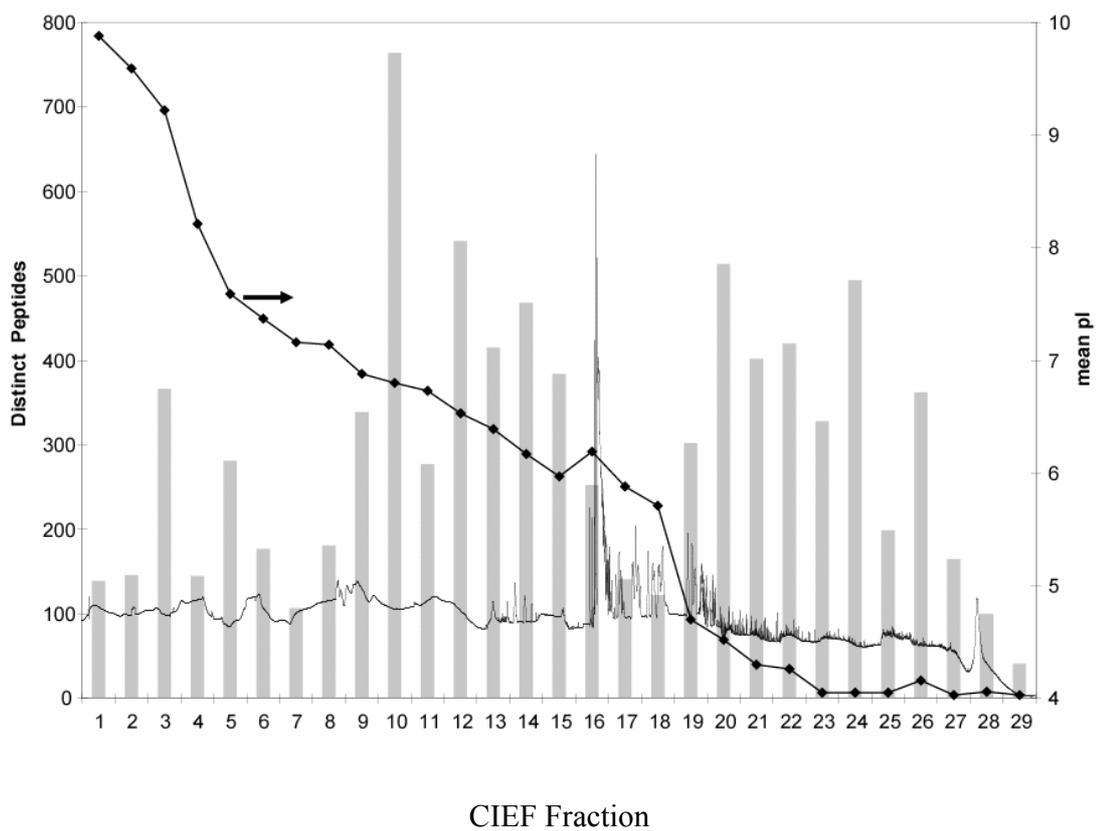


Figure 2-2 Overlaid plots containing the CIEF-UV trace monitored at 280 nm, the number of distinct peptides identified in each of the 29 CIEF fractions, and the distribution of the peptide's mean pI values in each fraction.

By comparing with immobilized pH gradient gels[52, 53], the significant advantages of performing pI-based separations using CIEF include the ability to use higher electric fields resulting in rapid and ultrahigh resolution separations and the amenability to automation and seamless coupling with nano-RPLC while avoiding any post-gel peptide extraction/ concentration procedures.

As shown in Fig. 2-3, both the false positive rate and the numbers of distinct peptide and protein identifications were plotted as functions of the E-value. An E-value of 1 by definition should indicate a 50% probability of a random match. As determined by a decoyed database search[59], an E-value of 1, however, returns a false positive rate of only 2.2%. It is likely that the actual number of possibilities for a random match in this case, which is defined by OMSSA as the number of MS/MS search queries and their fragment ion characteristics, is much less than the theoretical number of possibilities defined by all possible precursor ion masses for a qualifying peptide string within the search space and its fragment ion characteristics. An E-value threshold of 0.3, corresponding to 1% false positive, was therefore chosen as a cutoff in this study. This decoyed database search approach not only accurately reduces false negative identifications, but also controls the degree of false positives, while improving the predictive power of a typical OMSSA search. Due to the differences in the complexity of various proteome samples, the sample preparation procedures, the proteome measurements, and the search parameters, it is recommended to always perform a decoyed database search to determine specific score thresholds employed for peptide and protein identifications.

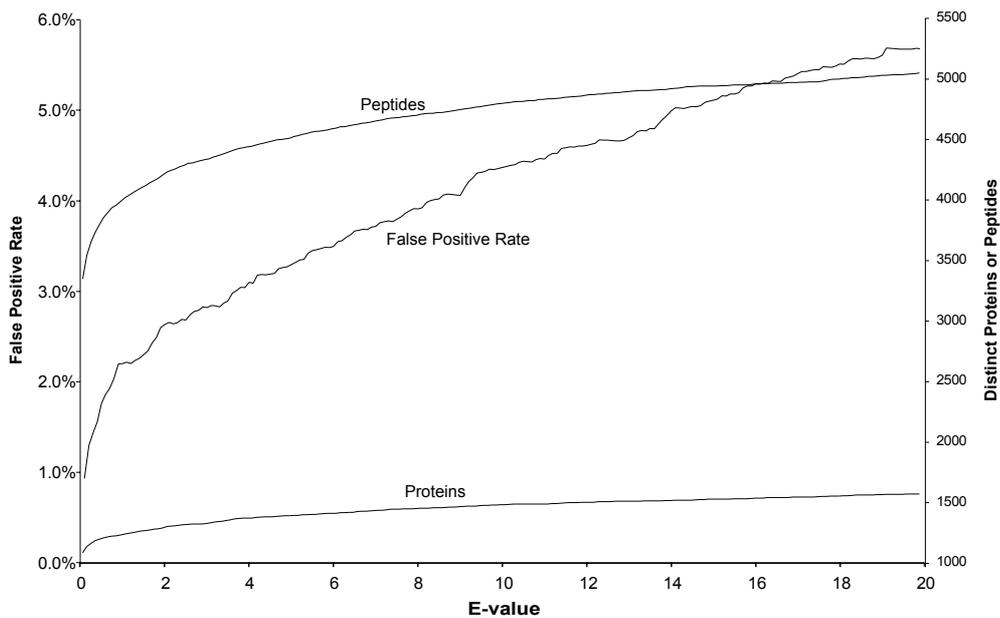


Figure 2-3 Plots of the false positive rate and the numbers of distinct peptide and protein identifications versus the E-value obtained from the search of the peak list files against a decoyed SwissProt human database using OMSSA[58].

In addition to the initial search for the identification of fully tryptic peptides, follow-up searches were performed for the detection of semi-tryptic and non-specific cleavages. The results for the semi-tryptic search constitute both peptides with an arginine or lysine at both the N- and C-terminal cleavage sites and peptides with an arginine or lysine at either one of the N- or C-terminal cleavage sites. The non-specific search results include both fully and semi-tryptic peptides as well as peptides which do not have an arginine or lysine at either cleavage site. As summarized in Table 2-1, the semi-tryptic search led to the identification of additional 1,314 peptides, of which 659 were semi-tryptic. The non-specific search resulted in 382 additional peptide identifications, of which 98 were non-tryptic. The semi-tryptic and non-tryptic searches contributed to the identification of 135 and 81 new proteins, respectively. By combining the fully tryptic, semi-tryptic, and non-specific searches, a total of 5,338 distinct peptides were identified for the coverage of 1,381 distinct SwissProt protein entries, giving average protein coverage of 3.9 peptides per protein.

In addition to performing searches against the SwissProt human database, a fully tryptic search was also conducted against the International Protein Index (IPI) human sequence database. As summarized in Table 2-2, this IPI-based search yielded 2,422 distinct protein sequence entry matches, more than two-fold of 1,165 proteins identified in the SwissProt search. Of these protein identifications, only 594 proteins were uniquely identifiable by a peptide which is unique to the IPI database, compared to the 876 proteins uniquely identifiable in the SwissProt search. On the other hand, the number of distinct peptides identified by either search differed by only 3%.

Table 2-1 Peptides and Proteins Identified at a False Positive Rate of 1%

Cleavage Rule	Total Peptide Hits	Distinct Tryptic Peptides	Distinct Semi-Tryptic Peptides	Distinct Non-specific Peptides	New Peptides	Total	Distinct Proteins	New Proteins
fully tryptic	18,992	3,642			3,642	3,642	1,165	1,165
semi-tryptic	16,674	3,088	659		1,314	3,747	1,090	135
non-specific	18,389	3,168	668	98	382	3,934	1,132	81
Total					5,338			1,381

Table 2-2 Proteins Identified from Fully Tryptic Peptides at a False Positive Rate of 1%

Human Database	Distinct Peptides	Distinct Protein Sequences Entries	Uniquely Identifiable	Non-Unique Clusters	Non-Identical Set
SwissProt	3,642	1,165	876	115	991
IPI	3,541	2,422	594	444	1,038

These comparisons clearly indicate that the IPI human database has a fairly large degree of protein sequence entry redundancy. While protein databases containing a large number of entries are often more comprehensive because they contain information for many isoforms and allele-specific translations, protein identification statistics cited using such databases should be examined with caution unless a minimization approach is employed to reduce the effects of inferring the presence of multiple proteins from single peptides where avoidable (i.e. Occam's Razor).

As shown in Table 2-2, proteins not identified by a unique peptide are clustered together in this study. Unique peptides are defined as distinct peptides which map to a single protein entry within the search database. A non-unique cluster can then be defined as a group of proteins which contain at least one distinct, but non-unique peptide. This non-unique cluster approach minimized the numbers of protein sequence entries containing overlapping peptide sequences. The non-identical set of proteins can therefore be defined as the sum of the uniquely identifiable and non-unique clusters categories. The number of proteins classified in the non-identical set from the two searches differed by only 4.5%, which is much less than the 2-fold difference observed in the number of distinct protein sequence entries and is comparable to the 3% difference in the distinct peptide identifications. The difference remaining between the non-identical set and the number of distinct protein sequence entries can be attributed to non-unique clusters which contain overlapping distinct peptides.

2.3.2 Detection of Small Proteins and Antimicrobial Proteins in Whole Saliva

Small proteins with a molecular mass of less than 20 kDa constitute a characteristic portion of the human saliva proteome. A total of 241 small proteins were identified in the combined fully tryptic, semi-tryptic, and non-specific searches. After removing structural proteins and proteins commonly found in serum, histones, and ribosomes, 103 small proteins remained and are summarized in Table 2-3. Tandem MS spectra of distinct peptides leading to the identification of representative small proteins are presented in Fig. 2-4. 48 of these 103 small proteins are PRPs which are involved in creating a protective film over teeth and promoting remineralization of the enamel[101]. Additionally, several antimicrobial proteins were also identified including β -defensin 125 precursor, β -defensin 1 precursor, FALL-39 precursor, mucins 5B and 2, 6 different cystatin precursors, and lysozyme.

2.3.3 Bacterial Proteins Identified in Whole Saliva

Besides the searches against human databases, a search was also conducted against all of the completed bacterial genomes to identify the presence of any unique bacterial proteins. This search of bacterial proteins is of particular interest as many species of bacterial flora colonize the oral cavity with important health implications. The resulting peptide hits were obtained using a rigorous score threshold at a 0.1% false positive rate. The identified peptide sequences were further filtered for isobaric matches (I \leftrightarrow L or K \leftrightarrow Q substitutions) and determined to be unique to that record within the NCBI non-redundant protein database. Proteins were then required to be identified by at least 2 peptide sequences unique to the NCBI non-redundant protein database.

Table 2-3 List of Small Proteins (< 20 kDa) Identified in Human Saliva Studies

Accession	Protein Name	Distinct Peptides	Mass
TYB4_HUMAN	Thymosin β -4 (T β 4) (Fx) [Contains: Hematopoietic system regulatory peptide (Seraspenside)].	2	4921.4
PRPE_HUMAN	Basic proline-rich peptide P-E (IB-9).	4	6023.7
ATOX1_HUMAN	Copper transport protein ATOX1 (Metal transport protein ATX1).	1	7401.6
BD01_HUMAN	β -defensin 1 precursor (BD-1) (Defensin, β 1) (hBD-1).	1	7419.7
SPR2E_HUMAN	Small proline-rich protein 2E (SPR-2E) (Small proline-rich protein II) (SPR-II).	2	7855.3
SPR2D_HUMAN	Small proline-rich protein 2D (SPR-2D) (Small proline-rich protein II) (SPR-II).	1	7905.3
SPR2A_HUMAN	Small proline-rich protein 2A (SPR-2A) (2-1).	2	7965.4
SPR2B_HUMAN	Small proline-rich protein 2B (SPR-2B).	2	7975.4
SPR2G_HUMAN	Small proline-rich protein 2G (SPR-2G).	1	8157.6
PROL3_HUMAN	Proline-rich protein 3 precursor (proline-rich peptide P-B) [contains: Peptide P-A; Peptide D1A].	21	8187.6
HOP_HUMAN	Homeodomain-only protein (Lung cancer-associated Y protein) (Odd homeobox protein 1) (Not expressed in choriocarcinoma protein 1).	1	8260.1
CRIP1_HUMAN	Cysteine-rich protein 1 (Cysteine-rich intestinal protein) (CRIP) (Cysteine-rich heart protein) (hCRHP).	1	8401.6
TFF3_HUMAN	Trefoil factor 3 precursor (Intestinal trefoil factor) (hP1.B).	1	8641.0
ECGR2_HUMAN	Esophagus cancer-related gene-2 protein precursor (ECRG-2).	1	9231.8
PRP5_HUMAN	Basic proline-rich peptide IB-1.	9	9530.4
NMES1_HUMAN	Normal mucosa of esophagus specific gene 1 protein (FOAP-11 protein).	1	9617.3
FDSCP_HUMAN	Follicular dendritic cell secreted peptide precursor (FDC-SP) (FDC secreted protein).	1	9700.3
NIC1_HUMAN	NICE-1 protein.	2	9735.8
SPR1A_HUMAN	Cornifin A (Small proline-rich protein IA)	1	9882.5

MAN	(SPR-IA) (SPRK) (19 kDa pancornulin).		
SPR1B_HUMAN	Cornifin B (Small proline-rich protein IB) (SPR-IB) (14.9 kDa pancornulin).	3	9899.6
ACBP_HUMAN	Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI) (Endozepine) (EP).	1	9913.2
S10A6_HUMAN	Calcyclin (Prolactin receptor associated protein) (PRA) (Growth factor-inducible protein 2A9) (S100 calcium-binding protein A6) (MLN 4).	1	10179.7
SH3L3_HUMAN	SH3 domain-binding glutamic acid-rich-like protein 3 (SH3 domain-binding protein SH3BP-1) (P1725).	1	10437.7
S10AC_HUMAN	Calgranulin C (CAGC) (CGRP) (Neutrophil S100 protein) (Calcium-binding protein in amniotic fluid 1) (CAAF1) (p6) [Contains: Calcitermin].	5	10443.8
CH10_HUMAN	10 kDa heat shock protein, mitochondrial (Hsp10) (10 kDa chaperonin) (CPN10) (Early-pregnancy factor) (EPF).	2	10800.5
S10A8_HUMAN	Calgranulin A (Migration inhibitory factor-related protein 8) (MRP-8) (Cystic fibrosis antigen) (CFAG) (P8) (Leukocyte L1 complex light chain) (S100 calcium-binding protein A8) (Calprotectin L1L subunit) (Urinary stone protein band A).	15	10834.5
S10A2_HUMAN	S100 calcium-binding protein A2 (S-100L protein) (CAN19).	5	10985.5
S10AA_HUMAN	Calpactin I light chain (S100 calcium-binding protein A10) (p10 protein) (p11) (Cellular ligand of annexin II).	3	11071.9
CYTB_HUMAN	Cystatin B (Liver thiol proteinase inhibitor) (CPI-B) (Stefin B).	7	11139.6
S10A7_HUMAN	S100 calcium-binding protein A7 (Psoriasin).	3	11325.7
RSN_HUMAN	Resistin precursor (Cysteine-rich secreted protein FIZZ3) (Adipose tissue-specific secretory factor) (ADSF) (C/EBP-epsilon regulated myeloid-specific secreted cysteine-rich protein) (Cysteine-rich secreted protein A12-alpha-like 2).	1	11419.3
S11Y_HUMAN	Putative S100 calcium-binding protein H NH0456N16.1.	2	11509.3
THIO_HUMAN	Thioredoxin (ATL-derived factor) (ADF) (Surface associated sulphhydryl protein) (SASP).	3	11606.3
S10AE_HUMAN	S100 calcium-binding protein A14 (S114).	5	11662.

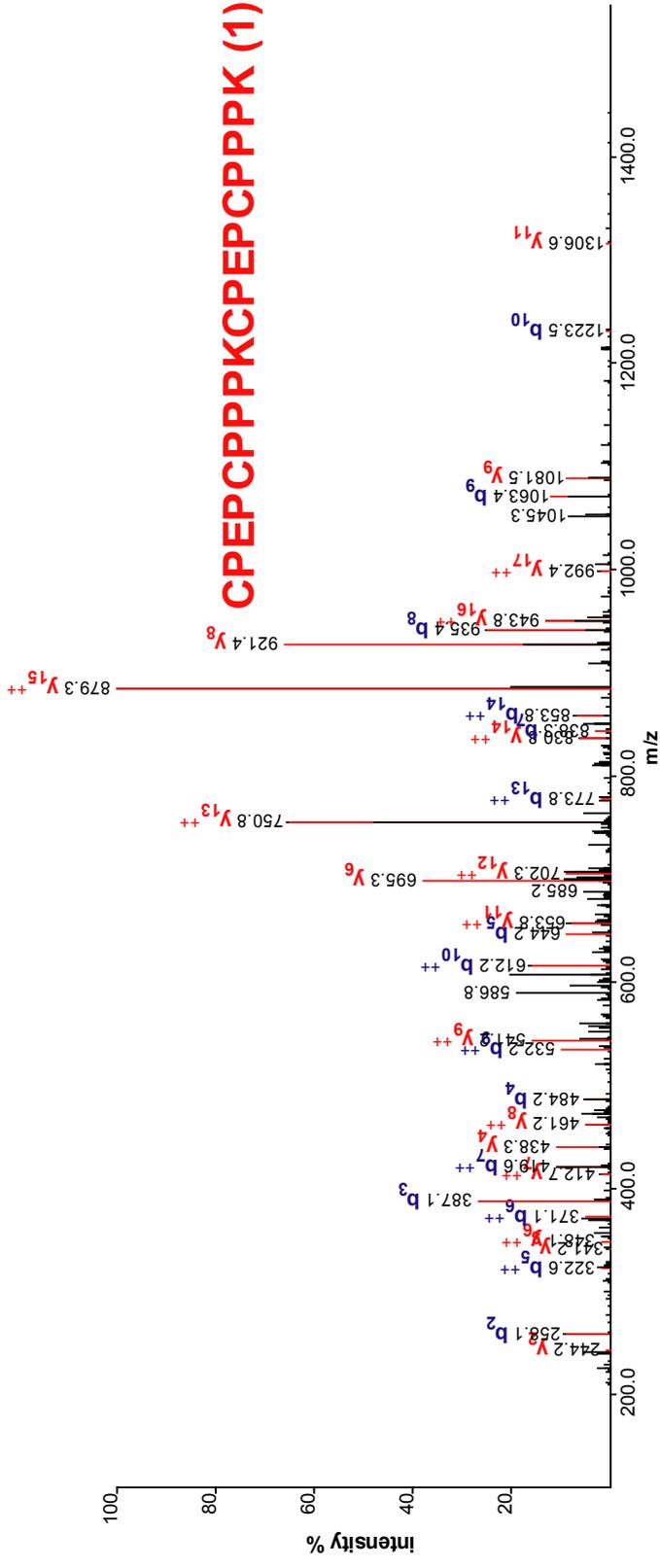
MAN			1
S10AB_HUMAN	Calgizzarin (S100 calcium-binding protein A11) (S100C protein) (MLN 70).	6	11740.4
S10AG_HUMAN	S100 calcium-binding protein A16 (S100F).	5	11801.4
FKB1A_HUMAN	FK506-binding protein 1A (EC 5.2.1.8) (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase) (12 kDa FKBP) (FKBP-12) (Immunophilin FKBP12).	1	11819.5
ERH_HUMAN	Enhancer of rudimentary homolog.	1	12258.9
ELAF_HUMAN	Elafin precursor (Elastase-specific inhibitor) (ESI) (Skin-derived antileukoproteinase) (SKALP) (WAP four-disulfide core domain protein 14) (Protease inhibitor WAP3).	1	12269.6
MGP_HUMAN	Matrix Gla-protein precursor (MGP).	1	12323.1
MIF_HUMAN	Macrophage migration inhibitory factor (MIF) (Phenylpyruvate tautomerase) (EC 5.3.2.1) (Glycosylation-inhibiting factor) (GIF).	2	12345.1
ELOC_HUMAN	Transcription elongation factor B polypeptide 1 (RNA polymerase II transcription factor SIII subunit C) (SIII p15) (Elongin C) (EloC) (Elongin 15 kDa subunit).	2	12473.1
PLAC8_HUMAN	Placenta-specific gene 8 protein (C15 protein).	1	12506.6
MTPN_HUMAN	Myotrophin (V-1 protein).	2	12763.6
WFDC2_HUMAN	WAP four-disulfide core domain protein 2 precursor (Major epididymis- specific protein E4) (Epididymal secretory protein E4) (Putative protease inhibitor WAP5).	1	12992.9
S10A9_HUMAN	Calgranulin B (Migration inhibitory factor-related protein 14) (MRP- 14) (P14) (Leukocyte L1 complex heavy chain) (S100 calcium-binding protein A9) (Calprotectin L1H subunit).	17	13242.0
SMD1_HUMAN	Small nuclear ribonucleoprotein Sm D1 (snRNP core protein D1) (Sm-D1) (Sm-D autoantigen).	1	13281.5
NUFM_HUMAN	NADH-ubiquinone oxidoreductase 13 kDa-B subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-13Kd-B) (CI-13Kd-B) (Complex I subunit B13).	1	13327.5
VATF_HUMAN	Vacuolar ATP synthase subunit F (EC 3.6.3.14) (V-ATPase F subunit) (Vacuolar	1	13358.1

	proton pump F subunit) (V-ATPase 14 kDa subunit).		
HINT1_HUMAN	Histidine triad nucleotide-binding protein 1 (Adenosine 5'-monophosphoramidase) (Protein kinase C inhibitor 1) (Protein kinase C-interacting protein 1) (PKCI-1).	2	13670.7
LYG6C_HUMAN	Lymphocyte antigen 6 complex locus G6C protein precursor (Protein NG24).	1	13821.2
CD59_HUMAN	CD59 glycoprotein precursor (Membrane attack complex inhibition factor) (MACIF) (MAC-inhibitory protein) (MAC-IP) (Protectin) (MEM43 antigen) (Membrane inhibitor of reactive lysis) (MIRL) (20 kDa homologous restriction factor) (HRF-20) (HRF20)	3	14177.2
CCL28_HUMAN	Small inducible cytokine A28 precursor (CCL28) (Mucosae-associated epithelial chemokine) (MEC) (CCK1 protein).	1	14279.6
ALK1_HUMAN	Antileukoproteinase 1 precursor (ALP) (HUSI-1) (Seminal proteinase inhibitor) (Secretory leukocyte protease inhibitor) (BLPI) (Mucus proteinase inhibitor) (MPI) (WAP four-disulfide core domain protein 4) (Protease inhibitor WAP4).	8	14326.0
PROF1_HUMAN	Profilin-1 (Profilin I).	4	14923.0
LEG7_HUMAN	Galectin-7 (Gal-7) (HKL-14) (PI7) (p53-induced protein 1).	6	14943.8
FABPE_HUMAN	Fatty acid-binding protein, epidermal (E-FABP) (Psoriasis-associated fatty acid-binding protein homolog) (PA-FABP).	15	15033.2
PEA15_HUMAN	Astrocytic phosphoprotein PEA-15 (Phosphoprotein enriched in diabetes) (PED).	1	15040.0
PROL4_HUMAN	Proline-rich protein 4 precursor (Lacrimal proline-rich protein) (Nasopharyngeal carcinoma-associated proline rich protein 4).	2	15096.7
NDK8_HUMAN	Putative nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase).	1	15529.0
RABP2_HUMAN	Retinoic acid-binding protein II, cellular (CRABP-II).	4	15561.8
MK_HUMAN	Midkine precursor (MK) (Neurite outgrowth-promoting protein) (Midgestation and kidney protein) (Amphiregulin-associated protein) (ARAP) (Neurite outgrowth-promoting factor 2).	1	15585.1
CYTC_HUMAN	Cystatin C precursor (Neuroendocrine basic	4	15799.

AN	polypeptide) (Gamma-trace) (Post-gamma-globulin).		2
SODC_HUMAN	Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).	2	15804.5
CYTD_HUMAN	Cystatin D precursor.	6	16080.4
PA2GA_HUMAN	Phospholipase A2, membrane associated precursor (EC 3.1.1.4) (Phosphatidylcholine 2-acylhydrolase) (Group IIA phospholipase A2) (GIIC sPLA2) (Non-pancreatic secretory phospholipase A2) (NPS-PLA2).	2	16082.6
CDD_HUMAN	Cytidine deaminase (EC 3.5.4.5) (Cytidine aminohydrolase).	2	16184.6
CYTS_HUMAN	Cystatin S precursor (Salivary acidic protein-1) (Cystatin SA-III).	19	16214.3
IF1AY_HUMAN	Eukaryotic translation initiation factor 1A, Y-chromosomal (eIF-1A Y isoform) (eIF-4C).	1	16311.1
IF1AH_HUMAN	Putative eukaryotic translation initiation factor 1A (eIF-1A) (eIF-4C).	1	16329.1
IF1AX_HUMAN	Eukaryotic translation initiation factor 1A, X-chromosomal (eIF-1A X isoform) (eIF-4C).	1	16329.2
CYTN_HUMAN	Cystatin SN precursor (Salivary cystatin SA-1) (Cystatin SA-I).	21	16361.6
CYTT_HUMAN	Cystatin SA precursor (Cystatin S5).	10	16444.6
LYSC_HUMAN	Lysozyme C precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramidase C).	46	16537.0
NPC2_HUMAN	Epididymal secretory protein E1 precursor (Niemann-Pick disease type C2 protein) (hE1).	1	16570.2
PIP_HUMAN	Prolactin-inducible protein precursor (Secretory actin-binding protein) (SABP) (Gross cystic disease fluid protein 15) (GCDFP-15) (gp17).	11	16572.4
GMFB_HUMAN	Glia maturation factor beta (GMF-beta).	1	16581.9
IF5A_HUMAN	Eukaryotic translation initiation factor 5A (eIF-5A) (eIF-4D) (Rev-binding factor).	3	16701.0
CALL3_HUMAN	Calmodulin-related protein NB-1 (Calmodulin-like protein) (CLP).	6	16759.5
COX5A_HUMAN	Cytochrome c oxidase polypeptide Va, mitochondrial precursor (EC 1.9.3.1).	3	16774.2
PRPC_HUMAN	Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].	22	17016.4
DB125_HUMAN	β -defensin 125 precursor (Beta-defensin 25)	1	17065.

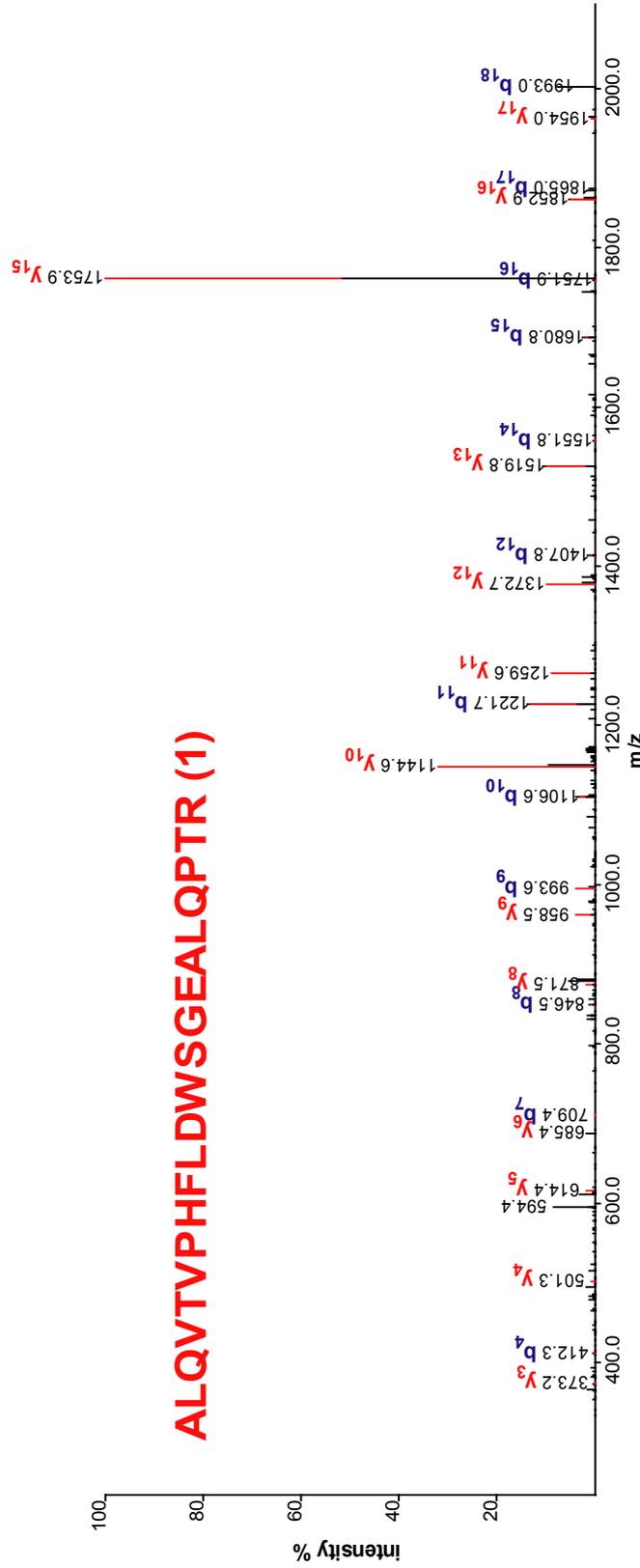
MAN	(DEFB-25).		4
NDKA_HUMAN	Nucleoside diphosphate kinase A (EC 2.7.4.6) (NDK A) (NDP kinase A) (Tumor metastatic process-associated protein) (Metastasis inhibition factor nm23) (nm23-H1) (Granzyme A-activated DNase) (GAAD).	2	17148.7
STMN1_HUMAN	Stathmin (Phosphoprotein p19) (pp19) (Oncoprotein 18) (Op18) (Leukemia-associated phosphoprotein p18) (pp17) (Prosolin) (Metablastin) (Pr22 protein).	1	17171.3
SSB_HUMAN	Single-stranded DNA-binding protein, mitochondrial precursor (Mt-SSB) (MtSSB) (PWP1-interacting protein 17).	1	17259.6
NDKB_HUMAN	Nucleoside diphosphate kinase B (EC 2.7.4.6) (NDK B) (NDP kinase B) (nm23-H2) (C-myc purine-binding transcription factor PUF).	2	17298.0
PRPP_HUMAN	Salivary proline-rich protein II-1 (Fragment).	3	17836.3
UB2L3_HUMAN	Ubiquitin-conjugating enzyme E2 L3 (EC 6.3.2.19) (Ubiquitin-protein ligase L3) (Ubiquitin carrier protein L3) (UbcH7) (E2-F1) (L-UBC).	1	17861.5
PPIA_HUMAN	Peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin A) (Cyclosporin A-binding protein).	7	17881.2
SPRR3_HUMAN	Small proline-rich protein 3 (Cornifin beta) (Esophagin) (22 kDa pancornulin).	20	18153.9
COF1_HUMAN	Cofilin-1 (Cofilin, non-muscle isoform) (18 kDa phosphoprotein) (p18).	9	18371.2
DEST_HUMAN	Destrin (Actin-depolymerizing factor) (ADF).	3	18374.5
ECP_HUMAN	Eosinophil cationic protein precursor (EC 3.1.27.-) (ECP) (Ribonuclease 3) (RNase 3).	3	18440.3
COF2_HUMAN	Cofilin-2 (Cofilin, muscle isoform).	2	18736.6
CG38_HUMAN	Protein CGI-38.	1	18985.4
VEGP_HUMAN	Von Ebner's gland protein precursor (VEG protein) (Tear prealbumin) (TP) (Tear lipocalin) (Lipocalin 1).	19	19249.9
FAL39_HUMAN	Antibacterial protein FALL-39 precursor (FALL-39 peptide antibiotic) (Cationic antimicrobial protein CAP-18) (hCAP-18) (HSD26) [Contains: Antibacterial protein LL-37].	4	19301.3
SFRS3_HUMAN	Splicing factor, arginine/serine-rich 3 (Pre-	1	19329.

MAN	mRNA splicing factor SRP20).		5
UB2G1_HUMAN	Ubiquitin-conjugating enzyme E2 G1 (EC 6.3.2.19) (Ubiquitin-protein ligase G1) (Ubiquitin carrier protein G1) (E217K) (UBC7).	1	19378.0
APT_HUMAN	Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).	1	19476.5
CETN1_HUMAN	Centrin-1 (Caltractin isoform 2).	1	19570.0
TCTP_HUMAN	Translationally controlled tumor protein (TCTP) (p23) (Histamine-releasing factor) (HRF).	1	19595.3
PARK7_HUMAN	DJ-1 protein (Oncogene DJ1).	1	19891.0

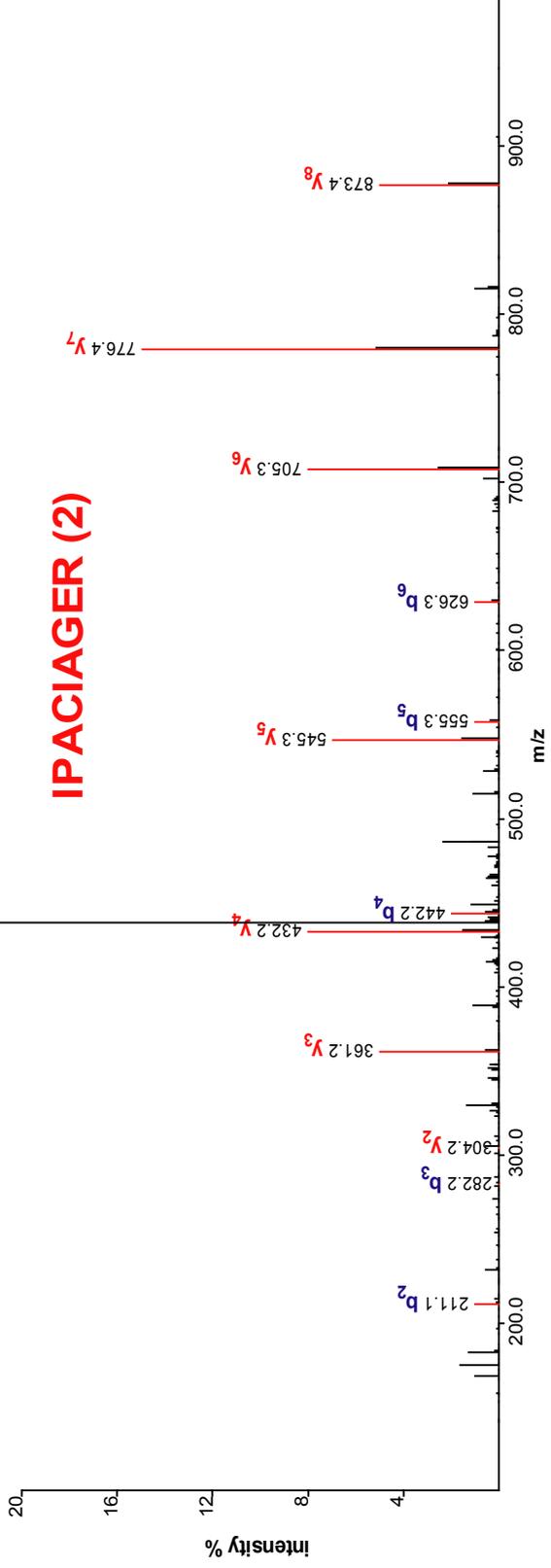


P35326 Small proline-rich protein 2A (SPR-2A) (2-1)

ALQVTPHFLDWSGEALQPTR (1)



Q8N4F0 Bactericidal/permeability-increasing protein-like 1 precursor



P59666 Neutrophil defensin 3 precursor

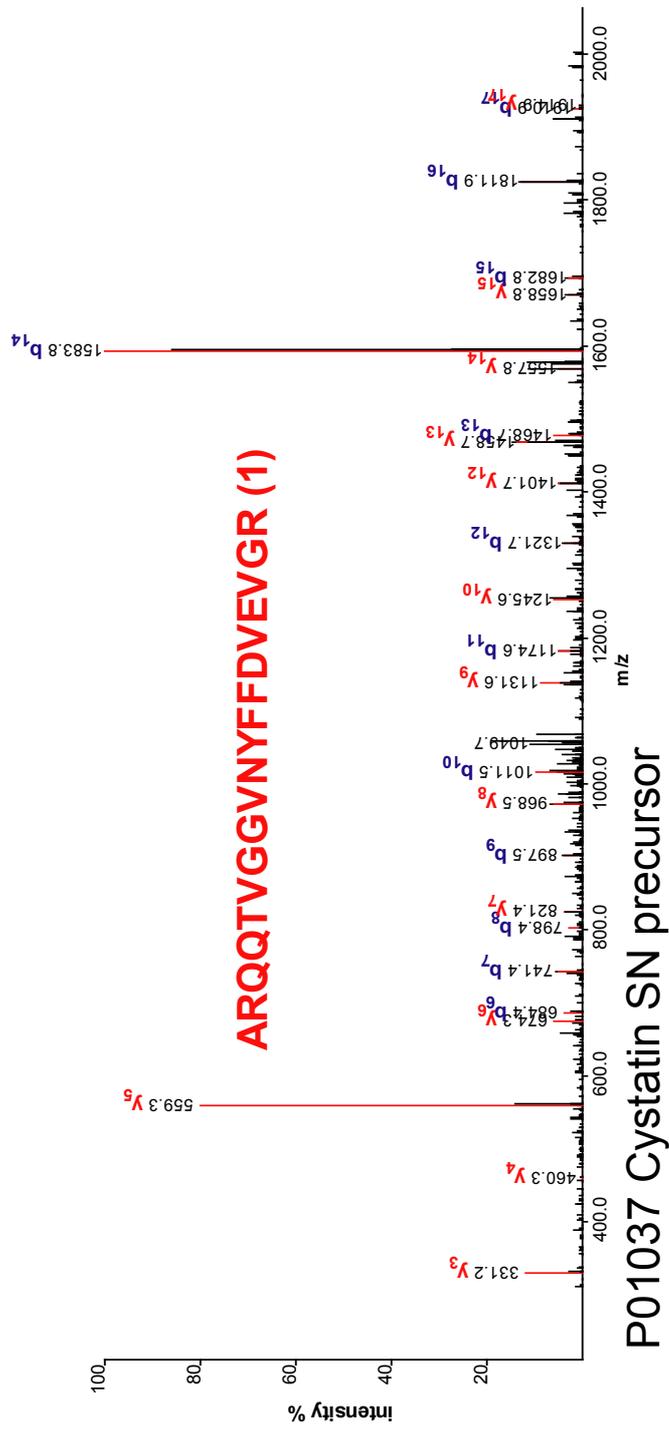


Figure 2-4 Tandem MS spectra of distinct peptide hits contributing to the identification of representative small proteins.

A total of 31 bacterial species, inferred from these protein identifications, are summarized in Table 2-4. Some of the bacteria on the list are rare and their presence in a whole saliva sample is unexpected. It should be noted that some of the peptide sequences may map to unsequenced peptide homologs unique to other species or missequenced proteins in either human or bacterial proteomes. The most represented species was *Streptococcus mutans* which is the bacteria known to cause dental caries and is ubiquitous in the human population[84]. *S. mutans* was identified by a total of 29 proteins which contained peptides unique in the database to this species. In particular, the secretome of *S. mutans* is of great interest as the proteins found in whole saliva likely constitute an important part of its arsenal in preventing the growth of other bacteria and developing a biofilm on dental surfaces.

Porphyromonas gingivalis, which is associated with periodontal disease, was identified by 4 proteins. *Helicobacter pylori*, which is associated with peptic ulcer disease and chronic gastritis, was identified by 2 proteins. Its rate of infection is estimated at 80% in the developed world versus 40% in the developing world. Gastric infection by *H. pylori* is currently diagnosed using techniques which rely on a biopsy of the gastric mucosa. Recently, it has been shown that PCR amplification of *H. pylori* DNA could be used to determine infection from both gastric mucosa biopsy and saliva samples of infected patients. An ELISA test is also available for the detection of *H. pylori* specific immunoglobulin G antibodies in serum.

Table 2-4 Bacterial Species Identified by Two or More Fully Tryptic Peptide Sequences (False Positive Rate of 0.1%) Which Are Unique to the NCBI Non-Redundant Protein Database

Acinetobacter sp. ADP1
Archaeoglobus fulgidus DSM 4304
Bacillus subtilis subsp. subtilis str.68 2
Bdellovibrio bacteriovorus HD100
Bordetella pertussis Tohama I
Burkholderia mallei ATCC 23344
Clostridium perfringens str.3 2
Clostridium tetani E88
Enterococcus faecalis V583
Fusobacterium nucleatum subsp. nucleatum ATCC 25586
Haemophilus ducreyi 35000HP
Helicobacter pylori 26695
Lactobacillus plantarum WCFS1
Mesorhizobium loti MAFF303099
Methanococcus maripaludis S2
Methanosarcina acetivorans C2A 2
Methylococcus capsulatus str. Bath
Mycoplasma penetrans HF-2
Mycoplasma pulmonis UAB CTIP
Parachlamydia sp. UWE25
Photobacterium profundum SS9
Porphyromonas gingivalis W83
Pyrobaculum aerophilum str. IM2
Pyrococcus horikoshii OT3
Streptococcus agalactiae NEM316
Streptococcus mutans UA159
Streptococcus pneumoniae TIGR4
Streptomyces avermitilis MA-4680
Thermoanaerobacter tengcongensis MB4
Vibrio vulnificus YJ016
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis

Antibodies present in saliva against *H. pylori* have been examined and resulted in a predictive value of 45%[102]. Our saliva proteome results offer unique peptides or proteins which can be targeted by antibodies in the diagnosis of *H. pylori* infection. However, the relationship between gastric and oral infection remains unclear.

2.4 CONCLUSION

There is a growing interest in using saliva as a diagnostic fluid due to its relatively simple and minimally invasive collection procedures. For example, immunoassays targeting saliva samples have been developed for HIV, hepatitis B, and measles [103-105]. Saliva has also been used for hormone and drug screening since analyte levels in saliva are indicative of the unbound analyte concentrations [84, 106]. However, as with any body fluids, the saliva proteome exhibits a large variation of protein relative abundances including high abundance proteins such as amylases, mucins, PRPs, and secretory IgA complex. Furthermore, salivary proteins present a rich array of PTMs such as glycosylation and phosphorylation which are important regulators of protein function [107-110]. Increasing evidence also suggests that proteolytic processing is another PTM of salivary proteins, illustrated by cleavages involving histatin family members, statherins, PRPs, and cystatins [94, 111, 112].

By using a minimal peptide loading of only 15.6 μg , a total of 5,338 distinct peptides were detected at a 1% false positive rate, leading to the identification of 1,381 distinct SwissProt protein entries from a single human saliva sample. Among these categorized proteins, 103 small proteins (< 20 kDa), representing a characteristic portion of saliva proteome, were identified including PRPs and

antimicrobial proteins such as β -defensin 125 precursor, β -defensin 1 precursor, FALL-39 precursor, mucins 5B and 2, 6 different cystatin precursors, and lysozyme.

Additionally, the tandem MS spectra were searched against a database of completed bacterial genomes, resulting in uniquely identifiable proteins possibly indicating the presence of 31 bacterial species. These proteome results highlight the potential use of saliva for the search of diagnostic markers of human disease as well as exposure to common pathogens. The results further underline the need to couple the CIEF-based multidimensional separation platform with other sample fractionation techniques such as ultrafiltration [90, 91] in order to enhance the ability for mining deeper into the saliva proteome. In analogy to the serum proteome studies, the application of affinity columns to remove highly abundant salivary proteins will certainly play an important role in aiding the enrichment of low abundant species while providing more complete proteome coverage.

2.5 ACKNOWLEDGEMENT

We thank the National Center for Research Resources (RR021239 and RR021862) and the National Cancer Institute (CA103086 and CA107988) for their support of portions of this research.

CHAPTER THREE

PROTEOME ANALYSIS OF MICRODISSECTED FORMALIN-FIXED AND PARAFFIN-EMBEDDED TISSUE SPECIMENS

Reproduced with permission from Tong Guo, Weijie Wang, Paul A. Rudnick, Tao Song, Jie Li, Zhengping Zhuang, Robert J. Weil, Don L. DeVoe, Cheng S. Lee, and Brian M. Balgley, *J. Histochem. Cytochem.* (2007), 55, 763-772

Copyright 2007 The Histochemical Society

3.1 INTRODUCTION

Because of the long history of the use of formalin as the standard fixative for tissue processing in histopathology, there is a large number of archival FFPE tissue banks worldwide. These FFPE tissue collections, with attached clinical and outcome information, present invaluable resources for conducting protein biomarker investigations. However, the high degree of covalently cross-linked proteins in FFPE tissues generally hinders efficient extraction of proteins, which has limited bioanalytical exploration of the potential information available in archival tissue banks.

The ability to identify proteins within FFPE tissue specimens is greatly enhanced by a simple and effective antigen retrieval (AR) technology, in which boiling the FFPE tissue sections in water or buffer solution dramatically reduces the detection thresholds (increases sensitivity) of IHC staining for a wide range of

antibodies[66-68, 113]. In this context, IHC as applied to the demonstration of antigens (primarily proteins) in tissue sections represents a first form of protein analysis. The mechanism of AR appeared to involve a re-naturation of the structure of fixed proteins through a series of conformational changes, including the possible breaking (hydrolysis) of formalin-induced cross-linkages, the entire process being driven by thermal energy from the heat source[67, 114]. Still, the high-temperature heating treatment may also induce negative results such as additional protein modifications and therefore require further studies and comparisons with low-temperature heating treatment and combined retrieval protocols involving heat and enzyme digestion[74, 75].

Based on the principle of heat-induced AR technique, Ikeda and co-workers[115] have extracted proteins from FFPE tissues by heating the sections in a radioimmunoprecipitation buffer containing 2% SDS. Subsequently, Yamashita and Okada[116] performed protein extraction from FFPE tissues by autoclaving the sections, followed by incubating the sections with a solution containing urea, 2-mercaptoethanol, and 2% SDS. A commercial Liquid Tissue™ kit apparently based on the same AR technique was recently introduced for processing FFPE tissues also in the presence of detergent while heating at 95 °C for 90 min[45, 70].

In addition to IHC-based tissue proteome studies, current protein/peptide separation platforms, including 2-D PAGE and multidimensional liquid chromatography systems[33, 77], all require large cellular samples in order to increase proteomic coverage. Thus, most tissue proteomic studies have employed multidimensional liquid chromatography separations[38-41] and are mainly based on

analysis of entire tissue sections instead of targeted subpopulations of microdissection-derived cells. However, the heterogeneous nature of most tissues, as well as the fact that in many cases the cells of interest may be in the minority and may be surrounded by normal or other abnormal cells limits the ultimate utility of whole tissue proteome studies in many instances.

Several tissue microdissection technologies, including LCM [10, 11], laser microdissection [12], and laser-free microdissection [13, 14], have been developed to provide a rapid, straightforward method for procuring homogeneous subpopulations of cells or structures for biochemical and molecular biological analyses. However, the smaller quantities of ample available for microdissected populations has, to this point, restricted protein analyses to the use of only a single chromatography separation prior to tandem MS analysis and limited the ability for mining deeper into the tissue proteome in recent studies[42-45, 72, 73]. Since the sizes of human tissue biopsies are becoming significantly smaller due to the advent of minimally-invasive methods and early detection and treatment of lesions, a more effective discovery-based proteome technology is critically needed to enable sensitive studies of protein profiles that will have diagnostic and therapeutic relevance.

The key to performing sensitive tissue proteome analysis, as demonstrated in our previous studies [15, 117], is to attain high analyte concentrations in small peak volumes. We have employed electrokinetic focusing and the high resolving power of a CIEF-based multidimensional separation platform to enhance the dynamic range and detection sensitivity of MS measurements. By coupling with the heat-induced AR technique, combined CIEF/nano-reversed-phase liquid chromatography (nano-RPLC)

separations have been demonstrated for the analysis of proteins extracted from entire FFPE tissue sections [71]. Instead of using whole sections, minute proteins procured from microdissected GBM FFPE tissues, as the model proteome system employed in this study, are processed, profiled, and compared with those extracted from fresh frozen tissues of the same, matched patient. The work reported here represents an important milestone toward the development, evaluation, and validation of a novel biomarker discovery paradigm on the basis of years of archived FFPE tissue collections.

3.2 EXPERIMENTAL SECTION

3.2.1 Clinical Materials.

Tissues and clinical (pathological) information were obtained as part of an Institutional Review Board-approved study at the Cleveland Clinic. At the time of craniotomy, tissue samples were split equally, one portion being sent for routine processing in the Pathology department, the other being snap-frozen, as noted below. For the portion sent to pathology, from which a clinical diagnosis was made, the sample was processed in the routine fashion, fixed in formalin overnight, embedded in paraffin, and stored at room temperature after use. Approximately 6-12 months later, 6- m thick unstained sections were cut from this block and used for analysis, as described below. The second portion of tissue was immediately snap-frozen in liquid nitrogen in the operating room, embedded in the Optimal Cutting Temperature Medium (Tissue-Tek, Sakura, Finetek, Torrence, CA), and stored at -80 °C.

3.2.2 Materials and Reagents.

Fused-silica capillaries (50 μm i.d./375 μm o.d. and 100 μm i.d./375 μm o.d.) were acquired from Polymicro Technologies. Acetic acid, ammonium acetate, ammonium hydroxide, ampholyte 3-10, DTT, formic acid, IAM, and octane were obtained from Sigma. Acetonitrile, hydroxypropyl cellulose (HPC, average MW 100,000), SDS, Tris, and urea were purchased from Fisher Scientific. Sequencing grade trypsin was obtained from Promega (Madison, WI). All solutions were prepared using water purified by a Nanopure II system (Dubuque, IA) and further filtered with a 0.22 μm membrane (Millipore, Billerica, MA).

3.2.3 Tissue Microdissection and Protein Sample Preparation.

FFPE tissues were de-paraffinized using octane, followed by vortexing and centrifugation [71]. Both de-paraffinized FFPE and fresh frozen tissues were microdissected by following the procedures described in our previous studies [13, 14] to gather approximately 100,000 tumor cells from each tissue specimen.

The microdissected cells obtained from fresh frozen tissues were placed directly into a microcentrifuge tube containing 8 M urea and 20 mM Tris-HCl at pH 8.0. The soluble proteins were collected in the supernatant by centrifugation at 20,000 g for 30 min. Proteins in the supernatant were reduced and alkylated by sequentially adding DTT and IAM with final concentrations of 10 mg/mL and 20 mg/mL, respectively. The solution was incubated at 37 °C for 1 hr in the dark and then diluted 8-fold with 100 mM ammonium acetate at pH 8.0. Trypsin was added at a 1:40 (w/w) enzyme to substrate ratio and the solution was incubated at 37 °C overnight. Tryptic digests were desalted using a Peptide MacroTrap column (Michrom Bioresources,

Auburn, CA), lyophilized to dryness using a SpeedVac (Thermo, San Jose, CA), and then stored at -80 °C.

In addition to acquiring the soluble protein fraction of targeted GBM cells procured from fresh frozen tissues, cell pellets were treated by a 1% SDS solution [118, 119] containing 20 mM Tris-HCl at pH 8.0, followed by centrifugation at 20,000 g for 30 min. The supernatant containing the membrane protein fraction was placed in a dialysis cup (Pierce, Rockford, IL) and dialyzed overnight at 4 °C against 100 mM Tris-HCl at pH 8.2. The extracted and dialyzed proteins were denatured, reduced, alkylated, digested, desalted, and lyophilized using the same sample preparation protocol as applied to the soluble protein fraction described previously.

Similar to the procedures described in our previous studies [71], the microdissected GBM cells obtained from FFPE tissues were treated with a 20 mM Tris buffer (pH 9) containing 2% SDS, followed by heating at 100 °C on a heat block (VWR Scientific Products, West Chester, PA) for 20 min, then incubation at 60 °C in a incubator (Robbins Scientific, Sunyvale, CA) for 2 hr. The soluble proteins were collected in the supernatant by centrifugation at 20,000 g for 30 min. The supernatant was placed in a dialysis cup (Pierce) and dialyzed overnight at 4 °C against 100 mM Tris-HCl at pH 8.2. The extracted and dialyzed proteins were denatured, reduced, alkylated, digested, desalted, and lyophilized using the same sample preparation protocol as applied to the membrane protein fraction of fresh frozen tissues described previously.

3.2.4 Integrated CIEF/Nano-RPLC Multidimensional Peptide Separations.

On-line integration of CIEF with nano-RPLC as a multidimensional peptide and protein separation platform has been described in detail in previous work [15, 46, 47, 117] and was employed for systematically resolving peptide digests based on their differences in pI and hydrophobicity. Briefly, an 80-cm long CIEF capillary (100 μm i.d./365 μm o.d.) coated with hydroxypropyl cellulose was initially filled with a solution containing 2% ampholyte 3-10 and 1.5 mg/mL tryptic peptides. Peptide focusing was performed by applying electric field strength of 300 V/cm and using solutions of 0.1 M acetic acid and 0.5% ammonium hydroxide as the anolyte and the catholyte, respectively. The current decreased continuously as the result of peptide focusing. Once the current reached $\sim 10\%$ of the original value, usually within 30 min, the focusing was considered to be complete.

Focused peptides were sequentially fractionated by hydrodynamically loading into individual trap columns (3 cm x 200 μm i.d. x 365 μm o.d.) packed with 5 μm porous C_{18} reversed-phase particles. A constant electric field of 300 V/cm was applied across the CIEF capillary for maintaining analyte band focusing in the capillary throughout the loading procedure. Each peptide fraction was subsequently analyzed by nano-RPLC equipped with an Ultimate dual-quaternary pump (Dionex, Sunnyvale, CA) and a dual nano-flow splitter connected to two pulled-tip fused-silica capillaries (50 μm i.d. x 365 μm o.d.). These two 15-cm long capillaries were packed with 3- μm Zorbax Stable Bond (Agilent, Palo Alto, CA) C_{18} particles.

Nano-RPLC separations were performed in parallel in which a dual-quaternary pump delivered two identical 2-hr organic solvent gradients with an offset

of 1 hr. Peptides were eluted at a flow rate of 200 nL/min using a 5-45% linear acetonitrile gradient (containing 0.02% formic acid) over 100 min with the remaining 20 min for column regeneration and equilibration. Full scans were collected from 400 - 1400 m/z using a linear ion-trap mass spectrometer (LTQ, ThermoFinnigan, San Jose, CA) and 5 data dependent MS/MS scans were gathered with dynamic exclusion set to 18 sec. A moving stage housing two nano-RPLC columns was employed to provide electrical contacts for applying electrospray voltages, and most importantly to position the columns in-line with the orifice of the heated metal capillary in the nano-ESI source at the start of each chromatography separation and data acquisition cycle.

3.2.5 Data Analysis.

The OMSSA developed at the National Center for Biotechnology Information [58] was used to search the peak list files against a decoyed SwissProt human database. This decoyed database was constructed by reversing all 12,484 real sequences and appending them to the end of the sequence library. Searches were performed using the following parameters: 1.5 Da precursor ion mass tolerance, 0.4 Da fragment ion mass tolerance, 1 missed cleavage, alkylated Cys as a fixed modification and variable modifications of acetylated N-terminus and Lys, and oxidated Met. Searches were run in parallel on a 12 node, 24 CPU Linux cluster (Linux Networx, Bluffdale, UT).

False positive rates were determined using the method of Elias and co-workers[59] . Briefly, false positive rates were calculated by multiplying the number of false positive identifications (hits to the reversed sequences scoring below a given threshold) by 2 and dividing by the number of total identifications. Peptides identified

below threshold, and also occurring as matches to the forward sequences, were not counted as false positives or true identifications. A curve was then generated by plotting E-value versus false positive rate and an E-value threshold corresponding to a 1% false positive rate was used as the cutoff in this analysis[60]. After generation of search data, the result files were parsed and loaded into a custom MySQL database for visualization and reporting using in-house software.

3.3 RESULTS AND DISCUSSION

In human disease research, where knowledge of disease outcome is critical for the evaluation of the significance of phenotypic or genotypic profiles, as well as response to therapy and outcome, it may take five, ten, or more years to gain a relatively complete picture of the pathophysiology of a disease. The ability to analyze well-characterized, archival cases is highly desirable. In addition, since the capacity to store large numbers of catalogued samples under optimal conditions is limited by cost, space, and personnel limitations, among others, the development of methods to analyze traditional pathological specimens, such as FFPE tissues, is an important priority. However, too often molecular analysis techniques are applied directly to these formalin-paraffin materials, or extracts thereof, without an understanding of the variables introduced by the effects of tissue fixation and processing, whether upon the structure and availability of DNA, RNA, and proteins. Accessibility of macromolecules in the fixed tissue specimens is therefore a critical issue, exemplified by the growth of IHC for protein antigens, and in situ hybridization for DNA and RNA.

Technological innovations already allow RNA profiling of FFPE tissues for the studies of patterns of altered gene expression caused by specific exposures or disease outcomes [120, 121]. Formalin-induced cross-linking of proteins serving as an efficient means of in-situ-preservation of proteins, however, greatly hinders efficient extraction of proteins from tissue sections and subsequent proteomic efforts. Building upon our initial success in the integration of the CIEF-based multidimensional separation platform with the AR technique[71], a selective, laser-free microdissection approach [13, 14] is further incorporated into the workflow in this study to enable comprehensive analysis of protein profiles within targeted tumor cells instead of whole FFPE tissue blocks.

As evaluated by SDS-PAGE (Fig. 3-1), the quality of protein pattern within the soluble fraction of targeted tumor cells procured from fresh frozen GBM tissue was superior to that extracted from microdissected FFPE tissue of the same patient. The smear among FFPE protein bands, particularly in the range of low - medium molecular masses, may be the result of protein fragmentation from the heat-induced AR process. Besides reinstating the condition of a formalin-modified protein back to its original structure, the high-temperature heating treatment has been reported to induce a variety of protein modifications including fragmentation [74-76].

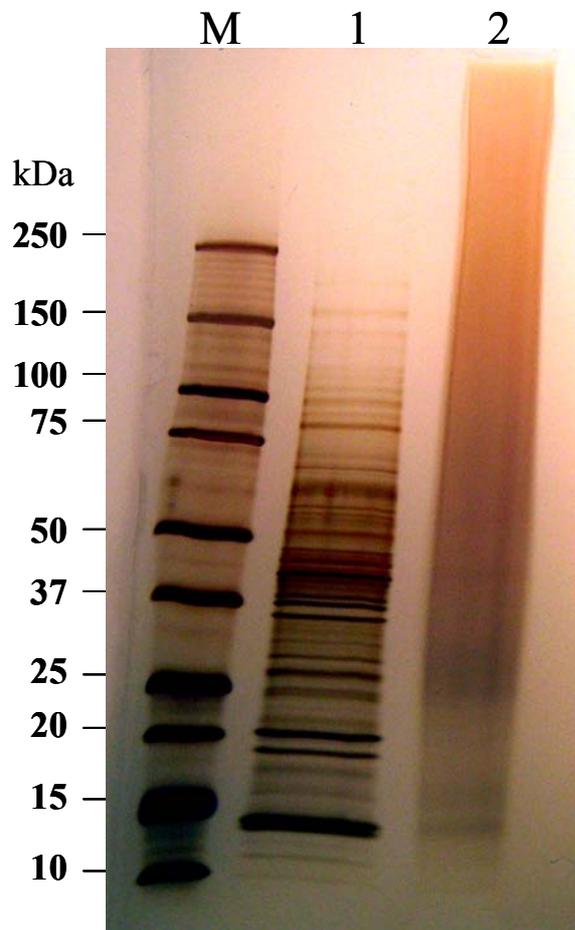


Figure 3-1 Comparison of protein profiles obtained from microdissected fresh frozen (Lane 1) and FFPE (Lane 2) GBM tissue specimens of the same patient using SDS-PAGE.

In addition to the evaluation of intact proteins using SDS-PAGE, combined CIEF/nano-RPLC separations were employed for the examination of protein digests to provide further in depth comparison of proteomes within microdissected fresh frozen and FFPE GBM tissue specimens. For profiling tryptic peptides obtained from a FFPE tissue sample, the entire content of focused peptides in the CIEF capillary was split into 19 individual fractions (Fig. 3-2) which were further resolved by nano-RPLC and identified using nano-ESI-LTQ-MS/MS. The number of distinct peptide identifications measured from each CIEF fraction is significantly greater than those typically reported in the literature using other IEF techniques including immobilized pH gradient gels [52, 53] and gel-free approaches [49, 122-124] such as chromatofocusing, immobilized pH membranes, Rotofor, and free-flow electrophoresis. A key feature of our CIEF-based multidimensional separation technology is the elimination of protein/peptide loss and dilution in an integrated platform while achieving comprehensive and ultrasensitive analysis of protein expression profiles within FFPE tissue specimens. By contrast, preparative-scale IEF techniques [49, 52, 53, 122-124] are incompatible with the smaller scale of the more selectively-procured proteomes obtained from microdissection-procured tissues.

As shown in Fig. 3-3, the peptide and protein false positive rates, and the numbers of total peptides, distinct peptides, and protein identifications were plotted as functions of the E-value of a typical OMSSA search. An E-value threshold of 0.17, corresponding to 1% false positive of total peptide identifications, was chosen as a cutoff in this study. A total of 14,748 distinct peptides were identified, leading to the identification of 2,733 non-redundant proteins from the SwissProt human database

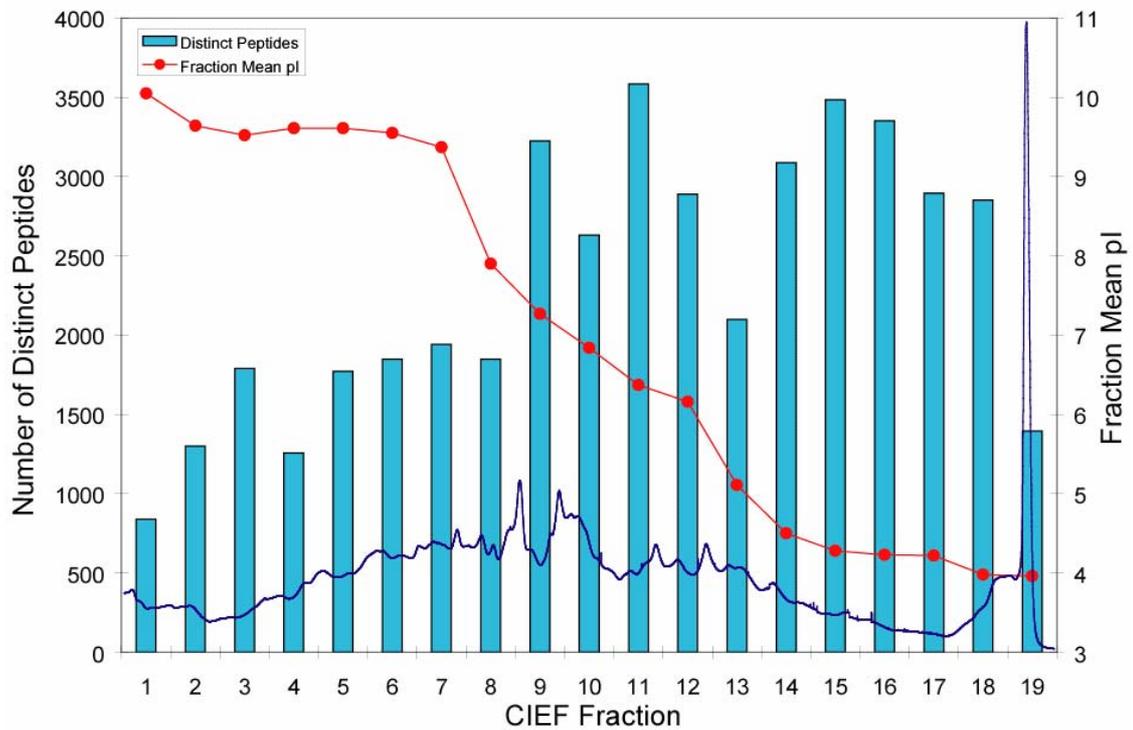


Figure 3-2 Overlaid plots containing the CIEF-UV trace monitored at 280 nm, the number of distinct peptides identified in each of the CIEF fractions, and the distribution of the peptide's mean pI values over the entire CIEF separation.

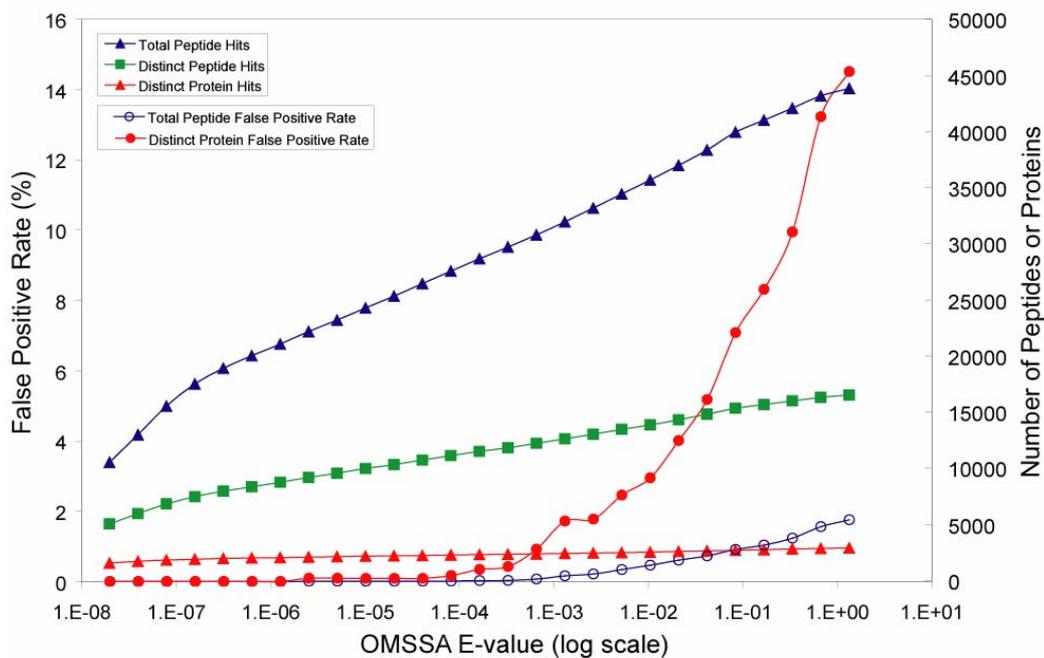


Figure 3-3 Plots of the false positive rates and the numbers of total peptide, distinct peptide, and distinct protein identifications versus the E-value obtained from the search of the peak list files against a decoyed SwissProt human database using OMSSA.

containing 12,484 non-redundant protein entries. This identity threshold also resulted in a protein false positive rate of 7.5% as indicated by the detection of peptides from 107 distinct reversed protein sequences in the decoy section of the search database.

The first reversed protein was detected at an E-value of 2×10^{-6} . At this threshold score, a total of 10,755 distinct peptides were identified, leading to the identification of 2,224 non-redundant proteins (Fig. 3-3). By tolerating at 1% false positive of total peptide identifications (E-value threshold of 0.17), additional 3,993 distinct peptides and 509 distinct proteins were measured at a cost of 112 and 107 predicted false identifications of distinct peptides and proteins, respectively. By further increasing the E-value threshold to 1.3, the false positive rates escalated to 1.8% and 15.0% for total peptide and protein identifications, respectively.

To better illustrate the impact of protein false positive rate on protein identification, it should be emphasized that new distinct proteins were added to search results at a ratio of approximately 50:1 relative to reversed distinct proteins at an E-value of 2×10^{-6} . At an E-value of 0.17, corresponding to a protein false positive rate of 7.5%, this ratio decreased to 2:1. This ratio was further reduced to 1:1 at an E-value of 1.3, meaning that new forward protein sequences were added at a rate equal to that of reversed protein sequences. The implication is that all new forward protein sequences were likely false positives at an E-value of 1.3 which corresponded to the false positive rates of 1.8% and 15.0% for total peptide and protein identifications, respectively.

Compared to several recently reported FFPE tissue-based proteome studies [45, 72, 73], our results present the largest catalog of proteins from a single

microdissected FFPE tissue specimen reported to date. Application of only a single-dimension separation in these recent studies[45, 72, 73] due to sample amount constraints has significantly limited the dynamic range and detection sensitivity of MS measurements, and greatly impacted their ability to mine deeper into the tissue proteome. In addition to the comprehensiveness of our tissue proteome analysis, the percentage of overlapping proteins obtained from repeated runs of the same FFPE GBM tissue sample was greater than 87% as illustrated by the Venn diagram shown in Fig. 3-4. A total of 2,845 non-redundant proteins were identified from combined proteome runs of a single FFPE tissue sample.

Among proteins identified from the microdissection-procured FFPE GBM tissue specimen, 488 proteins were predicted to contain at least one or more transmembrane domains using TMHMM (www.cbs.dtu.dk/services/TMHMM-2.0/)[125]. The subcellular location of proteins identified from the FFPE tissue was assigned using the protein subcellular localization prediction tool (PSLT)[126]. Still, the subcellular location of approximately half of the identified proteins can not be assigned by PSLT. Among proteins with assigned locations, 34% were found in plasma membrane and organelle categories (Fig. 3-5).

The sequence coverage of two representative transmembrane proteins, tenascin and basigin, is presented in Fig. 3-6 together with the examples of peptides' tandem mass spectra leading to their identifications. Tenascin, a glioma-associated-extracellular matrix antigen, is a substrate-adhesion molecule that appears to inhibit cell migration and may play a role in supporting the growth of epithelial tumors. Tenascin is also a ligand for integrins of α -8/ β -1, α -9/ β -1, α -V/ β -3, and α -V/ β -6 [127].

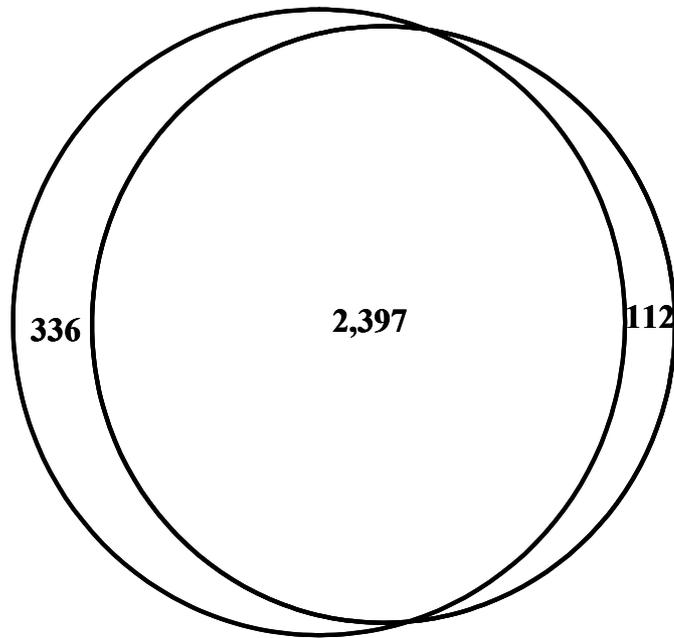


Figure3-4 The overlap in the proteins identified from repeated analyses using a single FFPE GBM tissue sample.

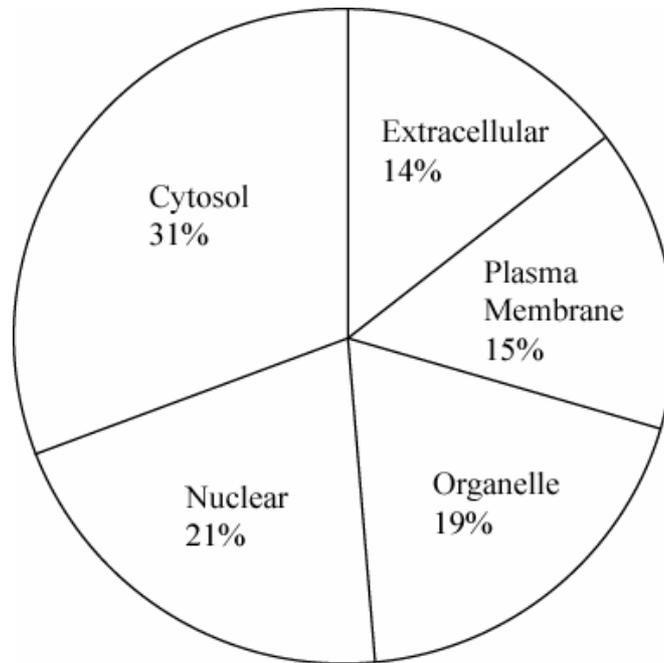
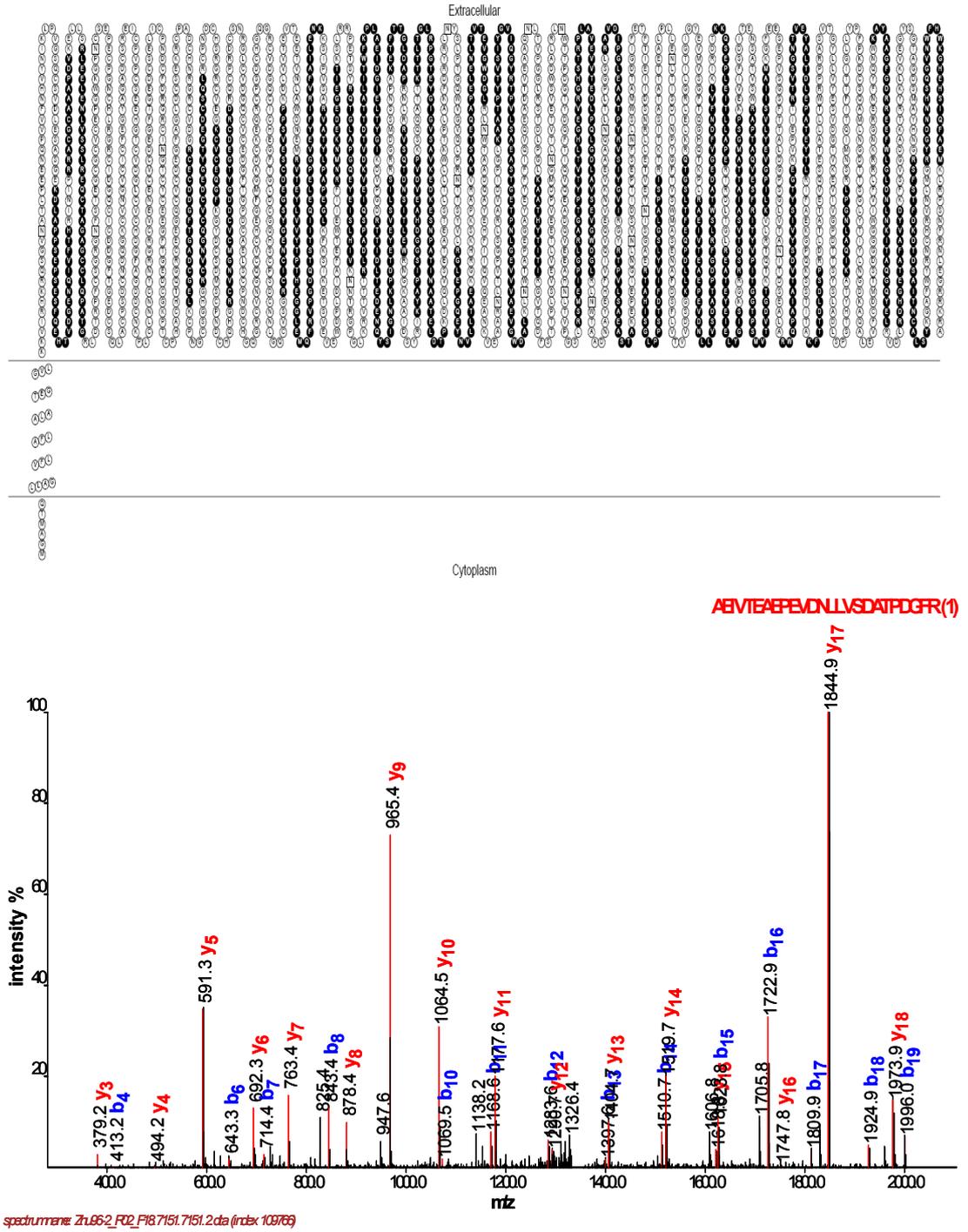
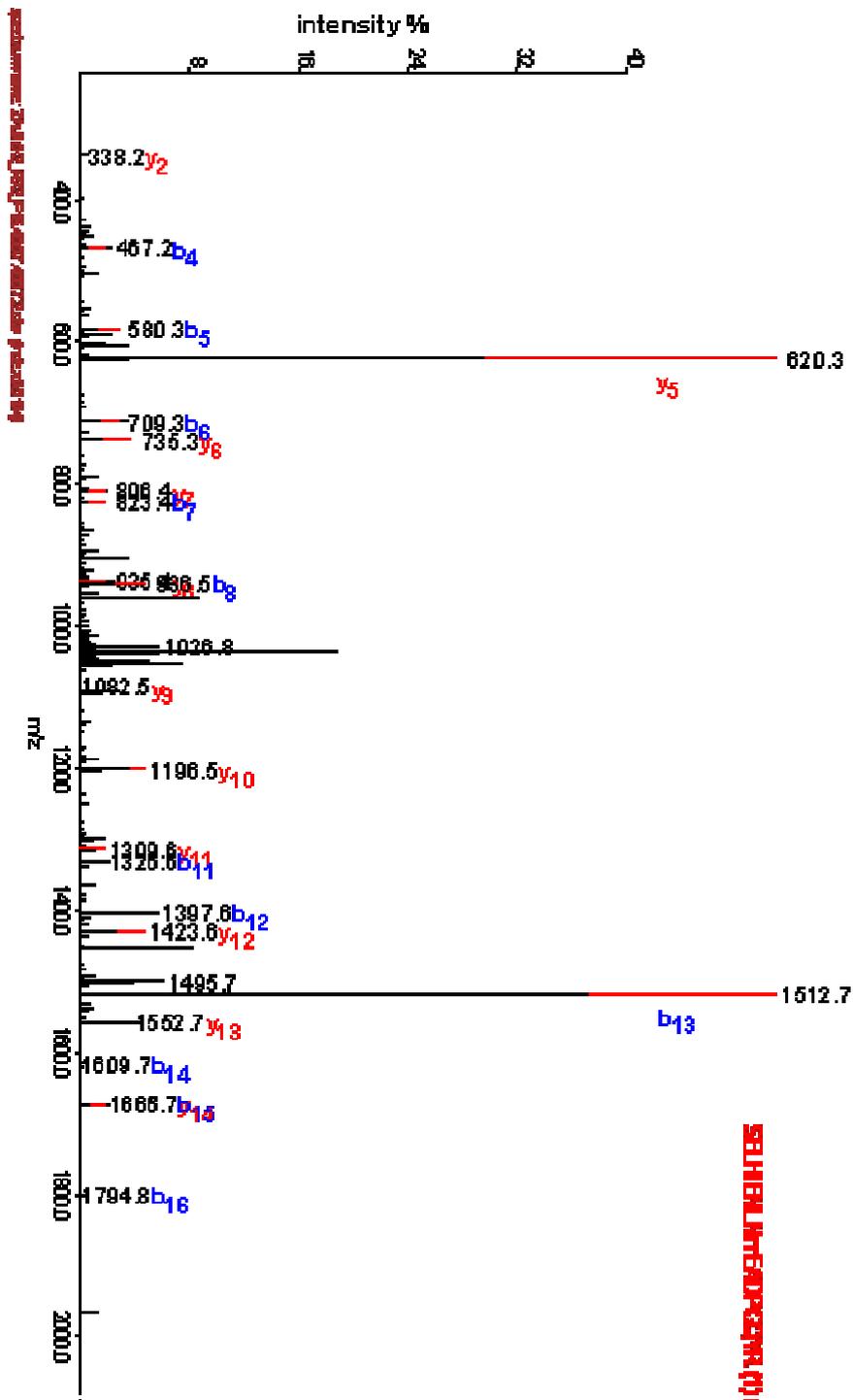


Figure 3-5 Distribution of PSLT-predicted subcellular localization of proteins identified from the microdissection-procured FFPE GBM tissue specimen.

A



B



B

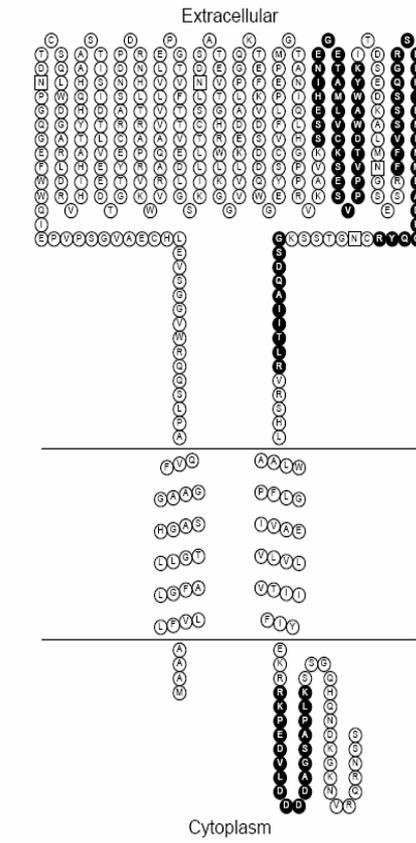


Figure 3-6 Peptide coverage of representative transmembrane proteins such as (A) tenasin and (B) basigin, and tandem MS spectra of unique peptides leading to their identifications.

Basigin, a tumor cell-derived collagenase stimulatory factor, is enriched on the surface of tumor cells and up-regulated in gliomas. Its expression level correlates with malignant potential of the tumor[128]. For its tissue specificity, basigin is only present in vascular endothelium in non-neoplastic regions of the brain, whereas it is present in tumor cells but not in proliferating blood vessels in malignant gliomas.

In addition to profiling the FFPE tissue proteome, combined CIEF/nano-RPLC separations coupled with nano-ESI-LTQ-MS/MS were employed in the analysis of protein digests obtained from the soluble and cell pellet fractions of microdissected fresh frozen tissue. Fresh tissue taken from the same case of GBM was microdissected and processed for the extraction of soluble proteins using urea, followed by a SDS-based protocol [118, 119] for the preparation of membrane proteins from remaining cell pellets. By using a 1% false positive rate for total peptide identifications, a total of 2,856 and 3,227 proteins were identified from the soluble and pellet fractions, respectively. By combining the proteome results obtained from the soluble and pellet fractions, the collective analysis yielded the identification of 3,902 non-redundant proteins with an average of 6.2 peptides per protein, corresponding to 31% coverage of the SwissProt human database.

Comparing the proteome results obtained from the fresh frozen and FFPE tissues (Fig. 3-7), most proteins identified from the FFPE slide were also detected in the corresponding fresh frozen section. Only 243 proteins, representing 8.5% of total protein identifications, were unique to the FFPE tissue. Among proteins identified from the FFPE tissue, 2,370 proteins or 83% of the total protein identifications, overlapped with those measured from the pellet fraction of fresh frozen tissue.

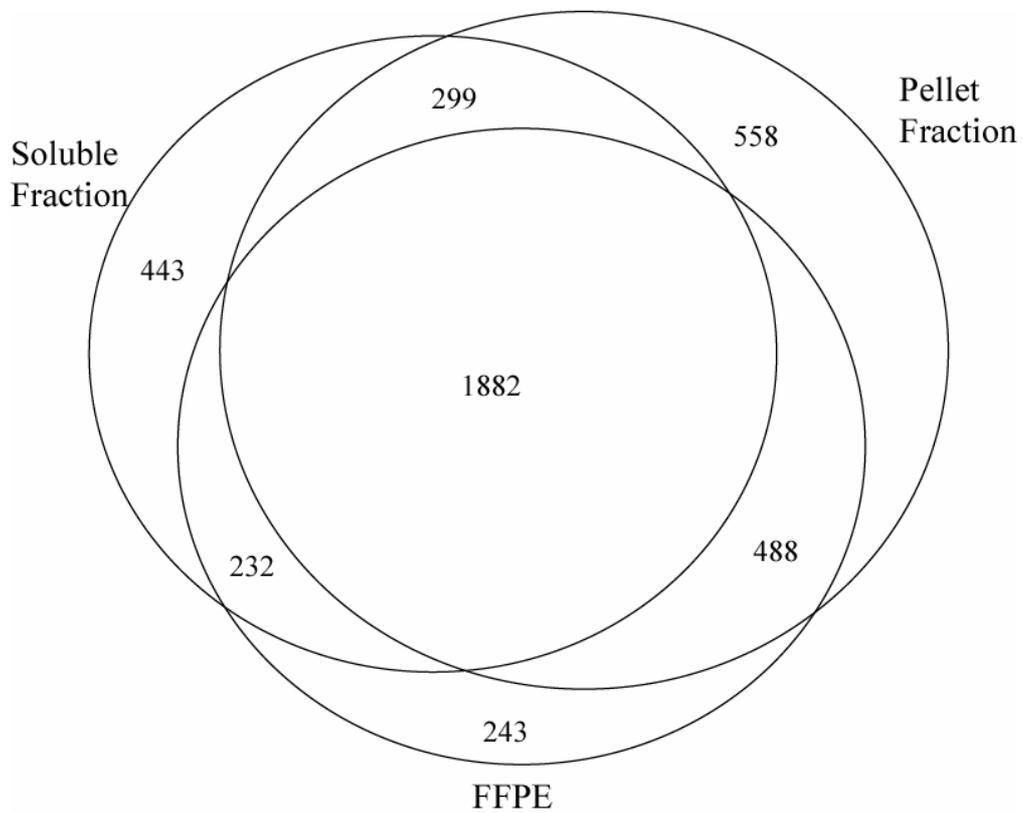


Figure 3-7 The overlap in the proteins identified from microdissected fresh frozen (the soluble and pellet fractions) and FFPE GBM tissue specimens of the same patient using combined CIEF/nano-RPLC separations coupled with nano-ESI-LTQ-MS/MS.

The percentage of overlap among membrane proteins (predicted to contain at least one or more transmembrane domains) identified from the FFPE and the pellet fraction of fresh frozen tissues was approximately 83% and was further increased to 88% by including membrane proteins measured from the soluble fraction (Fig. 3-8). We attribute our success with this high concordance in protein identification between FFPE and fresh-frozen GBM tissues to the combined effect of SDS extraction and heat-induced AR in concert with the exceptionally sensitive CIEF/nano-RPLC separations coupled with ESI-LTQ-MS/MS proteome strategy.

3.4 CONCLUSION

Based on the heat-induced and SDS-based AR technique, minute proteins extracted from the microdissection-procured FFPE GBM tissue specimen were processed and analyzed using combined CIEF/nano-RPLC separations coupled with nano-ESI-LTQ-MS/MS. By using a decoyed database search approach [59, 60], an E-value threshold of 0.17, corresponding to 1% false positive of total peptide identifications, was chosen as a cutoff in this study. A total of 14,478 distinct peptides were therefore detected, leading to the identification of 2,733 non-redundant SwissProt protein entries from a single proteome analysis of microdissected FFPE tissue.

Due to the ability of the CIEF-based multidimensional separation platform for achieving ultrahigh resolution of minute protein digests, our results present the largest catalog of proteins from a single microdissected FFPE tissue specimen reported to date.

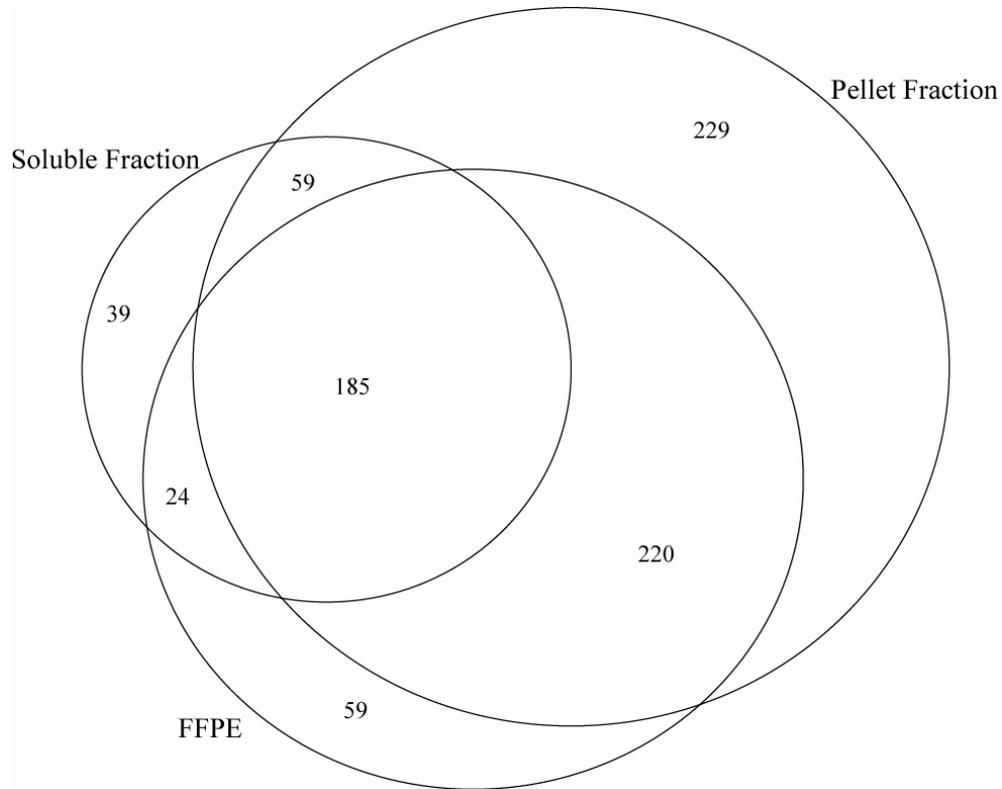


Figure 3-8 The overlap in the proteins identified from microdissected fresh frozen (the soluble and pellet fractions) and FFPE GBM tissue specimens predicted to contain at least one or more transmembrane domains using TMHMM (www.cbs.dtu.dk/services/TMHMM-2.0/)[125]

By comparing with those FFPE tissue proteome studies reported recently [45, 72, 73], the application of only a single-dimension separation due to sample amount constraints has significantly limited their capability to mine deeper into the tissue proteome. In addition to the large proteome coverage, the reproducibility of our protein identifications was greater than 87% by comparing proteins identified from repeated runs of the same GBM tissue sample.

A total of 2,370 FFPE tissue proteins or 83% of total protein identifications overlapped with those measured from the pellet fraction of fresh frozen GBM tissue of the same patient. This large degree of protein overlapping is the result of SDS extraction employed in both the cell pellet preparation protocol and the AR technique. The presence of SDS was critical in several AR protocols for achieving satisfactory protein extractions from FFPE tissue sections followed by IHC and SDS-PAGE analysis [66, 115, 116]. Instead of coupling high temperature heating with SDS treatment in a single step, a two-phase approach may be applied for protein extraction from the FFPE tissue in our future studies. The first phase intends to break down formalin-induced crosslinkings within proteins by heating. The second phase then involves the use of solubilization reagents such as urea and SDS for the extraction of the heat-retrieved proteins from the FFPE tissue.

Following the first medical school course in histology at Edinburgh in 1842, the diagnosis of cancer has come to be based upon the microscopic appearances of tissues (histopathology = morphologic phenotype). However, the last several decades have seen an accumulation of data describing the characteristics of cancer cells at a protein and nucleic acid level (molecular phenotype). Thus, advanced proteome

technologies employed and demonstrated in this study not only allow the rigorous evaluation of the quality and the reproducibility of proteins extracted from FFPE tissues for further optimization of AR methodology, but also provide significant opportunities in the pursuit of biomarker discovery using archived FFPE tissue collections. Additionally, the combination of histologic criteria with advanced molecular analysis techniques will permit pathologists and biomedical investigators to incorporate alterations at the molecular level into the histological diagnosis of cancer, advancing the rapidly growing field of “molecular morphology”.

3.5 ACKNOWLEDGEMENT

We thank the National Cancer Institute (CA103086 and CA107988) and the National Center for Research Resources (RR021239 and RR021862) for supporting portions of this research. This research was also supported in part by the Melvin Burkhardt chair in neurosurgical oncology and the Karen Colina Wilson research endowment fund within the Brain Tumor Institute at the Cleveland Clinic Foundation.

CHAPTER FOUR

EVALUATION OF ARCHIVAL TIME ON SHOTGUN PROTEOMICS OF FORMALIN-FIXED AND PARAFFIN- EMBEDDED TISSUES

Submitted for publication and reproduced with permission from J. Prot. Res. Unpublished work copyright 2008 American Chemical Society. Authors: Tong Guo, Kejia Zhao, Fattaneh A. Tavassoli, Cheng S. Lee, and Brian M. Balgley

4.1 INTRODUCTION

The criteria employed by pathologists for the diagnosis of numerous diseases, including essentially all cancers, have been established in formalin-fixed and paraffin-embedded (FFPE) tissue sections stained by hematoxylin and eosin (H&E). Furthermore, FFPE specimens, such as those collected under clinical trials, present invaluable resources for conducting retrospective investigations on molecular determinants associated with therapeutic response. Technological innovations already allow RNA profiling of FFPE sections for assessment of patterns of altered gene expression caused by specific exposures or disease outcomes[120, 129]. However, formalin-induced cross-linking of protein hinders efficient extraction of proteins from FFPE tissues for performing subsequent immunohistochemistry (IHC) and proteomic measurements.

Boiling the FFPE tissue sections in buffer solutions dramatically reduced the detection thresholds of IHC staining for a wide range of antibodies in the antigen retrieval methodology[66-68, 114]. To increase the amounts of protein extractable from FFPE sections, the heat-induced antigen retrieval approach was combined with the application of a radioimmunoprecipitation buffer containing sodium dodecyl sulfate (SDS)[115] or a denaturing solution containing both urea and SDS[130]. A commercial Liquid Tissue™ kit was introduced for processing FFPE tissues also in the presence of heating at 95 °C for 90 min[70]. Two recent FFPE proteome studies were based on the use of lysis buffers containing either guanidine hydrochloride[131] or organic solvent[132] at high temperature.

Combined capillary isoelectric focusing (CIEF)/nano-reversed phase liquid chromatography (nano-RPLC) separations coupled with electrospray ionization-mass spectrometry (ESI-MS) have been demonstrated in our laboratory to enable ultrasensitive analysis of minute proteins extracted from whole and microdissected FFPE tissues[57, 71]. A capillary isotachopheresis (CITP)-based proteome platform, capable of providing selective analyte enrichment and high resolving power, was utilized in a recent study to further address the challenges of protein complexity and relative abundance inherent in FFPE specimens[133]. From a practical point of view, one of the remaining issues in performing comparative proteomic measurements among FFPE tissues relates to potential variability in protein composition and retrieval during different storage periods. Thus, the CITP proteome technology coupled with the spectral counting approach[63-65] is employed in this work for the first time to investigate the effects of archival time on protein expression profiles over

a set of mesenchymal tumors, including nine leiomyomas dating from 1990 to 2002 and a single vaginal alveolar soft part sarcoma (ASPS) from 1980.

4.2 EXPERIMENTAL SECTION

4.2.1 Materials and Reagents.

Fused-silica capillaries (50 μm i.d./375 μm o.d. and 100 μm i.d./375 μm o.d.) were acquired from Polymicro Technologies (Phoenix, AZ). Acetic acid, dithiothreitol (DTT), formalin, iodoacetamide (IAM), and octane were obtained from Sigma (St. Louis, MO). Acetonitrile, ammonium acetate, hematoxylin, SDS, tris(hydroxymethyl)aminomethane (Tris), and urea were purchased from Fisher Scientific (Pittsburgh, PA). Pharmalyte 3-10 was acquired from Amersham Pharmacia Biotech (Uppsala, Sweden). Sequencing grade trypsin was obtained from Promega (Madison, WI). All solutions were prepared using water purified by a Neutron Ion system (Baltimore, MD) equipped with a UV sterilizing lamp and a 0.05 μm membrane final filter.

4.2.2 Tissue Sample Preparation.

A blinded set of ten mesenchymal tumors were obtained from Dr. Fattaneh Tavassoli's laboratory at the Yale University School of Medicine. FFPE tissues were de-paraffinized using octane, followed by vortexing and centrifugation[71]. Proteins were recovered from tissue sections of 1 cm x 1 cm x 10 μm by following heat-induced retrieval conditions described previously[57, 71, 133]. Briefly, sections of fixed tissues were treated with a 20 mM Tris buffer containing 2% SDS, followed by heating at 100 $^{\circ}\text{C}$ on a heat block (VWR Scientific Products, West Chester, PA) for

20 min, then incubation at 60 °C in an incubator (Robbins Scientific, Sunyvale, CA) for 2 hr. The proteins collected in the supernatant were placed in a dialysis cup (Pierce, Rockford, IL) and dialyzed overnight at 4 °C against 100 mM Tris-HCl at pH 8.2.

The retrieved and dialyzed proteins were denatured, reduced, and alkylated by sequentially adding urea, dithiothreitol, and iodoacetamide with final concentrations of 8 M, 10 mg/mL, and 20 mg/mL, respectively. The solution was incubated at 37 °C for 1 hr in the dark and then diluted 8-fold with 100 mM ammonium acetate at pH 8.0. Trypsin was added at a 1:40 (w/w) enzyme-to-substrate ratio, and the solution was incubated at 37 °C overnight. Tryptic digests were desalted using a Peptide MacroTrap column (Michrom Bioresources, Auburn, CA), lyophilized to dryness using a SpeedVac (Thermo, San Jose, CA), and then stored at -80 °C.

4.2.3 Transient CITP/Capillary Zone Electrophoresis (CZE)-Based Tissue Proteome Analysis.

Similar to the procedures described in our previous studies[133-135], a 80-cm long CITP capillary (100 µm i.d./365 µm o.d.) was initially filled with a background electrophoresis buffer of 0.1 M acetic acid at pH 2.8. The sample containing protein digests retrieved and processed from FFPE tumor tissues was prepared in a 2% pharmalyte solution. A 50-cm long sample plug, corresponding to 4.0 µL sample volume, was hydrodynamically injected into the capillary. A positive electric voltage of 24 kV was then applied to the inlet reservoir, which was filled with a 0.1 M acetic acid solution.

The cathodic end of the capillary was housed inside a stainless steel tube (460 μm i.d./785 μm o.d.) using a coaxial liquid sheath flow configuration[136]. A sheath liquid composed of 0.1 M acetic acid was delivered at a flow rate of 1 $\mu\text{L}/\text{min}$ using a Harvard Apparatus 22 syringe pump (South Natick, MA). The stacked and resolved peptides in the CITP/CZE capillary were sequentially fractionated and loaded into individual wells on a moving microtiter plate. The entire capillary content was separated and sampled into 30 individual fractions in less than 2 hr. The CITP separation and fractionation apparatus was operated on an in-house built robotic platform controlled by LabView (National Instruments, Austin, TX).

To couple transient CITP/CZE with nano-RPLC, peptides collected in individual wells were sequentially injected into dual trap columns (3 cm x 200 μm i.d. x 365 μm o.d.) packed with 5 μm porous C_{18} reversed-phase particles. Each peptide fraction was subsequently analyzed by nano-RPLC equipped with an Ultimate dual-quaternary pump (Dionex, Sunnyvale, CA) and a dual nano-flow splitter connected to two pulled-tip fused-silica capillaries (50 μm i.d. x 365 μm o.d.). These two 15-cm long capillaries were packed with 3- μm Zorbax Stable Bond (Agilent, Palo Alto, CA) C_{18} particles.

Nano-RPLC separations were performed in parallel in which a dual-quaternary pump delivered two identical 2-hr organic solvent gradients with an offset of 1 hr. Peptides were eluted at a flow rate of 200 nL/min using a 5-45% linear acetonitrile gradient over 100 min with the remaining 20 min for column regeneration and equilibration. The peptide eluants were monitored using a linear ion-trap mass spectrometer (LTQ, ThermoFinnigan, San Jose, CA) equipped with an electrospray

ionization interface and operated in a data dependent mode. Full scans were collected from 400 - 1400 m/z and 5 data dependent MS/MS scans were collected with dynamic exclusion set to 30 sec.

4.2.4 MS Data Analysis.

Raw LTQ data were converted to peak list files by `msn_extract.exe` (ThermoFinnigan). Open Mass Spectrometry Search Algorithm (OMSSA)[58] developed at the National Center for Biotechnology Information was used to search the peak list files against the UniProt sequence library (April 20, 2006) with decoyed sequences appended. This decoyed database was constructed by reversing all sequences and appending them to the end of the sequence library. Searches were performed using the following parameters: fully tryptic, 1.5 Da precursor ion mass tolerance, 0.4 Da fragment ion mass tolerance, 1 missed cleavage, alkylated Cys as a fixed modification and variable modification of Met oxidation. Searches were run in parallel on a 12 node, 24 CPU Linux cluster (Linux Networx, Bluffdale, UT).

False discovery rates (FDRs) were determined using a target-decoy search strategy introduced by Elias and co-workers[59] and employed in our previous study for a comparative evaluation among commonly used tandem MS identification search algorithms[137]. Briefly, FDRs were calculated by multiplying the number of false positive identifications (hits to the reversed sequences scoring below a given threshold) by 2 and dividing by the number of total identifications. Peptides identified below threshold, and also occurring as matches to the forward sequences, were not counted as false positives or true identifications. A curve was then generated by plotting E-value versus FDR and an E-value threshold corresponding to a 1% FDR

for total peptide identifications was chosen as a cutoff in this study. The UniProt sequence library consists of entries from both SwissProt and TrEMBL. Only peptide hits mapping to the SwissProt subset were reported.

4.2.5 Validation of Proteins Retrieved and Identified from FFPE Tissues Using IHC

Three proteins, including actin, desmin, and progesterone receptor, were chosen from MS-based proteomic measurements of leiomyomas for validation using IHC. Vacuolar proton translocating ATPase 116 kDa subunit isoform a3 (VPP3) as one of candidate markers whose expression profiles distinguish ASPS from leiomyoma was selected for validation using IHC. Primary antibodies against actin, desmin, and progesterone receptor were obtained from Dako (Carpinteria, CA). The mouse primary antibody against VPP3 was purchased from Abcam (Cambridge, MA) and incubated with a dilution ratio of 1:150 at 4 °C overnight.

The biotinylated anti-mouse immunoglobulins were employed as the link antibody and were incubated for 30 min. The avidin-biotin-peroxidase complex (Vector Laboratories, Burlingame, CA) was utilized as the labeling reagent and incubated for 30 min. A 10-min wash step with Tris buffered saline (TBS) was carried out between each step. Slides were counterstained by hematoxylin and mounted for examination. Negative controls were conducted using TBS instead of the primary antibody. Immunostained slides were evaluated by microscopy. The intensity of immunostaining was graded as + and – for positive and negative, respectively.

4.3 RESULTS AND DISCUSSION

Because the capacity to store large numbers of catalogued clinical fresh frozen samples under optimal conditions is limited by cost, space, and personnel limitations, it is important to explore the utility of a large number of archival FFPE tissue banks available worldwide for extraction of molecular data using novel technologies. These tissue banks particularly constitute invaluable resources for translational studies of cancer. In cancer research, where knowledge of disease outcome is critical in the evaluation of the significance of particular phenotypes or genotypes, and where it may take five or more years to complete expensive prospective studies, the ability to analyze documented archival cancer cases with known outcome is highly desirable.

From the onset, it was apparent that proper sample preparation is critical in application of IHC to FFPE tissues[138, 139]. Shi et al.[66, 140] clearly illustrated that hydrolysis of cross-linkages between formalin and protein can be disrupted by high temperature heating. As demonstrated in our previous studies[57, 71, 133], the ability to achieve excellent protein recovery from FFPE tissues for shotgun proteome analysis was attributed to the use of both heating and use of detergents such as SDS. In fact, Fowler et al.[141] have compared 6 different protocols including commercial reagents such as Liquid TissueTM[70] for protein extraction from FFPE tissues, and concluded that the most effective protein extraction buffer tested is a 20 mM Tris solution containing 2% SDS under high temperature heating protocol[57, 71, 133].

Following optimized protein extraction and digestion procedures for handling FFPE tissues[57, 71, 133], ten archived and blinded mesenchymal tumor

tissue blocks, including nine uterine leiomyomas and a single vaginal ASPS, were employed for performing shotgun-based tissue proteome analysis. The leiomyomas were categorized into three groups of three cases each by the archival year of 1990, 1997, and 2002. The sarcoma case was collected from 1980. It was unknown to the analyzers that this tissue was a significantly different type of mesenchymal tumor - an ASPS in contrast to the benign nature of the leiomyomas that are of smooth muscle derivation. As shown in Fig. 4-1, the morphology revealed by H&E staining is markedly different between leiomyoma and sarcoma.

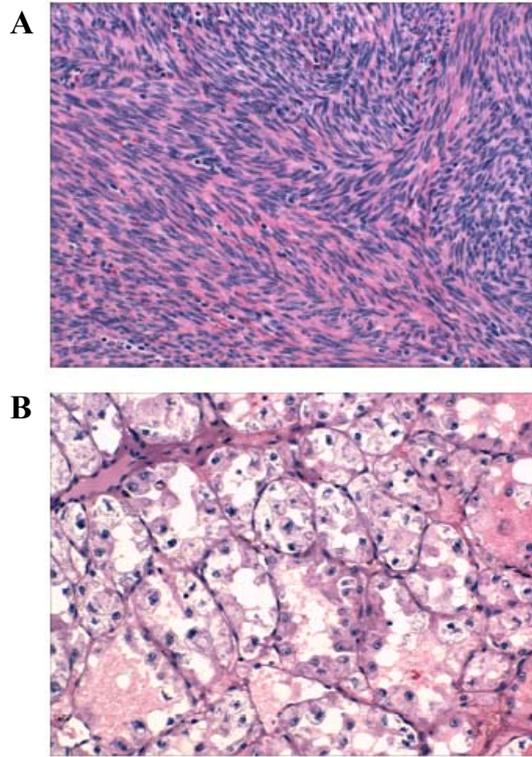


Figure 4-1 H&E staining of (A) uterine leiomyoma and (B) vaginal ASPS.

The proteomic results obtained from the CITP-based proteome platform were summarized in Table 4-1 for all ten mesenchymal tumor tissues examined in this study. The 1990 group of leiomyomas returned slightly poorer proteome performance compared to the leiomyomas cataloged in 1997 and 2002. There was no discernable difference between the 1997 and 2002 groups. The 1980 ASPS case yielded over 2,400 protein identifications which were near the upper end of what were detected for all 10 tumors, despite being the oldest FFPE tissue sample.

By including proteins identified by at least two distinct (different) peptide sequences, the Venn diagram shown in Fig. 4-2 illustrates the overlap among proteins identified in the sarcoma (ASPS) and leiomyomas. It is not surprising that, of the 2,583 proteins identified, 653 were found uniquely in the leiomyomas. This may be simply the result of the larger number of leiomyoma samples compared to a single example of sarcoma analyzed. The 80 proteins identified uniquely in the sarcoma case are significant, however, exactly because they were not discovered in the leiomyomas despite having evaluated nine different tumors. For example, VPP3 as one of unique proteins expressed in the ASPS was identified by three distinct peptides and validated by IHC (Fig. 4-3).

Table 4-1 Summary of Proteomic Results Obtained from Mesenchymal Tumor
Tissues

Tumor	Year	Spectral Counts	Distinct Peptides	Distinct Proteins
Sarcoma	1980	18,254	8,994	2,418
Leiomyoma	1990	15,193	5,887	1,714
Leiomyoma	1990	12,556	5,599	1,593
Leiomyoma	1990	12,976	5,064	1,707
Leiomyoma	1997	22,267	7,897	2,138
Leiomyoma	1997	34,810	12,413	2,671
Leiomyoma	1997	26,235	9,259	2,227
Leiomyoma	2002	41,711	10,020	2,406
Leiomyoma	2002	18,134	6,852	1,977
Leiomyoma	2002	29,335	10,961	2,825

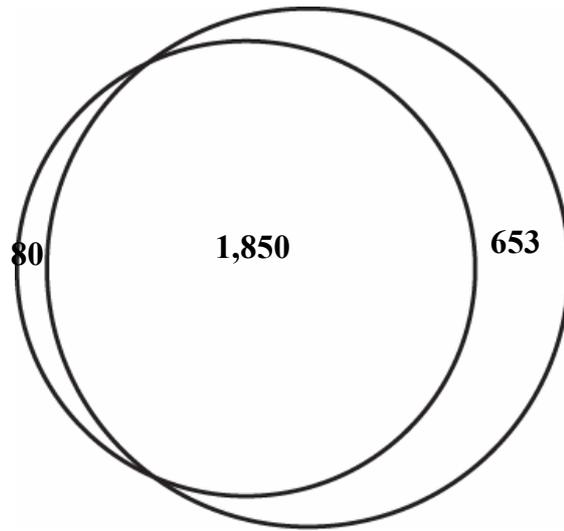
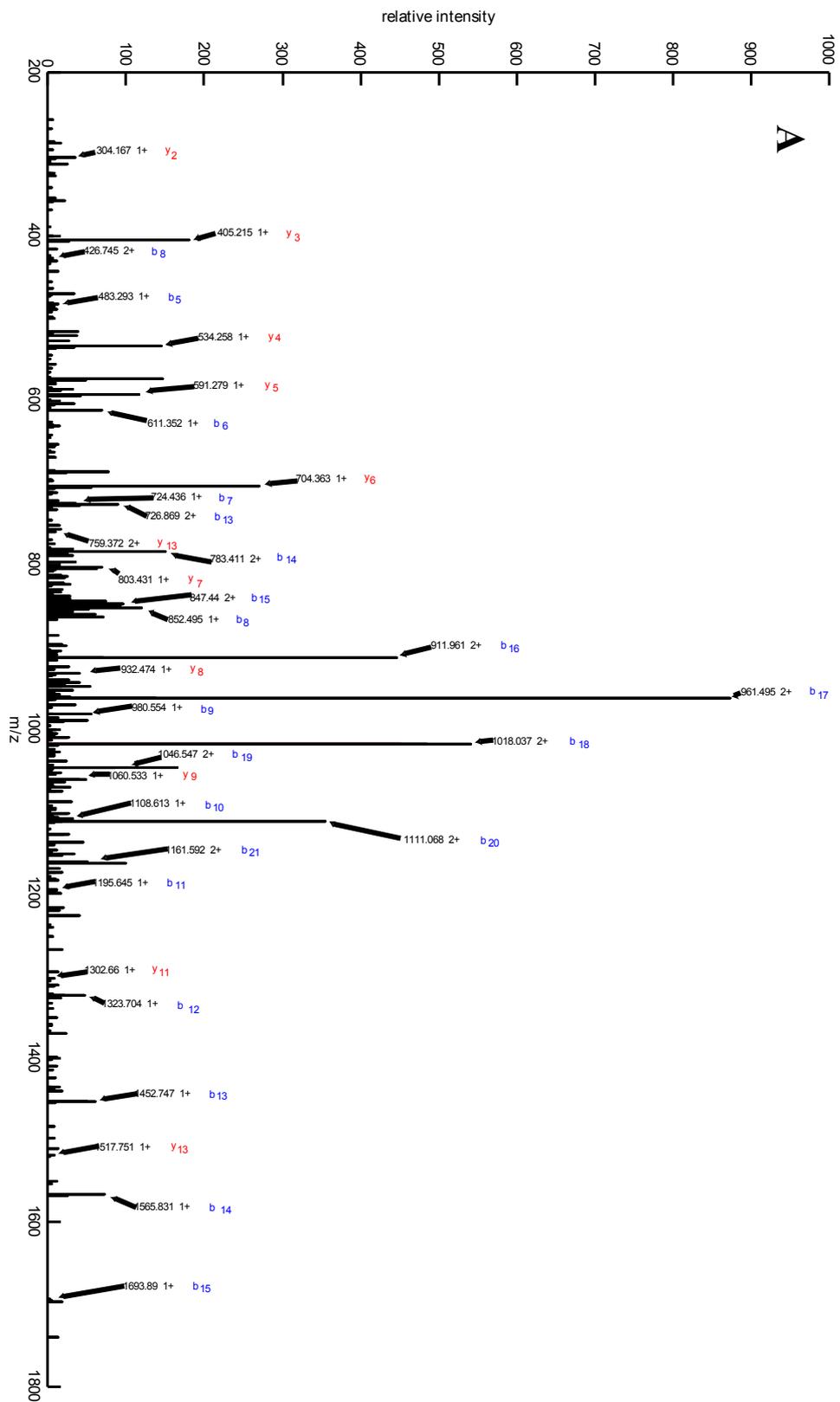


Figure 4-2 The Venn diagram comparing proteins identified from the sarcoma (small cycle) and leiomyomas (large cycle).



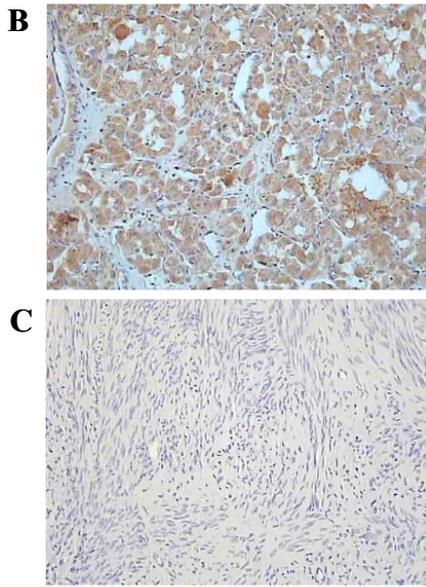


Figure 4-3 (A) Tandem MS spectrum of unique peptide LGALQQLQQSQELQEVLGETER leading to the identification of VPP3. IHC staining of VPP3 on (B) ASPS and (C) leiomyoma tissue sections.

Besides comparing identified proteins, the CITP-based proteome platform was coupled with the spectral counting approach[63-65] to quantify changes in protein expression between the ASPS and leiomyomas. The spectral counts of individual proteins were normalized against the total spectral counts of that run over the normalized total from all runs considered and termed expression values. All proteins with spectral counts > 1 were included in the determination of the Pearson correlation coefficient for the assessment of proteomic data quality and reproducibility.

Great correlation among quantitative proteome results obtained from leiomyomas was achieved in this study. For example, correlation of any of the three leiomyoma groups (archived in 1990, 1997, and 2002) with any of the other two sets yielded a Pearson R^2 value of greater than 0.97 (Fig. 4-4A). Despite correlating among different patient samples, this value compares favorably with a recent report of a Pearson R^2 correlation of 0.88 for performing technical replicates of the same protein digest using a multidimensional liquid chromatography system[142].

In contrast, the correlation between the vaginal sarcoma (ASPS) and the uterine leiomyomas was moderate, with a Pearson R^2 value of 0.53 (Fig. 4-4B). Despite sharing over 1,800 common proteins in a core set (Fig. 4-2), significant differences in protein expression were clearly observed between the two tumor types. Based on spectral counts of identified proteins, categorization of tumor specimens was achieved by an unsupervised hierarchical cluster analysis[143]. As shown in Fig. 4-4, the single sarcoma case was well-separated from the nine leiomyomas.

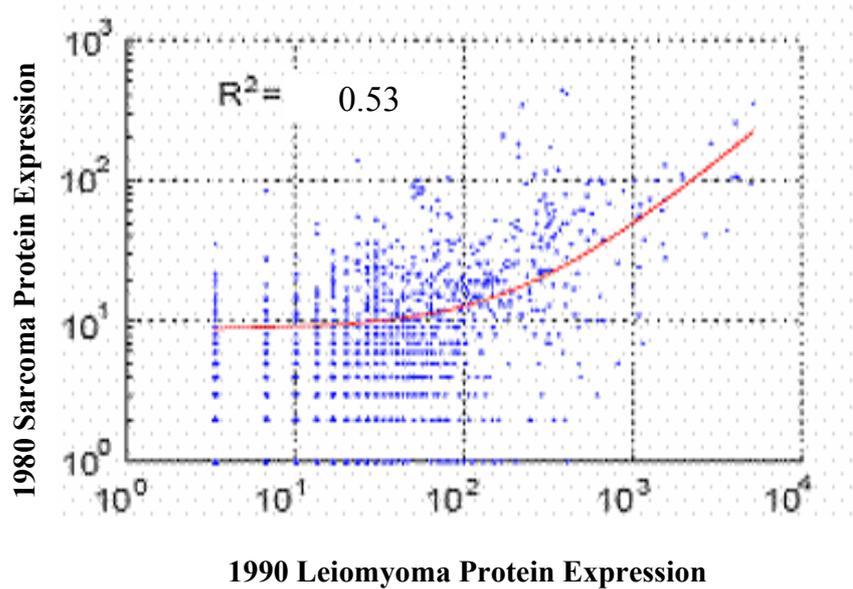
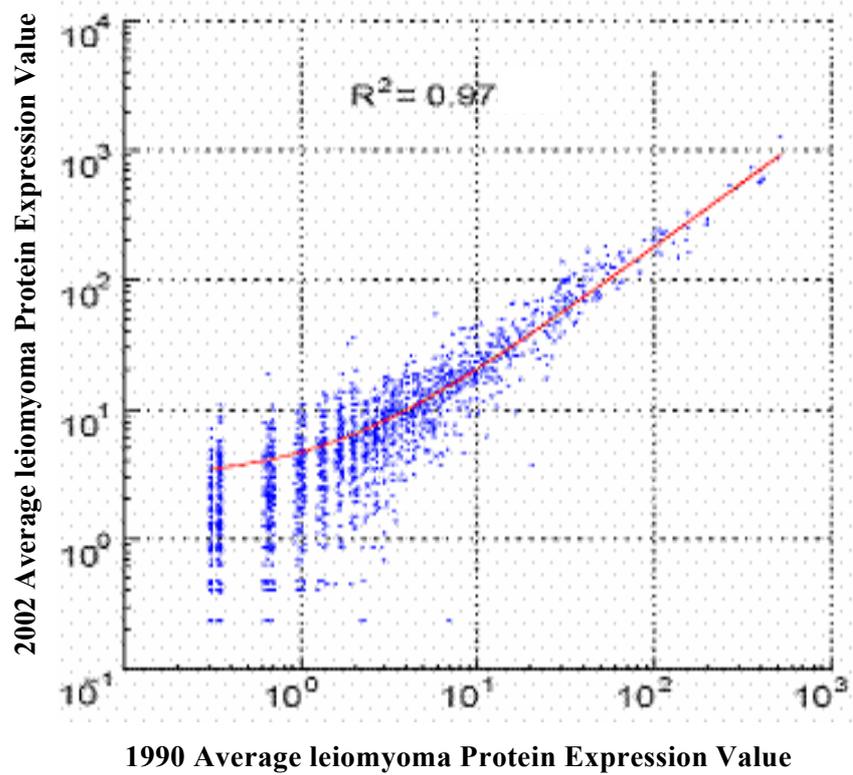


Figure 4-4 (A) Pearson correlation plot between 2002 and 1990 leiomyoma tumor groups. (B) Pearson correlation plot between 1980 sarcoma and 1990 leiomyoma tumor groups.

It is important to rule-out the possibility of comparison bias due to sample archival time. From this perspective it is interesting to note that the nine leiomyomas do not cluster by the archival year (Fig. 4-5). Neither do they group by diagnosis or patient age (data not shown). Furthermore, as discussed previously, each of the three leiomyoma groups (archived in 1990, 1997, and 2002) strongly correlate with the others in quantitative proteome measurements evaluated by the Pearson coefficient (Fig. 4-4A).

The expression values of three proteins, including actin, desmin, and progesterone receptor as commonly used markers for leiomyomas, were selected from proteome datasets to further investigate potential effects of archival time on the tissue proteome. As shown in Fig. 4-6, protein expression of these three markers was remarkably consistent over twelve years of archival time from 1990 to 2002. Representative IHC results shown in Fig. 4-7 validated MS-based proteome profiling of FFPE tumor tissues which were dated back as many as 18 years ago.

The modest disparity in the expression value of markers among nine leiomyomas (Fig. 4-6) could easily be attributed to case-to-case variation. The mean coefficient of variance (CV) for all quantified proteins across the leiomyoma cases was determined to be 48.9%. Considering a mean analytical CV of about 15-20% due solely to variation in sample preparation and instrumentation performance, a CV of less than 50% determined from nine patient samples does not seem excessive, nor does it indicate an archival effect.

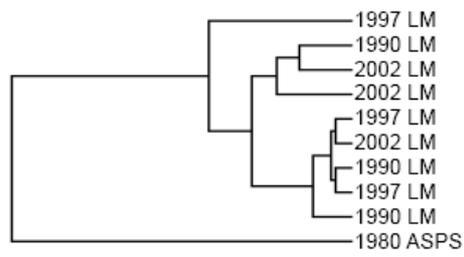


Figure 4-5 The hierarchical cluster analysis of all ten mesenchymal tumors.

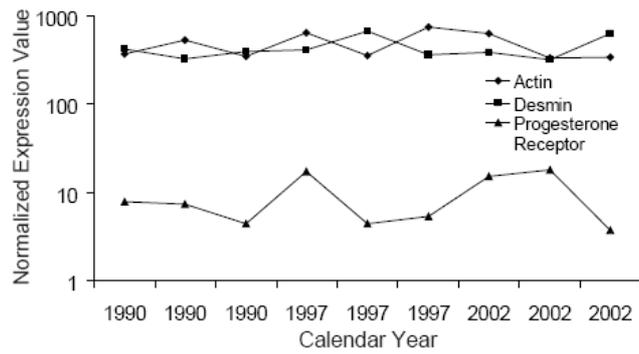


Figure 4-6 Distribution of protein expression of three leiomyoma markers, including actin, desmin, and progesterone receptor, over the archival years from 1990 to 2002.

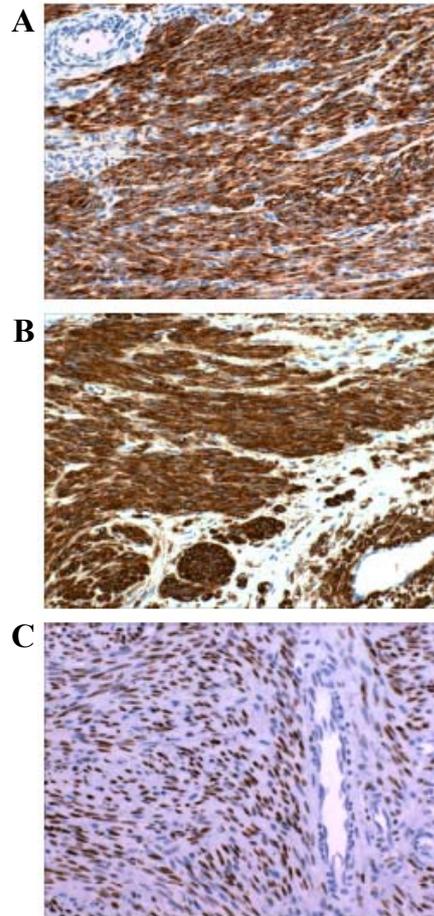


Figure 4-7 Representative IHC staining of three leiomyoma markers, including (A) desmin, (B) actin, and (C) progesterone receptor, on the same tissue blocks employed for MS-based proteome profiling.

In an attempt to evaluate the possibility of an archival effect on individual proteins or groups of proteins, k-means clustering was performed on all proteins across the average of the archival time points (Fig. 4-8). Proteins were organized into ten clusters with each cluster containing different range of protein abundance as determined by the number of spectral counts. The bottom right-hand corner of each panel displays the average change in protein expression with the top and bottom percentages for the comparisons between the 1990 and 2002 groups and among the 1997 and 2002 cases, respectively. Expression changes in all ten clusters were substantially less than 2-fold and were often within the analytical variability of the proteome platform.

While some individual proteins exhibited significant changes in expression between some cases on average, there was only modest change across time points consistent with an archival effect. That is, an archival effect may influence the retrieval of a protein or group of proteins over time and thus impact the subsequent proteomic measurements. Averaging the entire proteome data set for each averaged archival year shows an overall increase of protein expression by 23% from the 1990 to 2002 groups. The same calculation, however, indicates almost no difference in protein expression among the 1997 and 2002 cases. Furthermore, the k-means cluster suggests that there may be an archival effect for low abundance proteins detected and quantified by less than ten spectral counts in which these proteins are more difficult to be retrieved and extracted as the tissue block ages.

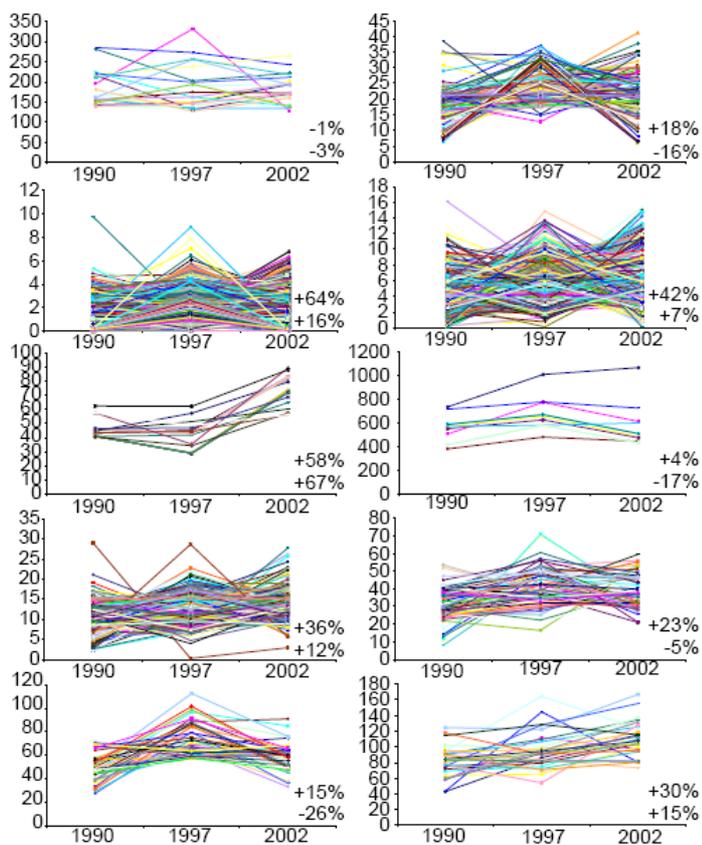


Figure 4-8 The k-means clustering of all leiomyoma proteins over the archival years from 1990 to 2002 on the x-axis. Proteins in each of ten clusters are varied in abundance as determined by the number of spectral counts on the y-axis.

Finally, the analysis of variance (ANOVA) was performed for comparing each of the three archived leiomyoma groups to the others to assess data normality. As shown in Fig. 4-9A, each comparison showed a consistent bias toward p -values below 0.5. When the cases were split for a two-way comparison such that each group had a roughly equal mix of cases from each archival date, the comparison data presented a more normal distribution (Fig. 4-9B). ANOVA results shown in Figs. 4-9A and 9B suggest that statistical comparisons across cases collected at different times may require the use of non-parametric methods or parametric comparisons should be performed only among groups archived around similar times.

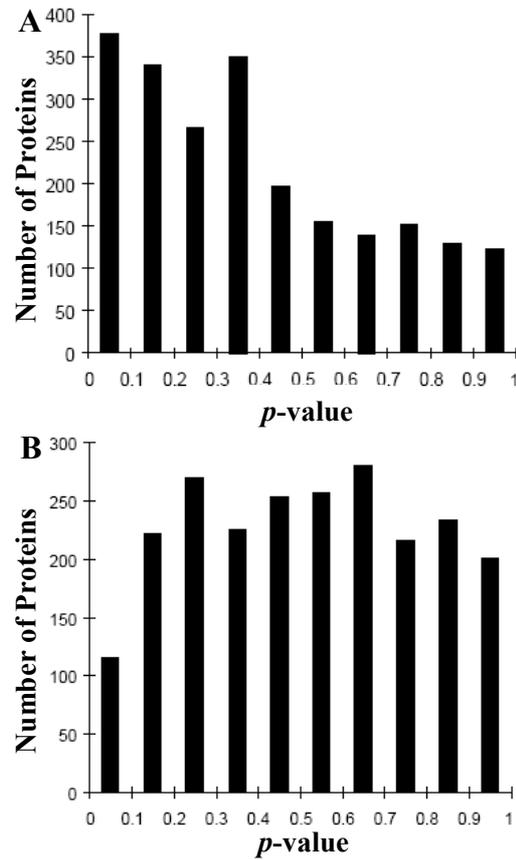


Figure 4-9 Performing ANOVA among proteomic results obtained from leiomyomas collected in 1990, 1997, and 2002. (A) Two-way comparison between 1990 and 2002 leiomyoma tumor groups. (B) Two-way comparison with each company containing a roughly equal mix of cases from each archival date.

4.4 CONCLUSION

By coupling optimized protein extraction and digestion procedures for handling FFPE tissues[57, 71, 133] with the CITP-based proteome platform[134, 135], ten archived mesenchymal tumor tissue blocks, including nine uterine leiomyomas dating from 1990 to 2002 and a single 1980 vaginal ASPS case, were employed for investigating potential effects of archival time on tissue proteome analysis. Although the 1990 group of leiomyomas demonstrated a slightly worse proteome performance in terms of total peptide, distinct peptide, and distinct protein identifications compared to leiomyomas cataloged in 1997 and 2002, great correlation among leiomyomas in quantitative proteomics was still achieved with a Pearson R^2 value of greater than 0.97 (Fig. 4-4A). Despite correlating among different patient samples, this value compares favorably with a recent report of a Pearson R^2 correlation of 0.88 for performing technical replicates of the same protein digest using a multidimensional liquid chromatography system[142].

Besides assessing proteomic data quality and reproducibility, several statistical measures, including the CV, k-means clustering, and ANOVA, were used to evaluate the possibility of an archival effect on individual proteins or groups of proteins within nine leiomyomas. That is, an archival effect may influence the retrieval of a protein or group of proteins over time and thus impact the subsequent proteomic measurements. For example, averaging the entire proteome data set for each averaged archival year shows an overall increase of protein expression by 23% from the 1990 to 2002 groups. Furthermore, the k-means cluster suggests that there may be an archival effect for low abundance proteins detected and quantified by less

than ten spectral counts in which these proteins are more difficult to be retrieved and extracted as the tissue block ages.

Still, the expression values of three commonly used leiomyoma markers, including actin, desmin, and progesterone receptor, were remarkably consistent over twelve years of archival time from 1990 to 2002 and validated using IHC measurements. Furthermore, high confidence and comparative proteome analysis between the uterine leiomyomas and the vaginal sarcoma (ASPS) was achieved using the sarcoma tissue block dating back as many as 28 years ago. Despite sharing over 1,800 common proteins in a core set, a total of 80 proteins were uniquely identified in the sarcoma tissues. Furthermore, the single 1980 sarcoma case was well-distinguished from the nine leiomyomas by an unsupervised hierarchical cluster analysis[143] even though the analyzers had no knowledge that one of the samples they had received was a substantially different tumor type.

4.5 ACKNOWLEDGEMENT

This work was supported by NIH grants GM073723 to CSL and CA 122715 to FAT and BMB.

CHAPTER FIVE

CONCLUSION

There is increasing acceptance of the critical importance of correlating the morphologic features of tissue with the data obtained from various molecular analytic techniques. Access to archived FFPE tissue specimens via shotgun-based proteomic analyses may, therefore, open new avenues for both prospective and retrospective translational research. However, one of the remaining issues in performing comparative proteomic measurements among FFPE tissues relates to potential variability in protein composition and retrieval based on length of storage periods.

The system which combines CIEF-based multidimensional separation system is evaluated using the human saliva sample secretory components as a model proteome system. Besides complexity, the greatest challenge facing comprehensive proteome analysis of saliva sample is related to the large variation of protein relative abundances, particularly in the identification of low abundance proteins. The combination of electrokinetic focusing/concentration with two highly resolving and orthogonal separation mechanisms in an integrated platform significantly enhances both the dynamic range and the detection sensitivity of MS toward the proteome analysis of microdissected tissue specimens. Instead of performing multiple runs of multidimensional separations, comparable or even better proteome results can be achieved by simply increasing the number of CIEF fractions due to the intrinsic high resolution nature of electrokinetic focusing. This unique feature is particularly important for the proteome analysis of limited tissue samples.

In the absence of protein amplification techniques, current proteome platforms including 2-D PAGE and shotgun-based multidimensional liquid chromatography separations, require substantially larger cellular samples which are generally incompatible with protein extract levels obtained from limited tissue samples, particularly in individual cell populations procured using tissue microdissection techniques. Thus, the work involved in this project has successfully established the proteome techniques for analyzing protein extracts obtained from microdissected Fresh Frozen and FFPE tissue specimens. By using the laser-free microdissection technology, we avoid potential detrimental effects of chemical reagents and laser heating on sample quality and protein separation. The capabilities of CIEF-based multidimensional separations for performing proteome analysis from minute samples create new opportunities in the pursuit of biomarker discovery using enriched and selected cell populations procured from tissue specimens. These proteome technological advances combined with recently developed tissue microdissection techniques provide powerful tools for those seeking to gain a greater understanding at the global level of the cellular machinery associated with human diseases such as cancer.

In addition to sample amount constraints imposed by current proteome techniques, the lack of optimized methodologies for protein extraction from FFPE tissues further restricts the ability to perform the proteomics analysis of archival tissues that exist in huge numbers. The goal of this stage of the research is to maximize the extent of protein recovery from FFPE tissues for subsequent comprehensive and comparative proteome analyses. It is well recognized that formalin-fixation causes major chemical changes in antigen by unknown protein cross linkage. A key question for the development of the original AR technique was

whether or not the cross-linkages of protein caused by formalin-fixation are reversible. This work involved in comparison between Urea extraction and SDS extraction. The SDS detergent is a widely used to solublize protein, and of the extraction is better than the Urea comparison. However, SDS will interfere with subsequent Tandem Mass detection, so we minimize the SDS effect on Mass Spec by dialyzing SDS out.

The evaluation of intact proteins using SDS-PAGE was conducted in this project, also the combined CIEF/nano-RPLC separations were employed for the examination of protein digests to provide further in depth comparison of proteomes within microdissected fresh frozen and FFPE GBM tissue specimens. OMSSA was used to search the peak list files against a decoyed SwissProt human database. The percentage of overlapping proteins obtained from repeated runs of the same FFPE GBM tissue sample was greater than 87%. A total of 2,845 non-redundant proteins were identified from comvined proteome runs of a single FFPE tissue sample. Comparing the proteome result obtained from the fresh frozen and FFPE tissues, most proteins identified from the FFPE slide were also detected in the corresponding fresh frozen section.

Last part of the research optimized the protein extraction and digestion procedures for handing FFPE tissue blocks stored for as many as 28 years. Optimized protein extraction and digestion procedures for handling FFPE tissues are coupled with the CIEF-based proteome technology to evaluate the effects of length of storage period on archival tissue proteome analysis across ten archived uterine mesenchymal tumor tissue blocks, including nine uterine leiomyomas dating from 1990 to 2002 and

a single case of ASPS from 1980. One of the unique proteins expressed in the ASPS, is further validated by IHC.

The results of comparative tissue proteome studies in combination with cancer pathology and biology are expected to provide significant details at the global level of the molecular mechanisms associated with cancer. Identification of differentially expressed proteins that are characteristic of a clearly defined disease state paves the way for defining the molecular and biochemical pathways by which normal cells progress to cancerous states in addition to nurturing discovery of biological markers and therapeutic targets for cancer. The greatest expectations for targeted proteomics research using enriched nonmalignant or malignant cells from high quality specimens reside in the identification of diagnostic, prognostic, and predictive biological markers in the clinical setting, as well as the discovery and validation of new protein targets in the biopharmaceutical industry.

LIST OF ABBREVIATIONS

AR: Antigen Retrieval

ASPS: Alveolar Soft Part Sarcoma

CIEF: Capillary Isoelectric Focusing

DNA: Deoxyribonucleic Acid

DTT: Dithiothreitol

ESI: Electrospray Ionization

FFPE: Formalin-Fixed Paraffin-Embedded

FDRs: False Discovery Rates

GBM: Glioblastoma Multiforme

H&E: Hematoxylin and Eosin

HIV: Human Immunodeficiency Virus

HPC: HydroxypropylCellulose

IAM: Iodoacetamide

IHC: Immunohistochemistry

LCM: Laser Capture Microdissection

LTQ: Linear Ion-Trap Mass Spectrometer

MALDI-MS: Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

MS: Mass Spectrometry

nano-RPLC: nano-Reversed Phase Liquid Chromatography

OMSSA: Open Mass Spectrometry Search Algorithm

PCR: Polymerase Chain Reaction

pI: Isoelectric Point

PSLT: Protein Subcellular Localization Prediction Tool

PTMs: Post-Translational Modifications

qTOF : quadrupole Time-Of-Flight

RNA: Ribonucleic Acid

RPLC: Reversed Phase Liquid Chromatography

SDS: Sodium Dodecyl Sulfate

SELDI-MS: surface-enhanced laser desorption/ionization-mass spectrometry

SNP: Single Nucleotide Polymorphism

TBS: Tris Buffered Saline

Tris: tris(hydroxymethyl)aminomethane

2-D PAGE: Two-Dimensional Polyacrylamide Gel Electrophoresis

VPP3: Vacuolar Proton translocating ATPase 116 kDa subunit isoform a3

REFERENCES

1. Group, B.D.W., "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework". *Clin. Pharmacol. Ther.* (2001), 69, 89-95.
2. Fox, R.;M. Hull, "Ultrasound diagnosis of polycystic ovaries". *Ann. N.Y. Acad. Sci.* (1993), 687, 217-223.
3. Diamandis, E.P.;H.A. Fritsche;H. Lilja, Chan;S. D. W.;M.K. (Eds.), *Tumor Markers: Physiology, Pathobiology, Technology, and Clinical Applications.* (2002), Washington D.C.: AACCC Press. pp. 33-63.
4. Bast, R.C.;P. Ravdin;D.F. Hayes;S. Bates;H. Fritsche;J.M. Jessup;N. Kemeny;G.Y. Locker;R.G. Mennel;M.R. Somerfield, "2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology". *J. Clin. Oncol.* (2001), 19, 1865-1878.
5. Arciero, C.;S.B. Somiari;C.D. Shriver;H. Brzeski;R. Jordan;H. Hu;D.L. Ellsworth;R.I. Somiari, "Functional relationship and gene ontology classification of breast cancer biomarkers". *Int. J. Biol. Markers* (2003), 18, 241-272.
6. Bland, K.I.;E.M.E. Copeland, *The breast: Comprehensive management of benign and malignant diseases.* (1998), Philadelphia: W. B. Saunders Co. pp. 499-517.
7. Bland, K.I.;E.M.E. Copeland, *The Breast: Comprehensive Management of Benign and Malignant Diseases.* (1998), Philadelphia: W. B. Saunders Co. pp. 458-498.
8. Cole, K.A.;D.B. Krizman;M.R. Emmert-Buck, "The genetics of cancer--a 3D model". *Nat. Genet.* (1999), 21, 38-41.
9. Wittliff, J.L.;S.T. Kunitake;S.S. Chu;J.C. Travis, "Applications of laser capture microdissection in genomics and proteomics". *J. Clin. Ligand Assay* (2000), 23, 66-73.
10. Emmert-Buck, M.R.;R.F. Bonner;P.D. Smith;R.F. Chuaqui;Z. Zhuang;S.R. Goldstein;R.A. Weiss;L.A. Liotta, "Laser capture microdissection". *Science* (1996), 274, 998-1001.
11. Bonner, R.F.;M. Emmert-Buck;K. Cole;T. Pohida;R. Chuaqui;S. Goldstein;L.A. Liotta, "Laser capture microdissection: molecular analysis of tissue". *Science* (1997), 278, 1481-1483.

12. De Souza, A.I.;E. McGregor;M.J. Dunn;M.L. Rose, "Preparation of human heart for laser microdissection and proteomics". *Proteomics* (2004), 4, 578-586.
13. Furuta, M.;R.J. Weil;A.O. Vortmeyer;S. Huang;J. Lei;T.-N. Huang;Y.-S. Lee;D.A. Bhowmick;I.A. Lubensky;E.H. Oldfield;Z. Zhuang, "Protein patterns and proteins that identify subtypes of glioblastoma multiforme". *Oncogene* (2004), 23, 6806-6814.
14. Zhuang, Z.;Y.S. Lee;W. Zeng;M. Furuta;T. Valyi-Nagy;M.D. Johnson;C.L. Vnencak-Jones;R.L. Woltjer;R.J. Weil, "Molecular genetic and proteomic analysis of synchronous malignant gliomas". *Neurology* (2004), 62, 2316-2319.
15. Wang, Y.;P.A. Rudnick;E.L. Evans;J. Li;Z. Zhuang;D.L. Devoe;C.S. Lee;B.M. Balgley, "Proteome analysis of microdissected tumor tissue using a capillary isoelectric focusing-based multidimensional separation platform coupled with ESI-tandem MS". *Anal. Chem.* (2005), 77, 6549-6556.
16. Banks, R.E.;M.J. Dunn;M.A. Forbes;A. Stanley;D. Pappin;T. Naven;M. Gough;P. Harnden;P.J. Selby, "The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis--preliminary findings". *Electrophoresis* (1999), 20, 689-700.
17. Ornstein, D.K.;J.W. Gillespie;C.P. Paweletz;P.H. Duray;J. Herring;C.D. Vocke;S.L. Topalian;D.G. Bostwick;W.M. Linehan;E.F. Petricoin;M.R. Emmert-Buck, "Proteomic analysis of laser capture microdissected human prostate cancer and in vitro prostate cell lines". *Electrophoresis* (2000), 21, 2235-2242.
18. Craven, R.A.;R.E. Banks, "Laser capture microdissection and proteomics: possibilities and limitation". *Proteomics* (2001), 1, 1200-1204.
19. Craven, R.A.;N. Totty;P. Harnden;P.J. Selby;R.E. Banks, "Laser capture microdissection and two-dimensional polyacrylamide gel electrophoresis: evaluation of tissue preparation and sample limitations". *Am. J. Pathol.* (2002), 160, 815-822.
20. Ahram, M.;M.J. Flaig;J.W. Gillespie;P.H. Duray;W.M. Linehan;D.K. Ornstein;S. Niu;Y. Zhao;E.F. Petricoin;M.R. Emmert-Buck, "Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications". *Proteomics* (2003), 3, 413-421.

21. Somiari, R.I.;A. Sullivan;S. Russell;S. Somiari;H. Hu;R. Jordan;A. George;R. Katenhusen;A. Buchowiecka;C. Arciero;H. Brzeski;J. Hooke;C. Shriver, "High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast". *Proteomics* (2003), 3, 1863-1873.
22. Somiari, R.I.;S. Somiari;S. Russell;C.D. Shriver, "Proteomics of breast carcinoma". *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* (2005), 815, 215-225.
23. Xu, B.J.;R.M. Caprioli;M.E. Sanders;R.A. Jensen, "Direct analysis of laser capture microdissected cells by MALDI mass spectrometry". *J. Am. Soc. Mass Spectrom.* (2002), 13, 1292-1297.
24. Bhattacharya, S.H.;A.A. Gal;K.K. Murray, "Laser capture microdissection MALDI for direct analysis of archival tissue". *J. Prot. Res.* (2003), 2, 95-98.
25. Chaurand, P.;M.E. Sanders;R.A. Jensen;R.M. Caprioli, "Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections". *Am. J. Pathol.* (2004), 165, 1057-1068.
26. Paweletz, C.P.;B. Trock;M. Pennanen;T. Tsangaris;C. Magnant;L.A. Liotta;E.F. Petricoin, "Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer". *Dis. Markers* (2001), 17, 301-307.
27. Petricoin, E.F.;K.C. Zoon;E.C. Kohn;J.C. Barrett;L.A. Liotta, "Clinical proteomics: translating benchside promise into bedside reality". *Nat. Rev. Drug Discov.* (2002), 1, 683-695.
28. Wulfkuhle, J.D.;L.A. Liotta;E.F. Petricoin, "Proteomic applications for the early detection of cancer". *Nat. Rev. Cancer* (2003), 3, 267-275.
29. Cottingham, K., "Clinical proteomics: are we there yet?" *Anal. Chem.* (2003), 75, 472-476.
30. Krieg, R.C.;N.T. Gaisa;C.P. Paweletz;R. Knuechel, "Proteomic analysis of human bladder tissue using SELDI approach following microdissection techniques". *Methods Mol. Biol.* (2005), 293, 255-267.
31. Melle, C.;G. Ernst;B. Schimmel;A. Bleul;R. Kaufmann;M. Hommann;K.K. Richter;W. Daffner;U. Settmacher;U. Claussen;F. von Eggeling, "Characterization of pepsinogen C as a potential biomarker for gastric cancer using a histo-proteomic approach". *J. Prot. Res.* (2005), 4, 1799-1804.
32. Gillette, M.A.;D.R. Mani;S.A. Carr, "Place of pattern in proteomic biomarker discovery". *J. Prot. Res.* (2005), 4, 1143-1154.

33. Wolters, D.A.;M.P. Washburn;J.R. Yates, "An automated multidimensional protein identification technology for shotgun proteomics". *Anal. Chem.* (2001), 73, 5683-5690.
34. VerBerkmoes, N.C.;J.L. Bundy;L. Hauser;K.G. Asano;J. Razumovskaya;F. Larimer;R.L. Hettich;J.L. Stephenson, "Integrating 'top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of *Shewanella oneidensis*". *J. Prot. Res.* (2002), 1, 239-252.
35. Peng, J.;J.E. Elias;C.C. Thoreen;L.J. Licklider;S.P. Gygi, "Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome". *J. Prot. Res.* (2003), 2, 43-50.
36. Washburn, M.P.;R. Ulaszek;C. Deciu;D.M. Schieltz;J.R. Yates, "Analysis of quantitative proteomic data generated via multidimensional protein identification technology". *Anal. Chem.* (2002), 74, 1650-1657.
37. Gygi, S.P.;B. Rist;T.J. Griffin;J. Eng;R. Aebersold, "Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags". *J. Prot. Res.* (2002), 1, 47-54.
38. Li, C.;Y. Hong;Y.-X. Tan;H. Zhou;J.-H. Ai;S.-J. Li;L. Zhang;Q.-C. Xia;J.-R. Wu;H.-Y. Wang;R. Zeng, "Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry". *Mol. Cell. Proteomics* (2004), 3, 399-409.
39. Zhang, J.;H. Hu;M. Gao;P. Yang;X. Zhang, "Comprehensive two-dimensional chromatography and capillary electrophoresis coupled with tandem time-of-flight mass spectrometry for high-speed proteome analysis". *Electrophoresis* (2004), 25, 2374-2383.
40. DeSouza, L.;G. Diehl;M.J. Rodrigues;J. Guo;A.D. Romaschin;T.J. Colgan;K.W.M. Siu, "Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry". *J. Prot. Res.* (2005), 4, 377-386.
41. Cagney, G.;S. Park;C. Chung;B. Tong;C. O'Dushlaine;D.C. Shields;A. Emili, "Human tissue profiling with multidimensional protein identification technology". *J. Prot. Res.* (2005), 4, 1757-1767.

42. Wu, S.-L.;W.S. Hancock;G.G. Goodrich;S.T. Kunitake, "An approach to the proteomic analysis of a breast cancer cell line (SKBR-3)". *Proteomics* (2003), 3, 1037-1046.
43. Zang, L.;D. Palmer Toy;W.S. Hancock;D.C. Sgroi;B.L. Karger, "Proteomic analysis of ductal carcinoma of the breast using laser capture microdissection, LC-MS, and 16O/18O isotopic labeling". *J. Prot. Res.* (2004), 3, 604-612.
44. Baker, H.;V. Patel;A.A. Molinolo;E.J. Shillitoe;J.F. Ensley;G.H. Yoo;A. Meneses-Garcia;J.N. Myers;A.K. El-Naggar;J.S. Gutkind;W.S. Hancock, "Proteome-wide analysis of head and neck squamous cell carcinomas using laser-capture microdissection and tandem mass spectrometry". *Oral Oncol.* (2005), 41, 183-199.
45. Hood, B.L.;M.M. Darfler;T.G. Guiel;B. Furusato;D.A. Lucas;B.R. Ringeisen;I.A. Sesterhenn;T.P. Conrads;T.D. Veenstra;D.B. Krizman, "Proteomic analysis of formalin-fixed prostate cancer tissue". *Mol. Cell. Proteomics* (2005), 4, 1741-1753.
46. Chen, J.;B.M. Balgley;D.L. DeVoe;C.S. Lee, "Capillary isoelectric focusing-based multidimensional concentration/separation platform for proteome analysis". *Anal. Chem.* (2003), 75, 3145-3152.
47. Wang, Y.;B.M. Balgley;P.A. Rudnick;E.L. Evans;D.L. DeVoe;C.S. Lee, "Integrated capillary isoelectric focusing/nano-reversed phase liquid chromatography coupled with ESI-MS for characterization of intact yeast proteins". *J. Prot. Res.* (2005), 4, 36-42.
48. Chen, H.;C. Horvath, "High-speed high-performance liquid chromatography of peptides and proteins". *J. Chromatogr. A* (1995), 705, 3-20.
49. Yan, F.;B. Subramanian;A. Nakeff;T.J. Barder;S.J. Parus;D.M. Lubman, "A comparison of drug-treated and untreated HCT-116 human colon adenocarcinoma cells using a 2-D liquid separation mapping method based upon chromatofocusing PI fractionation". *Anal. Chem.* (2003), 75, 2299-2308.
50. Wang, H.;S.G. Clouthier;V. Galchev;D.E. Misek;U. Duffner;C.-K. Min;R. Zhao;J. Tra;G.S. Omenn;J.L.M. Ferrara;S.M. Hanash, "Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids". *Mol. Cell. Proteomics* (2005), 4, 618-625.
51. Millea, K.M.;I.S. Krull;S.A. Cohen;J.C. Gebler;S.J. Berger, "Integration of multidimensional chromatographic protein separations with a combined "top-down" and "bottom-up" proteomic strategy". *J. Prot. Res.* (2006), 5, 135-146.

52. Cargile, B.J.;D.L. Talley;J.L. Stephenson, "Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides". *Electrophoresis* (2004), 25, 936-945.
53. Essader, A.S.;B.J. Cargile;J.L. Bundy;J.L. Stephenson, "A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics". *Proteomics* (2005), 5, 24-34.
54. Conti, M.;M. Galassi;A. Bossi;P.G. Righetti, "Capillary Isoelectric Focusing: the Problem of Protein Solubility". *J. Chromatogr. A* (1997), 757, 237-245.
55. Chevallet, M.;V. Santoni;A. Poinas;D. Rouquie;A. Fuchs;S. Kieffer;M. Rossignol;J. Lunardi;J. Garin;T. Rabilloud, "New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis". *Electrophoresis* (1998), 19, 1901-1909.
56. Pedersen, S.K.;J.L. Harry;L. Sebastian;J. Baker;M.D. Traini;J.T. McCarthy;A. Manoharan;M.R. Wilkins;A.A. Gooley;P.G. Righetti;N.H. Packer;K.L. Williams;B.R. Herbert, "Unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins". *J. Prot. Res.* (2003), 2, 303-311.
57. Guo, T.;W. Wang;P.A. Rudnick;T. Song;J. Li;Z. Zhuang;R.J. Weil;D.L. DeVoe;C.S. Lee;B.M. Balgley, "Proteome analysis of microdissected formalin-fixed and Paraffin-embedded tissue specimens". *J. Histochem. Cytochem.* (2007), 55, 763-772.
58. Geer, L.Y.;S.P. Markey;J.A. Kowalak;L. Wagner;M. Xu;D.M. Maynard;X. Yang;W. Shi;S.H. Bryant, "Open mass spectrometry search algorithm". *J. Prot. Res.* (2004), 3, 958-964.
59. Elias, J.E.;W. Haas;B.K. Faherty;S.P. Gygi, "Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations". *Nat. Methods* (2005), 2, 667-675.
60. Rudnick, P.A.;Y. Wang;E. Evans;C.S. Lee;B.M. Balgley, "Large scale analysis of MASCOT results using a Mass Accuracy-based THreshold (MATH) effectively improves data interpretation". *J. Prot. Res.* (2005), 4, 1353-1360.
61. Wang, W.;H. Zhou;H. Lin;S. Roy;T.A. Shaler;L.R. Hill;S. Norton;P. Kumar;M. Anderle;C.H. Becker, "Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards". *Anal. Chem.* (2003), 75, 4818-4826.

62. Chelius, D.;T. Zhang;G. Wang;R.-F. Shen, "Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry". *Anal. Chem.* (2003), 75, 6658-6665.
63. Liu, H.;R.G. Sadygov;J.R. Yates, "A model for random sampling and estimation of relative protein abundance in shotgun proteomics". *Anal. Chem.* (2004), 76, 4193-4201.
64. Rappsilber, J.;U. Ryder;A.I. Lamond;M. Mann, "Large-scale proteomic analysis of the human spliceosome". *Genome Res.* (2002), 12, 1231-1245.
65. Ishihama, Y.;Y. Oda;T. Tabata;T. Sato;T. Nagasu;J. Rappsilber;M. Mann, "Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein". *Mol. Cell. Proteomics* (2005), 4, 1265-1272.
66. Shi, S.R.;M.E. Key;K.L. Kalra, "Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections". *J. Histochem. Cytochem.* (1991), 39, 741-748.
67. Shi, S.R.;R.J. Cote;C.R. Taylor, "Antigen retrieval immunohistochemistry: past, present, and future". *J. Histochem. Cytochem.* (1997), 45, 327-343.
68. Shi, S.R.;R.J. Cote;C.R. Taylor, "Antigen retrieval techniques: current perspectives". *J. Histochem. Cytochem.* (2001), 49, 931-937.
69. Taylor, C.R.;R.J. Cote, *Immunomicroscopy: A Diagnostic Tool for the Surgical Pathologist.* (2005), Philadelphia: Elsevier Sanders.
70. Prieto, D.A.;B.L. Hood;M.M. Darfler;T.G. Guiel;D.A. Lucas;T.P. Conrads;T.D. Veenstra;D.B. Krizman, "Liquid Tissue: proteomic profiling of formalin-fixed tissues". *Biotechniques* (2005), 38, 32-35.
71. Shi, S.-R.;C. Liu;B.M. Balgley;C. Lee;C.R. Taylor, "Protein extraction from formalin-fixed, paraffin-embedded tissue sections: quality evaluation by mass spectrometry". *J. Histochem. Cytochem.* (2006), 54, 739-743.
72. Crockett, D.K.;Z. Lin;C.P. Vaughn;M.S. Lim;K.S.J. Elenitoba-Johnson, "Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS". *Lab. Invest.* (2005), 85, 1405-1415.
73. Palmer-Toy, D.E.;B. Krastins;D.A. Sarracino;J.B. Nadol;S.N. Merchant, "Efficient method for the proteomic analysis of fixed and embedded tissues". *J. Prot. Res.* (2005), 4, 2404-2411.

74. Shi, S.R.;J. Gu;C.R.E. Taylor, *Antigen retrieval techniques : immunohistochemistry and molecular morphology*. (2000), Natick: Eaton Publishing. pp.165-179.
75. Shi, S.R.;J. Gu;C.R.E. Taylor, *Antigen retrieval techniques : immunohistochemistry and molecular morphology*. (2000), Natick: Eaton Publishing. pp.7-13.
76. Shi, S.R.;J. Gu;C.R.E. Taylor, *Antigen retrieval techniques : immunohistochemistry and molecular morphology*. (2000), Natick: Eaton Publishing. pp.275-285.
77. Washburn, M.P.;D. Wolters;J.R. Yates, "Large-scale analysis of the yeast proteome by multidimensional protein identification technology". *Nat. Biotechnol.* (2001), 19, 242-247.
78. Streckfus, C.F.;L.R. Bigler, "Saliva as a diagnostic fluid". *Oral Dis.* (2002), 8, 69-76.
79. Banks, R.E.;M.J. Dunn;D.F. Hochstrasser;J.C. Sanchez;W. Blackstock;D.J. Pappin;P.J. Selby, "Proteomics: new perspectives, new biomedical opportunities". *Lancet* (2000), 356, 1749-1756.
80. Drake, R.R.;L.H. Cazare;O.J. Semmes;J.T. Wadsworth, "Serum, salivary and tissue proteomics for discovery of biomarkers for head and neck cancers". *Expert Rev. Mol. Diagn.* (2005), 5, 93-100.
81. Amado, F.M.L.;R.M.P. Vitorino;P.M.D.N. Domingues;M.J.C. Lobo;J.A.R. Duarte, "Analysis of the human saliva proteome". *Expert Rev. Proteomics* (2005), 2, 521-539.
82. Fischer, H.P.;W. Eich;I.J. Russell, "A possible role for saliva as a diagnostic fluid in patients with chronic pain". *Semin. Arthritis Rheum.* (1998), 27, 348-359.
83. Kaufman, E.;I.B. Lamster, "Analysis of saliva for periodontal diagnosis--a review". *J. Clin. Periodontol.* (2000), 27, 453-465.
84. Kaufman, E.;I.B. Lamster, "The diagnostic applications of saliva--a review". *Crit. Rev. Oral Biol. Med.* (2002), 13, 197-212.
85. Nagler, R.M.;O. Hershkovich;S. Lischinsky;E. Diamond;A.Z. Reznick, "Saliva analysis in the clinical setting: revisiting an underused diagnostic tool". *J. Investig. Med.* (2002), 50, 214-225.

86. Ghafouri, B.;C. Tagesson;M. Lindahl, "Mapping of proteins in human saliva using two-dimensional gel electrophoresis and peptide mass fingerprinting". *Proteomics* (2003), 3, 1003-1015.
87. Yao, Y.;E.A. Berg;C.E. Costello;R.F. Troxler;F.G. Oppenheim, "Identification of protein components in human acquired enamel pellicle and whole saliva using novel proteomics approaches". *J. Biol. Chem.* (2003), 278, 5300-5308.
88. Vitorino, R.;M.J.C. Lobo;A.J. Ferrer-Correira;J.R. Dubin;K.B. Tomer;P.M. Domingues;F.M.L. Amado, "Identification of human whole saliva protein components using proteomics". *Proteomics* (2004), 4, 1109-1115.
89. Huang, C.-M., "Comparative proteomic analysis of human whole saliva". *Arch. Oral Biol.* (2004), 49, 951-962.
90. Hardt, M.;L.R. Thomas;S.E. Dixon;G. Newport;N. Agabian;A. Prakobphol;S.C. Hall;H.E. Witkowska;S.J. Fisher, "Toward defining the human parotid gland salivary proteome and peptidome: identification and characterization using 2D SDS-PAGE, ultrafiltration, HPLC, and mass spectrometry". *Biochemistry* (2005), 44, 2885-2899.
91. Hu, S.;Y. Xie;P. Ramachandran;R.R. Ogorzalek Loo;Y. Li;J.A. Loo;D.T. Wong, "Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry". *Proteomics* (2005), 5, 1714-1728.
92. Wilmarth, P.A.;M.A. Riviere;D.L. Rustvold;J.D. Lauten;T.E. Madden;L.L. David, "Two-dimensional liquid chromatography study of the human whole saliva proteome". *J. Prot. Res.* (2004), 3, 1017-1023.
93. Xie, H.;N.L. Rhodus;R.J. Griffin;J.V. Carlis;T.J. Griffin, "A catalogue of human saliva proteins identified by free flow electrophoresis-based peptide separation and tandem mass spectrometry". *Mol. Cell. Proteomics* (2005), 4, 1826-1830.
94. Ayad, M.;B.C. Van Wuyckhuysse;K. Minaguchi;R.F. Raubertas;G.S. Bedi;R.J. Billings;W.H. Bowen;L.A. Tabak, "The association of basic proline-rich peptides from human parotid gland secretions with caries experience". *J. Dent. Res.* (2000), 79, 976-982.
95. Chen, Y.-C.;T.-Y. Li;M.-F. Tsai, "Analysis of the saliva from patients with oral cancer by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry". *Rapid Commun. Mass Spectrom.* (2002), 16, 364-369.

96. Mizukawa, N.;K. Sugiyama;J. Fukunaga;T. Ueno;K. Mishima;S. Takagi;T. Sugahara, "Defensin-1, a peptide detected in the saliva of oral squamous cell carcinoma patients". *Anticancer Res.* (1998), 18, 4645-4649.
97. Mizukawa, N.;K. Sugiyama;T. Ueno;K. Mishima;S. Takagi;T. Sugahara, "Defensin-1, an antimicrobial peptide present in the saliva of patients with oral diseases". *Oral Dis.* (1999), 5, 139-142.
98. Mathews, M.;H.P. Jia;J.M. Guthmiller;G. Losh;S. Graham;G.K. Johnson;B.F. Tack;P.B. McCray, "Production of beta-defensin antimicrobial peptides by the oral mucosa and salivary glands". *Infect. Immun.* (1999), 67, 2740-2745.
99. Dale, B.A.;S. Krisanaprakornkit, "Defensin antimicrobial peptides in the oral cavity". *J. Oral Pathol. Med.* (2001), 30, 321-327.
100. Ganz, T., "Defensins: antimicrobial peptides of innate immunity". *Nat. Rev. Immunol.* (2003), 3, 710-720.
101. Van Nieuw Amerongen, A.;J.G.M. Bolscher;E.C.I. Veerman, "Salivary proteins: protective and diagnostic value in cariology?" *Caries Res.* (2004), 38, 247-253.
102. Czesnikiewicz-Guzik, M.;W. Bielanski;T.J. Guzik;B. Loster;S.J. Konturek, "Helicobacter pylori in the oral cavity and its implications for gastric infection, periodontal health, immunology and dyspepsia". *J. Physiol. Pharmacol.* (2005), 56 Suppl 6, 77-89.
103. Hodinka, R.L.;T. Nagashunmugam;D. Malamud, "Detection of human immunodeficiency virus antibodies in oral fluids". *Clin. Diagn. Lab. Immunol.* (1998), 5, 419-426.
104. Fisker, N.;J. Georgsen;T. Stolborg;M.R. Khalil;P.B. Christensen, "Low hepatitis B prevalence among pre-school children in Denmark: saliva anti-HBc screening in day care centres". *J. Med. Virol.* (2002), 68, 500-504.
105. Nigatu, W.;L. Jin;B.J. Cohen;D.J. Nokes;M. Etana;F.T. Cutts;D.W. Brown, "Measles virus strains circulating in Ethiopia in 1998-1999: molecular characterisation using oral fluid samples and identification of a new genotype". *J. Med. Virol.* (2001), 65, 373-380.
106. Cone, E.J.;L. Presley;M. Lehrer;W. Seiter;M. Smith;K.W. Kardos;D. Fritch;S. Salamone;R.S. Niedbala, "Oral fluid testing for drugs of abuse: positive prevalence rates by Intercept immunoassay screening and GC-MS-MS confirmation and suggested cutoff concentrations". *J. Anal. Toxicol.* (2002), 26, 541-546.

107. Isemura, S.;E. Saitoh;K. Sanada;K. Minakata, "Identification of full-sized forms of salivary (S-type) cystatins (cystatin SN, cystatin SA, cystatin S, and two phosphorylated forms of cystatin S) in human whole saliva and determination of phosphorylation sites of cystatin S". *J. Biochem. (Tokyo)* (1991), 110, 648-654.
108. Lamkin, M.S.;F.G. Oppenheim, "Structural features of salivary function". *J. Dent. Res.* (1994), 73, 191-194.
109. Drzymala, L.;A. Castle;J.C. Cheung;A. Bennick, "Cellular phosphorylation of an acidic proline-rich protein, PRP1, a secreted salivary phosphoprotein". *Biochemistry* (2000), 39, 2023-2031.
110. Driscoll, J.;Y. Zuo;T. Xu;J.R. Choi;R.F. Troxler;F.G. Oppenheim, "Functional comparison of native and recombinant human salivary histatin 1". *J. Dent. Res.* (1995), 74, 1837-1844.
111. Lupi, A.;I. Messina;G. Denotti;M.E. Schininà;G. Gambarini;M.B. Fadda;A. Vitali;T. Cabras;V. Piras;M. Patamia;M. Cordaro;B. Giardina;M. Castagnola, "Identification of the human salivary cystatin complex by the coupling of high-performance liquid chromatography and ion-trap mass spectrometry". *Proteomics* (2003), 3, 461-467.
112. Messina, I.;T. Cabras;R. Inzitari;A. Lupi;C. Zuppi;C. Olmi;M.B. Fadda;M. Cordaro;B. Giardina;M. Castagnola, "Characterization of the human salivary basic proline-rich protein complex by a proteomic approach". *J. Prot. Res.* (2004), 3, 792-800.
113. Taylor, C.R.;S.R. Shi;R.J. Cote, "Antigen retrieval for immunohistochemistry: Status and need for greater standardization". *Appl. Immunohistochem.* (1996), 4, 144-166.
114. Gown, A.M., "Unmasking the mysteries of antigen or epitope retrieval and formalin fixation". *Am. J. Clin. Pathol.* (2004), 121, 172-174.
115. Ikeda, K.;T. Monden;T. Kanoh;M. Tsujie;H. Izawa;A. Haba;T. Ohnishi;M. Sekimoto;N. Tomita;H. Shiozaki;M. Monden, "Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections". *J. Histochem. Cytochem.* (1998), 46, 397-403.
116. Yamashita, S.;Y. Okada, "Mechanisms of Heat-induced Antigen Retrieval : Analyses In Vitro Employing SDS-PAGE and Immunohistochemistry". *J Histochem. Cytochem.* (2005), 53, 13-21.

117. Wang, Y.;B.M. Balgley;C.S. Lee, "Tissue proteomics using capillary isoelectric focusing-based multidimensional separations". *Expert Rev. Proteomics* (2005), 2, 659-667.
118. Han, D.K.;J. Eng;H. Zhou;R. Aebersold, "Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry". *Nat. Biotechnol.* (2001), 19, 946-951.
119. Wei, J.;J. Sun;W. Yu;A. Jones;P. Oeller;M. Keller;G. Woodnutt;J.M. Short, "Global proteome discovery using an online three-dimensional LC-MS/MS". *J. Prot. Res.* (2005), 4, 801-808.
120. Lewis, F.;N.J. Maughan;V. Smith;K. Hillan;P. Quirke, "Unlocking the archive--gene expression in paraffin-embedded tissue". *J. Pathol.* (2001), 195, 66-71.
121. Nyska, A.;C.R. Moomaw;L. Lomnitski;P.C. Chan, "Glutathione S-transferase pi expression in forestomach carcinogenesis process induced by gavage-administered 2,4-hexadienal in the F344 rat". *Arch. Toxicol.* (2001), 75, 618-624.
122. Zuo, X.;L. Echan;P. Hembach;H.Y. Tang;K.D. Speicher;D. Santoli;D.W. Speicher, "Towards global analysis of mammalian proteomes using sample prefractionation prior to narrow pH range two-dimensional gels and using one-dimensional gels for insoluble and large proteins". *Electrophoresis* (2001), 22, 1603-1615.
123. Wall, D.B.;M.T. Kachman;S. Gong;R. Hinderer;S. Parus;D.E. Misek;S.M. Hanash;D.M. Lubman, "Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry". *Anal. Chem.* (2000), 72, 1099-1111.
124. Moritz, R.L.;H. Ji;F. Schutz;L.M. Connolly;E.A. Kapp;T.P. Speed;R.J. Simpson, "A proteome strategy for fractionating proteins and peptides using continuous free-flow electrophoresis coupled off-line to reversed-phase high-performance liquid chromatography". *Anal. Chem.* (2004), 76, 4811-4824.
125. Krogh, A.;B. Larsson;G. von Heijne;E.L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes". *J. Mol. Biol.* (2001), 305, 567-580.
126. Scott, M.S.;S.J. Calafell;D.Y. Thomas;M.T. Hallett, "Refining protein subcellular localization". *PLoS Comput. Biol.* (2005), 1, 5518-5528.

127. Olson, E.N.;D. Srivastava, "Molecular pathways controlling heart development". *Science* (1996), 272, 671-676.
128. Spring, F.A.;C.H. Holmes;K.L. Simpson;W.J. Mawby;M.J. Mattes;Y. Okubo;S.F. Parsons, "The Oka blood group antigen is a marker for the M6 leukocyte activation antigen, the human homolog of OX-47 antigen, basigin and neurothelin, an immunoglobulin superfamily molecule that is widely expressed in human cells and tissues". *Eur. J. Immunol.* (1997), 27, 891-897.
129. Hong, H.-H.L.;J. Dunnick;R. Herbert;T.R. Devereux;Y. Kim;R.C. Sills, "Genetic alterations in K-ras and p53 cancer genes in lung neoplasms from Swiss (CD-1) male mice exposed transplacentally to AZT". *Environ. Mol. Mutagen.* (2007), 48, 299-306.
130. Yamashita, S.;Y. Okada, "Application of heat-induced antigen retrieval to aldehyde-fixed fresh frozen sections". *J. Histochem. Cytochem.* (2005), 53, 1421-1432.
131. Jiang, X.;X. Jiang;S. Feng;R. Tian;M. Ye;H. Zou, "Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues". *J. Prot. Res.* (2007), 6, 1038-1047.
132. Hwang, S.I.;J. Thumar;D.H. Lundgren;K. Rezaul;V. Mayya;L. Wu;J. Eng;M.E. Wright;D.K. Han, "Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues". *Oncogene* (2007), 26, 65-76.
133. Xu, H.;L. Yang;W. Wang;S.R. Shi;C. Liu;Y. Liu;C.R. Taylor;C.S. Lee;B.M. Balgley, "Antigen retrieval for proteomic characterization of formalin-fixed and paraffin-embedded tissues". *J. Prot. Res.* (2008), in press.
134. Fang, X.;L. Yang;W. Wang;T. Song;C. Lee;D. Devoe;B. Balgley, "Comparison of Electrokinetics-Based Multidimensional Separations Coupled with Electrospray Ionization-Tandem Mass Spectrometry for Characterization of Human Salivary Proteins". *Anal. Chem.* (2007), 79, 5785-5792.
135. Fang, X.;W. Wang;L. Yang;K. Chandrasekaran;T. Kristian;B.M. Balgley;C.S. Lee, "Application of capillary isotachopheresis-based multidimensional separations coupled with electrospray ionization-tandem mass spectrometry for characterization of mouse brain mitochondrial proteome". *Electrophoresis* (2008).
136. Yang, L.;C.S. Lee;S.A. Hofstadler;R.D. Smith, "Characterization of microdialysis acidification for capillary isoelectric focusing-microelectrospray ionization mass spectrometry". *Anal. Chem.* (1998), 70, 4945-4950.

137. Balgley, B.M.;T. Laudeman;L. Yang;T. Song;C.S. Lee, "Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy". *Mol. Cell. Proteomics* (2007), 6, 1599-1608.
138. Taylor, C.R., "Immunohistologic studies of lymphomas. New methodology yields new information and poses new problems". *J. Histochem. Cytochem.* (1979), 27, 1189-1191.
139. Taylor, C.R., "Immunohistologic studies of lymphoma: past, present, and future". *J. Histochem. Cytochem.* (1980), 28, 777-787.
140. Shi, S.R.;C. Cote;K.L. Kalra;C.R. Taylor;A.K. Tandon, "A technique for retrieving antigens in formalin-fixed, routinely acid-decalcified, celloidin-embedded human temporal bone sections for immunohistochemistry". *J. Histochem. Cytochem.* (1992), 40, 787-792.
141. Fowler, C.B.;R.E. Cunningham;T.J. O'Leary;J.T. Mason, "'Tissue surrogates' as a model for archival formalin-fixed paraffin-embedded tissues". *Lab. Invest.* (2007), 87, 836-846.
142. Lu, P.;C. Vogel;R. Wang;X. Yao;E.M. Marcotte, "Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation". *Nat. Biotechnol.* (2007), 25, 117-124.
143. Eisen, M.B.;P.T. Spellman;P.O. Brown;D. Botstein, "Cluster analysis and display of genome-wide expression patterns". *Proc. Natl. Acad. Sci. U.S.A.* (1998), 95, 14863-14868.