ABSTRACT

Title of Document:	**TARGETING BIOMARKERS VIA CITP-BASED SELECTIVE PROTEOME ENRICHMENT**
	Xueping Fang, Doctor of Philosophy, 2008
Directed By:	Professor Cheng S. Lee Department of Chemistry and Biochemistry

Besides the complexity of protein samples, probably the greatest challenge presently facing comprehensive proteome analysis is related to the large variation of protein relative abundances (>6 orders of magnitude), having potential biological significance in mammalian systems. To achieve comprehensive proteome analysis including the identification of low abundance proteins, this project aims to develop and demonstrate a capillary isotachophoresis (CITP)-based proteome platform, capable of providing selective analyte enrichment and extremely high resolving power toward complex protein mixtures. In contrast to universally enriching all proteins by a similar degree, the result of the CITP stacking process is that major components may be diluted, but trace compounds are concentrated.

By employing combined CITP/nano-reversed phase liquid chromatography (nano-RPLC) separations, a total of 6,112 fully tryptic peptides are sequenced by electrospray ionization mass spectrometry (ESI-MS), leading to the identification of 1,479 distinct human

SwissProt protein entries from a single proteome sample of whole unstimulated human saliva. By comparing with capillary isoelectric focusing as another electrokinetics-based stacking approach, CITP not only offers a broad field of application, but also is less prone to protein/peptide precipitation during the analysis.

The CITP-based proteome platform is further employed for the analysis of protein expression within synaptic mitochondria isolated from mouse brain. The ultrahigh resolving power of CITP separation is evidenced by the large number of distinct peptide identifications measured from each CITP fraction together with the low peptide fraction overlapping among identified peptides. Furthermore, the collective proteome datasets yield the identification of 2,191 distinct mitochondrial protein entries, corresponding to 76% coverage of the MitoP2-database reference set.

Comparisons among CITP and multidimensional liquid chromatography techniques are conducted using a single processed protein digest from brain cancer stem cells, identical second dimension separation (nano-RPLC) and ESI-MS conditions, and consistent search parameters and cutoff established by the target-decoy determined false discovery rate. Besides achieving superior overall proteome performance in total peptide, distinct peptide, and distinct protein identifications, analytical reproducibility of the CITP proteome platform is determined by a Pearson R^2 value of 0.98 and a coefficient of variation of 15% across all proteins quantified.

TARGETING BIOMARKERS VIA
CITP-BASED SELECTIVE PROTEOME ENRICHMENT

By

Xueping Fang

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 L. DeVoe
Professor Douglas S. English
Professor Catherine C. Fenselau
Professor Douglas A. Julin

Table of Contents

List of Figures

List of Abbreviations

2-D PAGE	Two-dimensional polyacrylamide gel electrophoresis
AR	Antigen retrieval
BSA	Bovine serum albumin
CE	Capillary electrophoresis
CEC	Capillary electrochromatography
CIEF	Capillary isoelectric focusing
CITP	Capillary isotachophoresis
CV	Coefficient of variation
CZE	Capillary zone electrophoresis
DTT	Dithiothreitol
EGTA	Ethylene glycol tetraacetic acid
ESI	Electrospray ionization
E-value	Expectation-value
FDRs	False discovery rates
FFPE	Formalin-fixed and paraffin-embedded
GBM	Glioblastoma multiforme
HPLC	High performance liquid chromatography
IAM	Iodoacetamide
pI	Isoelectric point
LE	Leading electrolyte
LIFD	Laser-induced fluorescence detection

LTQ MS	Linear ion-trap mass spectrometer
MALDI	Matrix-assisted laser desorption/ionization
MEKC	Micellar electrokinetic chromatography
MS	Mass spectrometry or mass spectrometer
MudPIT	Multidimensional protein identification technology
PRPs	Praline rich proteins
ROS	Reactive oxygen species
RPLC	Reversed-phase liquid chromatography
OMSSA	Open mass spectrometry search algorithm
SCX	Strong cation-exchange
SELDI	Surface-enhanced laser desorption/ionization
SDS	Sodium dodecyl sulfate
TE	Terminating electrolyte

Chapter 1: Introduction

Sections of 1.1 Introduction - 1.3 Capillary Electrophoresis-Based Multidimensional Separations are reproduced with permission from Balgley, B. M., Wang, W., Fang, X., DeVoe, D. L., Lee, C. S., Clinical Proteomics - from Diagnosis to Therapy (2007), 73-88.

1.1 Introduction

Mass spectrometry (MS) employing electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) interfaces has evolved to become an essential tool for protein identification and sequencing. The vast number of proteins present in the proteome of a typical organism, tissue, and organ, however, requires that separations be performed on the mixture prior to introduction into a mass spectrometer. Capillary electrophoresis is based on a family of electrokinetically resolving mechanisms which are performed in narrow-bore capillaries under the influence of an electric field. Besides two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and liquid chromatography, capillary electrophoretic separations are attractive and promising alternatives due to their excellent efficiency, high resolving power, and ability to handle small cell populations and limited tissue samples in clinical proteomics.

1.2 Single Dimension Capillary Electrophoretic Separation

Capillary zone electrophoresis (CZE) resolves proteins and peptides based on their differences in electrophoretic mobility which is a function of the charge-to-size ratio. Due to its high throughput and excellent resolving power, the coupling of CZE with ESI-MS is therefore a particularly attractive approach for the discovery of biomarkers in human urine and serum. Peptides from healthy volunteers' urine were prepared and concentrated using a combination of anion-exchange chromatography and lyophilization and then compared with those from kidney disease patients' dialysate [1]. The application of CZE-ESI-MS has been shown to locate more potential kidney disease biomarkers [2, 3] than those obtained from surface-enhanced laser desorption/ionization (SELDI)-MS [4]. Additionally, CZE-ESI-MS has been employed to obtain peptide patterns from urine samples of urothelial cancer patients and healthy volunteers [5].

One characteristic inherent to CZE is the small sample volume that can be successfully applied to obtain a high-resolution analysis. As a good portion of the important proteins/peptides contained in a proteomic sample may be present at low levels, it becomes important to combine sample concentration/preparation in-line with the subsequent electrophoretic separations while avoiding analyte loss or dilution due to manual handling and transfer procedures. Thus, transient capillary isotachophoresis (CITP) was employed for on-capillary preconcentration of serum peptides prior to CZE separation [6]. CITP, unlike most separation modes in capillary electrophoresis, employs a discontinuous electrolyte system. The stacking and separation of peptides and proteins occur in the boundary region between the leading (high-mobility) and terminating (low-

2

mobility) electrolytes under the influence of an applied electric field. The application of transient CITP/CZE separations significantly impacts the analysis of low abundance peptides by effectively enhancing the dynamic range and detection sensitivity of MS as demonstrated by An and co-workers [7].

1.3 Capillary Electrophoresis-Based Multidimensional Separations

In addition to protein complexity, probably the greatest challenge facing comprehensive proteome analysis is related to the large variation of relative protein abundance (greater than 6 orders or magnitude) in human serum/plasma and urine samples. The use of only a single electrophoretic separation such as CZE prior to MS detection greatly limits the ability for mining deeper into the clinical proteome. Assuming the separation techniques used in the two dimensions are orthogonal, i.e., the two separation techniques are based on different physicochemical properties of proteins/peptides, the peak capacity of 2-D separations is the product of the peak capacities of individual one-dimensional approaches. Thus, various capillary electrophoresis-based 2-D separation schemes have been developed in an effort to improve the overall peak capacity and increase the number of detected peptides and proteins identified due to better use of the MS dynamic range and reduced discrimination during ionization.

1.3.1 Capillary liquid chromatography-capillary electrophoresis

A 2-D reversed-phase liquid chromatography (RPLC)-CZE separation platform has been coupled with ESI-tandem MS for the identification of proteins in complex mixtures [8]. A

serum sample was depleted of high-abundance proteins, digested with trypsin, and then separated by RPLC and collected into 96 fractions. These fractions, collected off-line from the RPLC separation, were subjected to sequential CZE separations through a sheathless ESI interface that was integrated on the separation capillary. A key feature of this work is the employment of RPLC in lieu of ion exchange fractionation as the first separation dimension, where there is a large degree of overlap. The CZE-MS results demonstrated less than 15% overlap between neighboring RPLC fractions, leading to a more sensitive proteome analysis.

Instead of resolving complex protein digests, intact proteins from liver cancer tissue have been separated using RPLC followed by array-based capillary isoelectric focusing (CIEF) equipped with whole column image detection [9]. In addition to the laser-induced fluorescence detection (LIFD), the CIEF protein fractions could be subsequently deposited on to MALDI-targets. On-target tryptic digestion of collected CIEF fractions was performed to generate peptide mass fingerprint for the identification of proteins in rat liver tissue extracts [10].

1.3.2 Capillary electrophoresis-capillary electrophoresis

Michaels and co-workers [11] detailed a 2-D electrokinetic separation system by on-line combination of micellar electrokinetic chromatography (MEKC) with CZE for resolving complex protein mixtures based on their differences in hydrophobicity and charge-to-size ratio, respectively. Protein fractions from the first dimension MEKC separation were transferred to a second capillary through an injection interface, where CZE was performed and followed by LIFD. Hu et al. [12] further combined capillary size-sieving electrophoresis on the basis of a sodium dodecyl sulfate (SDS)-pullulan buffer system with MEKC for the

study of protein expression in single mammalian cells. After a 6-min-long size-based separation, over 100 transfers of fractions from the first capillary were introduced into a second capillary for further separation by MEKC over an approximately 3.5-hr-long period.

1.3.3 Capillary electrophoresis-liquid chromatography

On-line combination of CIEF with capillary electrochromatography (CEC) via a nanoinjector valve has been reported by Zheng and El Rassi [13] for the analysis of albumin-depleted human serum. The serum proteins were systematically resolved by exploiting their differences in isoelectric point (pI) and hydrophobicity, and detected using UV absorbance. The key challenge associated with using CEC to perform peptide and protein separations is the requirement of charged surfaces that generate reproducible electroosmotic flow as an electrokinetically-driven pump for the delivery of the mobile phase.

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 global studies of protein profiles that will have diagnostic and therapeutic relevance. A CIEF-based clinical proteomics platform, which combines CIEF with nano-RPLC (Figure 1-1), has recently been developed to achieve comprehensive and ultrasensitive analysis of minute protein amounts extracted from small cell populations and limited tissue samples [14-17]. The key to performing sensitive clinical proteome analysis is to attain high analyte concentrations in small peak volumes prior to MS. The application of electrokinetic focusing together with the high resolving power of combined CIEF/nano-RPLC

separations greatly enhance the dynamic range and detection sensitivity of MS measurements. Furthermore, the microcolumn format of CIEF and nano-RPLC provides the amenability to automation and seamless coupling with ESI-MS in an integrated platform while avoiding potential sample loss and analyte dilution.

Combined CIEF/nano-RPLC separations have been employed for the analysis of protein profiles within small and selected tumor cell populations obtained from the a glioblastoma multiforme (GBM) tissue specimen [14]. By using ESI-quadrupole time-of-flight-MS for the analysis of peptides eluted from nano-RPLC (Figure 1-2), a total of 6,866 fully tryptic peptides have been detected, leading to the identification of 1,820 distinct proteins each corresponding to a unique human International Protein Index database entry.

These identifications were 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 μg of total protein. These identifications were based on high mass accuracy and high confidence hits to fully tryptic peptides. 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, were identified from the GBM tissue samples.
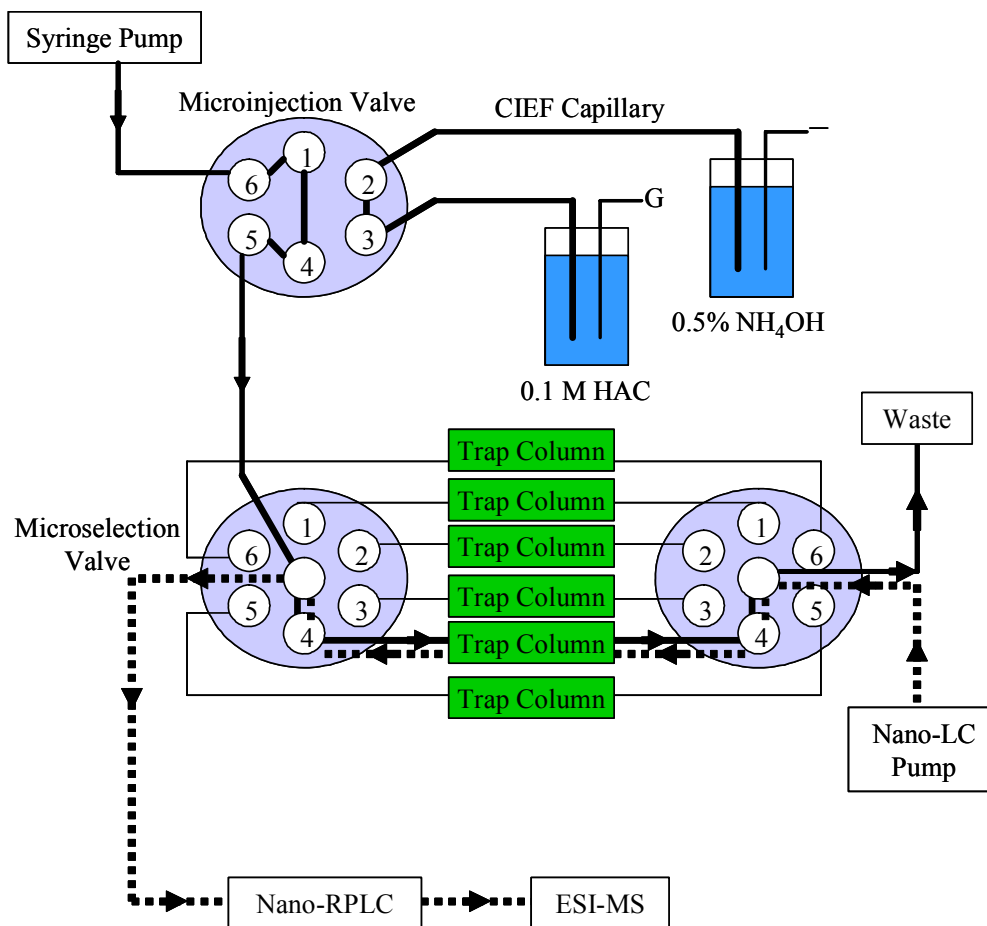
Figure 1-1 Schematic of on-line integration of CIEF with nano-RPLC as a concentrating and multidimensional protein/peptide separation platform. Solid and dashed lines represent the flow paths for the loading of CIEF fractions and the injection of fractions into a nano-RPLC column, respectively [18].
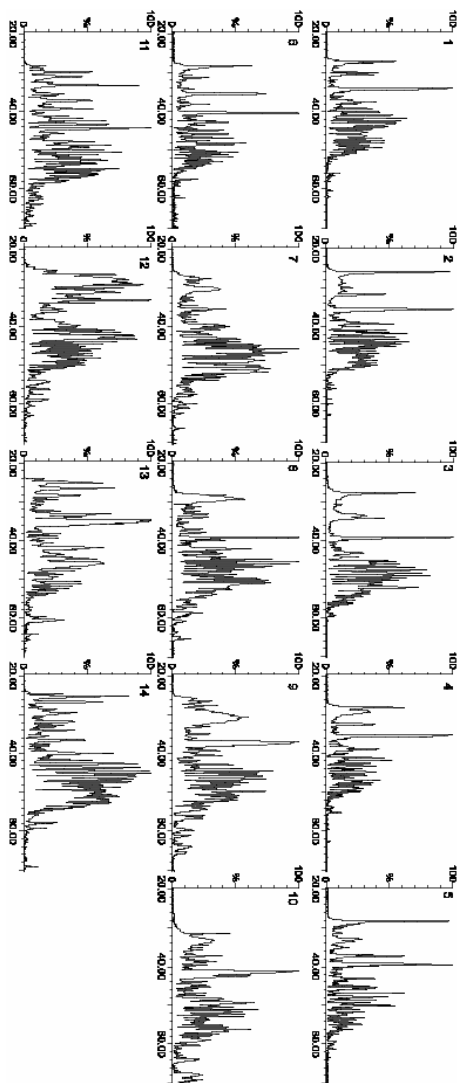
Figure 1-2 Base peak chromatograms of a representative CIEF/nano-RPLC multidimensional separation of tryptic peptide digest prepared from GBM tissue sample. Each number represents the sequence of CIEF fractions further analyzed by nano-RPLC from basic to acidic pHs [14].

8

In addition to the analysis of soluble proteins within tumor cells, the tissue proteome capabilities of combined CIEF/nano-RPLC separations have been further expanded for the identification of membrane proteins extracted from cell pellets using SDS detergent. Epithelial ovarian carcinoma tissue specimens containing the serous cell type were processed and employed for the membrane proteome analysis [17]. The peptide and protein false discovery rates (FDRs), and the numbers of total peptides, distinct peptides, and protein identifications were plotted as functions of the E-value using the Open Mass Spectrometry Search Algorithm (OMSSA) [19] (Figure 1-3). An E-value threshold of 0.05, corresponding to 1% FDR of total peptide identifications, was chosen as a cutoff and resulted in the identification of 18,861 distinct peptides together with 3,303 non-redundant proteins from the SwissProt human database. This cutoff led to a protein false positive rate of 6.8% as indicated by the detection of peptides from 112 distinct reversed protein sequences in the decoy section of the search database.

Among 12,484 protein entries in the SwissProt human database, a total of 3,514 proteins are predicted to contain at least one or more transmembrane domains using TMHMM (www.cbs.dtu.dk/services/TMHMM-2.0/). Out of these membrane protein candidates, 773 proteins were identified from the pellets of targeted ovarian tumor cells, corresponding to 22% membrane proteome coverage. Of the 773 proteins, only 291 proteins or 37% were also found in the soluble protein fraction of ovarian tissues using the urea solubilization protocol. Moreover, there were only 69 additional membrane proteins identified in the soluble protein fraction, affirming the utility of SDS extraction coupled with combined CIEF/nano-RPLC separations for profiling the membrane proteome within tissue specimens.
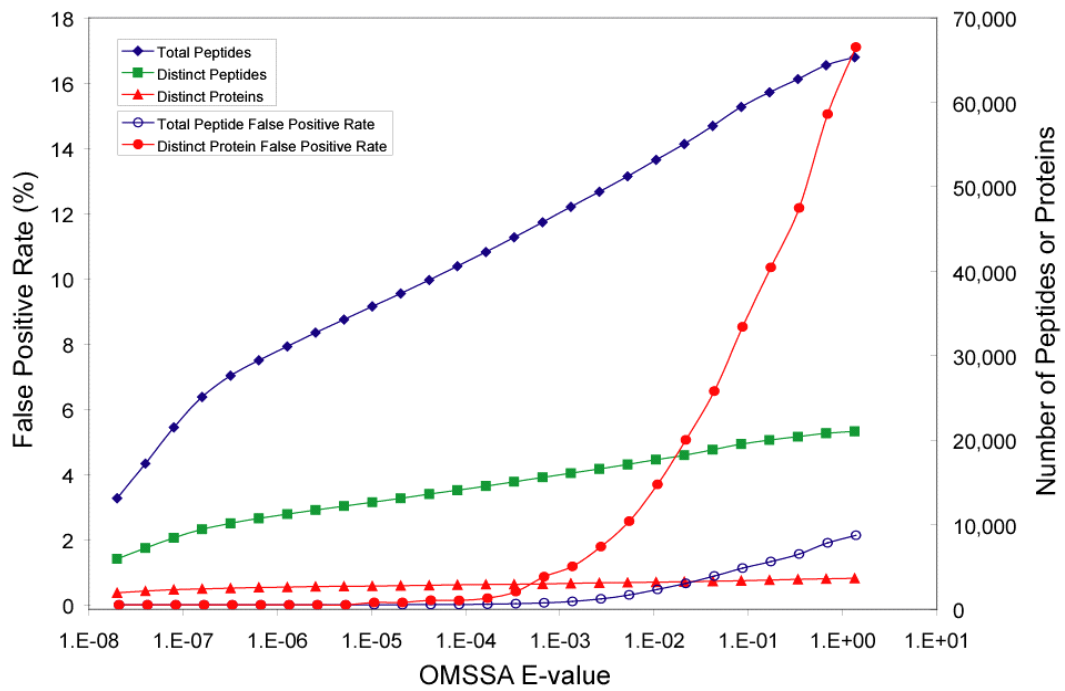
Figure 1-3 Plots of the false positive rates and the numbers of total peptide, distinct peptide, and distinct protein identifications versus the E-value obtained from the search of the peak list files against a decoyed SwissProt human database using OMSSA [17].

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 formalin-fixed and paraffin-embedded (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 proteomic efforts from retrieving information sequestered in archival tissues. By coupling with the antigen retrieval (AR) technique originally developed for immunohistochemistry [20], multidimensional CIEF/nano-RPLC separations have been employed to examine protein profiles extracted from a whole FFPE tissue section of human renal carcinoma [21]. A total of 3,336 distinct peptides were identified with a false positive rate of 1% from the tissue section treated with the heat-induced AR approach, leading to the identification of 1,830 distinct proteins. 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.

Instead of examining whole tissue sections, minute proteins procured from GBM FFPE tissues have been processed, profiled, and compared with those extracted from fresh frozen tissues of the same, matched patient [22]. The entire content of focused peptides in the CIEF capillary was split into 19 individual fractions (Figure 1-4) which were further resolved by nano-RPLC and identified using nano-ESI-linear ion trap-MS/MS. 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 2,845 non-redundant proteins from the SwissProt human database. 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.

Combined CIEF/nano-RPLC separations were also employed in the analysis of protein digests obtained from the soluble and cell pellet fractions of patient-matched fresh frozen GBM tissue. 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.

Comparing the proteome results obtained from the fresh frozen and FFPE GBM tissues (Figure 1-5), 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.
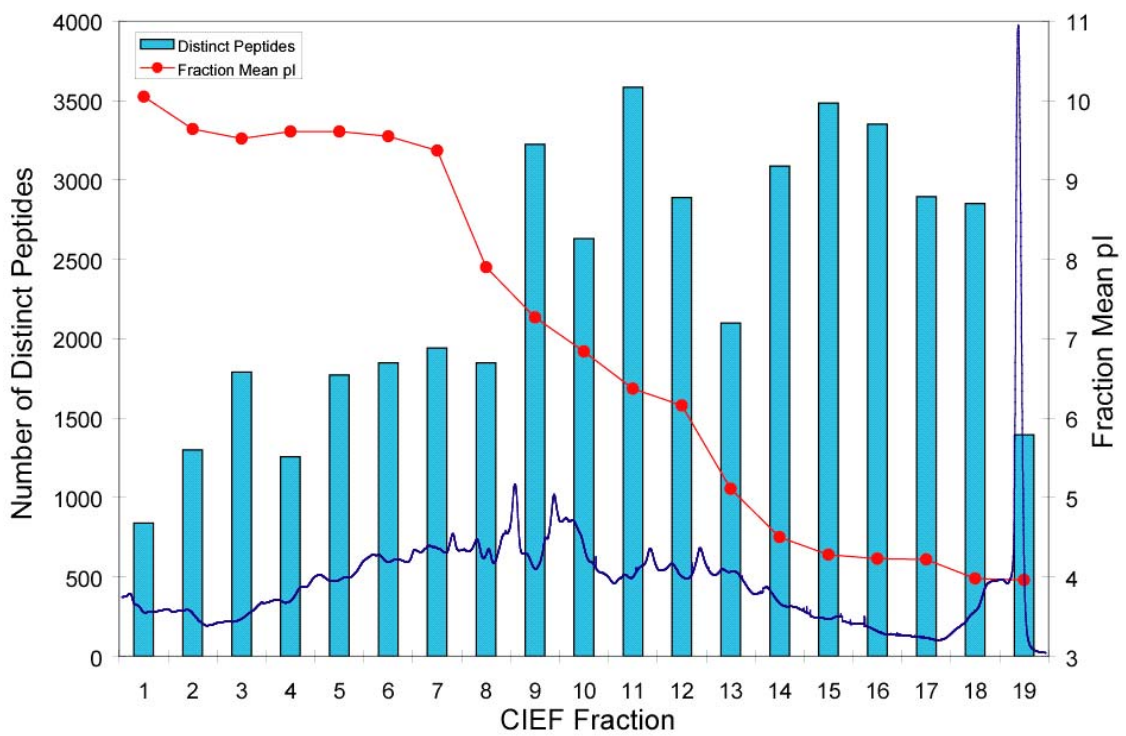
Figure 1-4 Overlaid plots containing the CIEF-UV trace monitored at 280 nm, the number of distinct peptides identified in each of the CIEF fractions, and the distribution of the peptide's mean pI values over the entire CIEF separation [22].
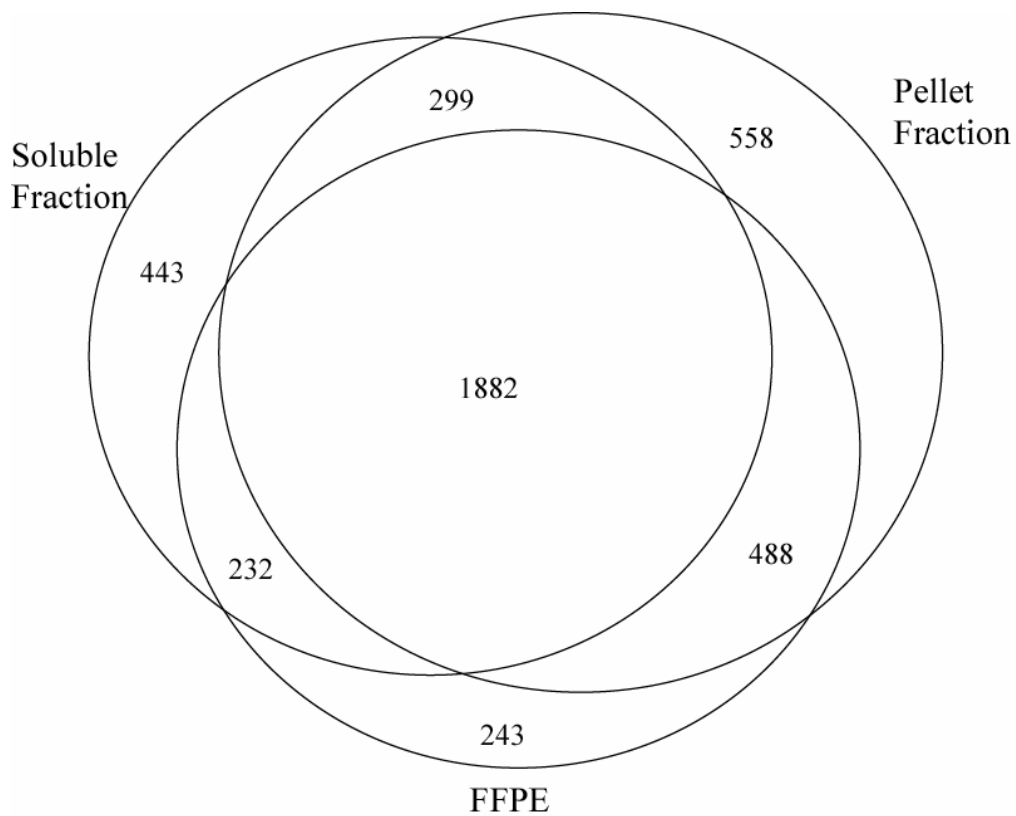
Figure 1-5 The overlap in the proteins identified from fresh frozen (the soluble and pellet fractions) and FFPE GBM tissue specimens of the same patient using combined CIEF/nano-RPLC separations coupled with nano-ESI-linear ion trap-MS/MS [22].

1.4 Project Description

Predictions of disease behavior and therapeutic response have been complicated by the great complexity of the human genome and the heterogeneity among different cell types. To address the issue of cell heterogeneity, several microdissection techniques [23-25] have been developed to provide a rapid and straightforward method for isolating selected subpopulations of diseased cells for subsequent genetic and proteomic analyses. Developments in capillary electrophoresis-based multidimensional separations are particularly important for their roles in achieving ultrasensitive and comprehensive analysis of minute proteins extracted from isolated cell populations.

Besides the complexity of protein samples, probably the greatest challenge presently facing comprehensive proteome analysis is related to the large variation of protein relative abundances (>6 orders of magnitude), having potential biological significance in mammalian systems. One obvious approach to increase the detection capability of low abundant proteins is to significantly raise the sample loading along with many additional protein/peptide fractionation and separation procedures, and therefore may be impractical for proteomic studies of small cell populations or limited tissue samples. To achieve comprehensive proteome analysis including the identification of low abundance proteins, my research efforts therefore focus on the development of transient CITP/CZE-based multidimensional separation technique and the demonstration of its unique bioanalytical capabilities toward the analysis of complex mammalian proteomes.

While CITP has been widely used for analyte preconcentration prior to electrophoretic separation, the application of CITP to selectively enrich trace amounts of proteins in complex mixtures, is completely untapped revenue in current proteome technology

development. CITP, unlike most separation modes in CE, employs a discontinuous electrolyte system. A CITP capillary is initially filled with a leading electrolyte such as ammonium acetate with a mobility greater than any of the sample components. Following sample injection, the inlet end of the capillary is placed in a terminating electrolyte such as acetic acid. The stacking and separation of analytes occurred in the boundary region between the leading and terminating electrolytes by applying an electric voltage (Figure 1-6).

Under the influence of constant current across the entire capillary, the sections containing the high-mobility leading electrolyte and the low-mobility terminating electrolyte would therefore experience lower and higher electric fields, respectively. The analyte ions in the sample zone initially migrating at fast speed under the influence of high electric field are slowed down and stacked at the interface of the sample and leading electrolyte sections. Stable boundaries formed between analytes and are based on their individual mobilities. In CITP, all bands move at the same velocity as the bands are focused. The individual band velocities are self-normalizing because the band containing the highly mobile analyte ions had high conductivity and, as a result, exhibited a lower voltage drop across the band. Stacking is also a consequence of velocity normalization. For example, if a band diffuses into a neighboring zone, it will either speed up or slow down, based on the field strength encountered, and rejoin the original band.

| TE Anode (+) | $E_{TE} >$ | E_{sample} | $> E_{LE}$ | LE Cathode(-) |

$$v = \mu\, E$$

$$\mu_{sample}\, E_{sample} > \mu_{sample}\, E_{LE}$$

| TE | $\mu_{sample}\, E_{sample}$ $\oplus$ | $\oplus$ $\mu_{sample} E_{LE}$ | LE |

| TE | C_{sample} | LE |
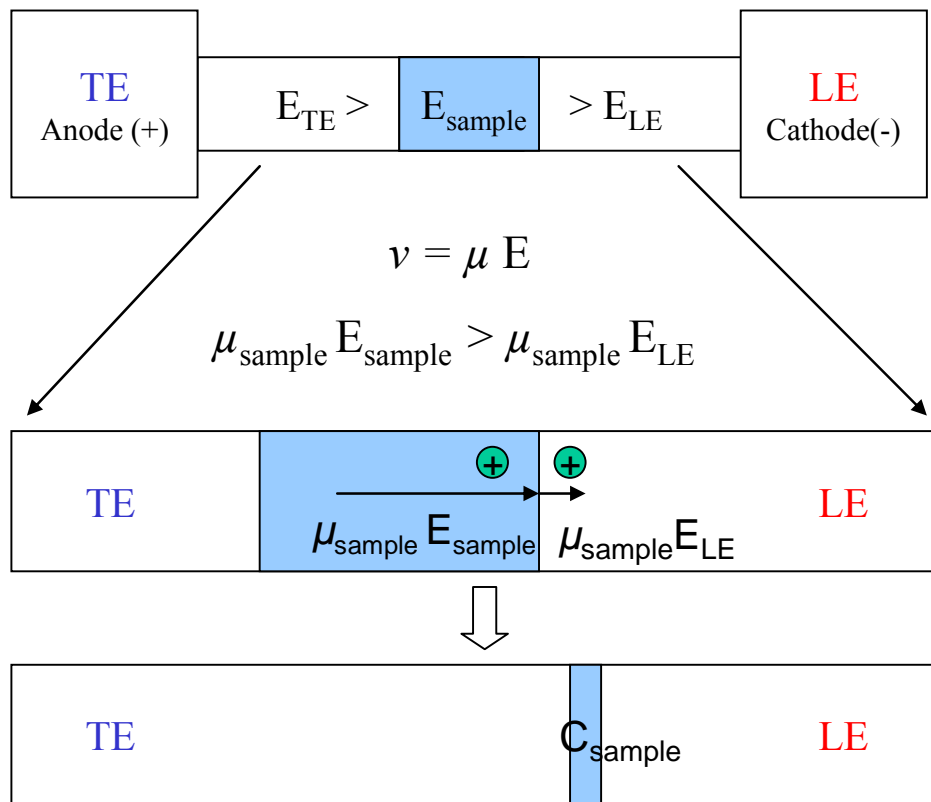
Figure 1-6 Schematic view of the CITP-based electrokinetic stacking mechanism. C: concentration; E: applied electric field strength; LE: leading electrolyte; TE: terminating electrolyte; μ: mobility; ν: velocity.

In contrast to universally enriching all proteins by a similar degree, the result of the CITP stacking process is that major components may be diluted, but trace compounds are concentrated. Such selective enhancement toward low abundance proteins can drastically reduce the range of relative protein abundances in complex proteomes, and greatly improve proteome coverage using the CITP-based multidimensional protein/peptide separation technology.

Chapter 2 applied a multidimensional CITP/CZE-nano-RPLC separation platform coupled with ESI-tandem MS for the characterization of whole unstimulated human saliva. Besides initial evaluation of the CITP proteome technology, the foundation for saliva-based diagnostics is the development of a complete catalog of secreted proteins detectable in saliva. Besides protein complexity, the greatest challenge facing comprehensive analysis of saliva samples is related to the large variation of protein relative abundances including the presence of high abundance proteins such as amylases, mucins, praline rich proteins (PRPs), and secretory IgA complex.

A total of 6,112 fully tryptic peptides were sequenced at a 1% FDR, leading to the identification of 1,479 distinct human SwissProt protein entries. By comparing with CIEF as another electrokinetics-based stacking approach, CITP/CZE not only offered a broad field of application, but also was less prone to protein/peptide precipitation during the analysis. The ultrahigh resolving power of CITP/CZE was evidenced by the large number of distinct peptide identifications measured from each CITP fraction together with the low peptide fraction overlapping among identified peptides. Furthermore, when evaluating the protein sequence coverage by the number of distinct peptides mapping to each protein identification, the CITP-based proteome

technology similarly achieved the superior performance with 674 proteins (46%) having 3 or more distinct peptides, 288 (19%) having 2 distinct peptides, and 517 (35%) having a single distinct peptide.

In addition to their biological and medical significance, mitochondria are ideal targets for global proteome analysis because they have a manageable level of complexity as a consequence of their apparent prokaryotic ancestry. Still, proteins cataloged from reported proteomic studies comprise only small part of the estimated total human mitochondrial proteins. A CITP-based multidimensional separation platform, capable of providing selective analyte enrichment and high resolving power toward complex protein/peptide mixtures, was therefore employed in Chapter 3 for the analysis of synaptic mitochondria isolated from the brains of mice.

A total of 1,795 distinct mouse SwissProt protein entries were identified from synaptic mitochondria isolated from mouse brain. The ultrahigh resolving power of CITP/CZE was evidenced by the large number of distinct peptide identifications measured from each CITP fraction together with the low peptide fraction overlapping among identified peptides. The degree of peptide overlapping among CITP fractions was even lower than that achieved using combined CIEF/nano-RPLC separations for the analysis of the same mitochondrial sample. When evaluating the protein sequence coverage by the number of distinct peptides mapping to each mitochondrial protein identification, CITP/CZE similarly achieved superior performance with 1,041 proteins (58%) having 3 or more distinct peptides, 233 (13%) having 2 distinct peptides, and 521 (29%) having a single distinct peptide.

The reproducibility of protein identifications was found to be around 86% by comparing proteins identified from repeated runs of the same mitochondrial sample. The analysis of the mouse mitochondrial proteome by two CITP/CZE runs resulted in the detection of 2,095 distinct mouse SwissProt protein entries, corresponding to 59% coverage of the updated Maestro mitochondrial reference set. The collective analysis from combined CITP/CZE and CIEF-based proteomic studies yielded the identification of 2,191 distinct mitochondrial protein entries, corresponding to 76% coverage of the MitoP2-database reference set.

Comparisons among shotgun proteome technologies, including CITP-based multidimensional separations and multidimensional liquid chromatography system, were performed in Chapter 4 regarding their abilities to address the challenges of protein complexity and relative abundance inherent in GBM derived cancer stem cells. Comparisons were conducted using a single processed protein digest with equal sample loading, identical second dimension separation (nano-RPLC) and MS conditions, and consistent search parameters and cutoff established by the target-decoy determined FDR. Besides achieving superior overall proteome performance in total peptide, distinct peptide, and distinct protein identifications, analytical reproducibility of the CITP proteome platform coupled with the spectral counting approach was determined by a Pearson R^2 value of 0.98 and a coefficient of variation (CV) of 15% across all proteins quantified. In contrast, extensive fraction overlapping in strong cation exchange greatly limited the ability of multidimensional liquid chromatography separations for mining deeper into the tissue proteome as evidenced by the poor coverage in various protein functional categories.

Chapter 2: Comparison of Electrokinetics-Based Multidimensional Separations Coupled with Electrospray Ionization-Tandem Mass Spectrometry for Characterization of Human Salivary Proteins

2.1 Introduction

Whole saliva is derived predominantly from three paired major salivary glands, accounting for about 90% of fluid production, as well as from minor salivary glands. Besides playing a role in digestion, lubrication, and formation of a pellicle that protects teeth and mucosal surfaces, salivary components are important determinants of oral health and essential in regulating the oral ecology[26]. Alterations in the composition of the oral flora and/or immune dysfunction are often linked to common oral diseases. 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[27-33].

There has been growing interest in identifying "salivary biomarkers" as a means of monitoring general health and for the early diagnosis of diseases, such as bacterial infection, human immunodeficiency virus, and oral cancer, or systematic diseases that

affect the function of salivary glands, e.g. Sjogren's syndrome, alcoholic cirrhosis, cystic fibrosis, diabetes mellitus, diseases of the adrenal cortex, cardiovascular diseases, and dental caries[31, 34-36]. For example, differences in proteolytic processing of basic proline rich proteins (PRPs) from the human parotid gland have been associated with caries[37]. Saliva samples obtained from oral cancer patients have been profiled and compared with those from healthy control subjects[38]. Studying the expression of specific salivary proteins has demonstrated that the family of defensins exhibits antibiotic, antifungal, and antiviral properties[39-43]. Additionally, detailed knowledge of salivary composition will be helpful in the construction of artificial salivas for use by patients with xerostomia and salivary hypofunction. These conditions are commonly associated with head and neck radiotherapy, Sjogren's syndrome, and the use of antidepressants[44-46]. In fact, the efficacy and safety of including antimicrobial proteins in artificial salivas have recently been under evaluation[47].

Several research groups have employed the two-dimensional polyacrylamide gel electrophoresis-mass spectrometry (2-D PAGE-MS) technique for the analysis of whole saliva and parotid gland secretions and have identified less than 200 proteins, which only represents a small portion of the salivary proteome[48-53]. In addition to the separation of intact proteins using 2-D PAGE, Yates and co-workers[54, 55] developed and 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. A total of 102 and 266 proteins were identified from tryptic digests of human whole saliva using single[56] and multidimensional liquid chromatography[50],

respectively. Instead of using strong cation exchange chromatography as the first separation dimension, free flow electrophoresis has been coupled with reversed-phase liquid chromatography (RPLC) for achieving multidimensional peptide separations and leading to the identification of 437 salivary proteins[57]. The overlap among these cataloged proteins[50, 56, 57] was relatively small and included most of the common proteins present in high abundance and associated with housekeeping functions.

The greatest bioanalytical challenge facing comprehensive proteome analysis of saliva samples is related to the large variation of protein relative abundances. Similar to other body fluids such as serum and plasma, the large variation of protein relative abundances together with the complexity of the saliva proteome continuously challenges the development of proteome technologies from sample processing to separation and MS detection[58]. One obvious approach to enhance the proteome coverage and the detection capability of low abundance proteins is to significantly raise the required amounts of protein along with many additional sample preparation/fractionation/separation processes [59-62], including the depletion of high abundance proteins such as albumins in serum/plasma samples. For example, N-linked glycoproteins in human saliva were captured and enriched using the hydrazide coupling and release approach which functions as a pre-fractionation technique for handling highly complex biofluids[63]. A total of 84 N-glycosylated peptides were identified and associated with 45 unique N-glycoproteins, including 16 glycoproteins that have not been reported previously in saliva.

The high resolving power of capillary bioelectric focusing (CIEF) as the first dimension in combined CIEF/nano-RPLC separations[14, 17, 18] was evidenced by the low fraction overlapping among identified peptides. The percentage of identified

peptides present in more than one CIEF fraction was only around 20-25%[14, 17], significantly less than 40-80% obtained from multidimensional liquid chromatography system using strong cation exchange coupled with RPLC[55, 64]. By employing a CIEF-based multidimensional separation platform coupled with electrospray ionization (ESI)-tandem MS, a total of 3,642 fully tryptic peptides were sequenced at a 1% false discovery rate (FDR), leading to the identification of 1,165 distinct SwissProt protein entries from a single human saliva sample[65].

Instead of using CIEF as the first dimension, a capillary isotachophoresis (CITP)-based multidimensional separation platform, capable of providing selective analyte enrichment and high resolving power toward complex protein/peptide mixtures, is employed in this study to further address the challenges of protein complexity and protein relative abundance inherent in human saliva proteome. In contrast to universally enriching all analytes by a similar degree, the result of the CITP stacking process is that major components may be diluted, but trace compounds are concentrated[7, 66-68]. As demonstrated in our recent studies[7], such selective enhancement toward low abundance proteins can drastically reduce the range of relative protein abundances and greatly improve the sequence analysis of low abundance proteins in complex proteomes. Furthermore, CITP offers the benefits of speed and straightforward manipulation/switching between the stacking and separation modes in transient CITP/capillary zone electrophoresis (CZE)[7, 66].

2.2 Experimental Section

2.2.1 Material 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), iodoacetamide (IAM), and protease inhibitors were obtained from Sigma (St. Louis, MO). Acetonitrile, hydroxypropyl cellulose (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 Saliva 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 °C against 100 mM Tris at pH 8.2. 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 (Thermo Savant, San Jose, CA), and then stored at -80 °C.

2.2.3 Transient CITP/CZE-Based Multidimensional Separations

The CITP apparatus was constructed in-house using a CZE 1000R high-voltage power supply (Spellman High-Voltage Electronics, Plainview, NY). A 80-cm long CITP capillary (100 μm i.d./365 μm o.d.) coated with hydroxypropyl cellulose was initially filled a background electrophoresis buffer of 0.1 M acetic acid at pH 2.8. The sample containing saliva protein digests was prepared in a 2% pharmalyte solution. A 50 cm long sample plug, corresponding to 4.0 μl of sample volume, was injected into the capillary by pressure. 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.

UV absorbance detection at 214 nm was placed at 14 cm from the cathodic end of the capillary. The cathodic end of the capillary was housed inside a stainless steel needle using a coaxial liquid sheath flow configuration[69]. A sheath liquid composed of 0.1 M acetic acid was delivered by using a Harvard Apparatus 22 syringe pump (South Natick, MA). The stacked and resolved peptides in the CITP capillary were sequentially fractionated and loaded into individual wells on a moving microtiter plate.

To couple transient CITP/CZE with nano-RPLC, peptides collected in individual wells were sequentially injected 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. 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) 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). Open Mass Spectrometry Search Algorithm (OMSSA)[19] 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, 0.4 Da

fragment ion mass tolerance, 1 missed cleavage, alkylated Cys as a fixed modification and variable modification of Met oxidation. 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).

FDRs were determined using the method of Elias and co-workers[70]. 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 occurring as matches to the forward sequences were not counted as false positives. A curve was then generated by plotting FDR versus E-value and an E-value threshold corresponding to a 1% FDR for total peptide identifications was used as the cutoff in this study. OMSSA E-value is the expectation value that a tandem MS event represents the predicted peptide given a number of factors, including the quality of the match between the experimental and theoretical fragment ion spectra, the search space as determined by the sequence library, the precursor and fragment ion mass accuracy settings, enzyme specificity, and missed cleavages and modifications[19].

2.3 Result and Discussion

2.3.1 Optimization of Transient CITP/CZE Separations

One of key issues in performing CE separations involves the elimination of protein/peptide adsorption onto the capillary wall. The adsorption of proteins and peptides not only results in changes in the electroosmotic flow, but also contributes to

analyte band broadening. Application of polyacrylamide coating as described in our previous CIEF work[18] produced satisfactory performance in both the separation efficiency and the resolving power. However, the polyacrylamide coating has limited column life at acidic pH and the entire coating procedure is quite time consuming. Thus, the use of hydroxypropyl cellulose[71] as the capillary coating was evaluated and demonstrated for the simplicity in the coating protocol, the robustness in the multiple usage, and the superior performance in the elimination of electroosmotic pumping and protein/peptide adsorption.

For the separation of cationic analytes at acidic pHs, ammonium acetate has been reported as the most common leading electrolyte used in the CITP process[7, 66-68]. In contrast to the generation of a pH gradient in CIEF, carrier ampholytes such as commercially available pharmalyte 3-10, which exhibit high electrophoretic mobilities, were also investigated as the potential candidate for the leading electrolyte employed in transient CITP/CZE separations. By comparing with the use of ammonium acetate, pharmalyte ions not only offered lower electric currents to avoid the joule heating during the separation, but also greatly enhanced the solubility of stacked proteins and peptides.

It is also interesting to note that the final concentration of the analyte in CITP is independent from the applied electric voltage. However, the application of low electric field strengths would lead to poorer resolution among the stacked zones and longer migration time in CITP[7, 66-68]. We have determined to routinely apply an electric field strength of 300 V/cm across the entire capillary in order to reduce the joule heating while enhancing the overall resolving power.

Due to the limitation of a 30-kV high-voltage power supply available in our laboratory, a 80-cm long coated capillary was typically employed for performing transient CITP/CZE separations. The inner diameter of coated capillary was chosen from initial 50 μm to 100 μm in order to increase the sample loading by a four-fold. However, the joule heating, which also increases with increasing the inner diameter, prevented the use of even larger capillary columns in transient CITP/CZE.

Furthermore, the use of formic acid has been reported to greatly improve the solubility of hydrophobic proteins and peptides[72, 73]. Besides the use of acetic acid, formic acid was also employed as the terminating electrolyte in the background electrophoresis buffer and investigated for its effect on transient CITP-CZE separations while enhancing analyte solubility in the stacked zones. However, there was no apparent enhancement in the solubility of protein digests associated with the use of formic acid as the terminating electrolyte. The presence of pharmalyte ions as the leading electrolyte may already provide sufficient enhancement in peptide solubility.

By comparing with CIEF, CITP offers a broader field of application, particularly for protein separation and concentration. Proteins with extreme pI values may be outside the working pH range of CIEF due to limited availability of commercial ampholytes for the creation of a pH gradient inside the capillary. Furthermore, proteins focused in CIEF reside at their pIs where they have an increased tendency to precipitate. In contrast, the stacked proteins in CITP are less prone to precipitation due to their charged nature.

2.3.2 Analysis of Saliva Sample by CITP-Based Proteome Technology

An important aspect of any multidimensional proteome separation platform is its ability to improve the detection of analytes present in low quantities during the analyses of complex

protein/peptide mixtures. As opposed to universally enriching all proteins by a similar degree in most sample concentration techniques, the CITP stacking process specifically targets trace amounts of proteins[7, 66-68] and thus reduces the range of relative protein abundances for providing unparallel advantages towards the identification of low abundance proteins. Such selective proteome enrichment enabled by the CITP process can be represented by the Kohlrausch equation[68]:

$$C_A = C_L / \{[\mu_L/(\mu_L + \mu_R)] \ [(\mu_A + \mu_R)/ \mu_A]\} \tag{1}$$

Where C_A is the analyte concentration in the stacked zone; C_L is the molarity of the leading electrolyte; and μ is the electrophoretic mobility (the subscript R refers to the counter ion). It can be concluded that the final concentration of the analyte, C_A, is largely proportional to the molarity of the leading buffer, C_L. Depending on the initial concentrations of individual sample components relative to the molarity of the leading buffer, the result of the CITP stacking process is that major compounds may be diluted, but trace peptides are concentrated.

As a selective enrichment technique, it is also interesting to note that the final concentration of the analyte in CITP is independent from the applied electric voltage. However, the application of low electric field strengths would lead to poorer resolution among the stacked zones and longer migration time in CITP[7, 66-68]. In order to obtain excellent separation resolution in CITP, the operational electrolyte system should be optimized to give appropriate differences among the effective mobilities of analytes. As the pH and ionic strength of the stacked zones in CITP are different from each other, optimization of the separation in CITP is not straightforward in comparison with CZE.

Thus, on-column transition of CITP to CZE was employed as the first separation dimension in this study to combine selective analyte enrichment with improved zone resolution. Two important requirements[7, 66] were fulfilled in order to induce transient CITP/CZE separation of complex proteome mixtures. The mobility of the co-ion of the background electrolyte (0.1 M acetic acid at pH 2.8) was lower than the sample components (tryptic digest of human salivary proteins) during the transient CITP step so that it served as the terminating ion. The proteome sample also contained an additional co-ion (2% pharmalyte)[74, 75] with high electrophoretic mobility as the leading electrolyte. Under the influence of applied electric field, the higher mobility pharmalyte ions rushed to the frontal boundary of the sample zone and sample stacked between the high-mobility leading electrolyte and the low-mobility terminating electrolyte in the background electrophoresis buffer. Once the leading electrolyte has migrated away, and the peptides were completely stacked, transient CITP was followed by CZE of concentrated peptides in the background electrolyte.

In addition to a transient CITP/CZE-UV trace, the number of distinct peptides identified from each of the CITP/CZE fractions is summarized in Figure 2-1. In all, 24 peptide fractions were sampled from the first dimension of transient CITP/CZE separation with a total peptide loading of only 8.0 μg (injected volume of 4.0 μL x peptide concentration of 2.0 μg/μL).

Each peptide fraction was further resolved by the second dimension nano-RPLC separation and the eluants were detected using ESI-tandem MS. As shown in Figure 2-2, the peptide and protein FDRs and the numbers of total peptides, distinct peptides, and protein identifications were plotted as functions of the E-value of a typical OMSSA search.

OMSSA E-value is the expectation value that a tandem MS event represents the predicted peptide given a number of factors, including the quality of the match between the experimental and theoretical fragment ion spectra, the search space as determined by the sequence library, the precursor and fragment ion mass accuracy settings, enzyme specificity, and missed cleavages and modifications. An E-value threshold of 0.003, corresponding to 1% FDR of total peptide identifications, was chosen as the cutoff in this study and led to the identification of 6,112 fully tryptic peptides covering 1,479 distinct human SwissProt protein entries. This cutoff led to a protein FDR of 8.8% as indicated by the detection of peptides from 66 distinct reversed protein sequences in the decoy section of the search database.

By increasing the number of CIEF fractions from 29 in our previous study[65] to 54 in this work, a total of 4,589 fully tryptic peptides were sequenced at a 1% FDR of total peptide hits, leading to the identification of 1,374 distinct SwissProt protein entries from the same saliva sample. By increasing the number of CIEF fractions in combined CIEF/nano-RPLC separations, the numbers of distinct peptides and protein identifications were clearly higher than those (3,642 fully tryptic peptides and 1,165 distinct proteins) reported previously[65]. Even though only 24 transient CITP/CZE fractions were taken in this study (Figure 2-1), the CITP-based multidimensional separation platform, equipped with selective analyte enrichment, yielded the largest catalog of proteins measured from a single saliva sample to date.
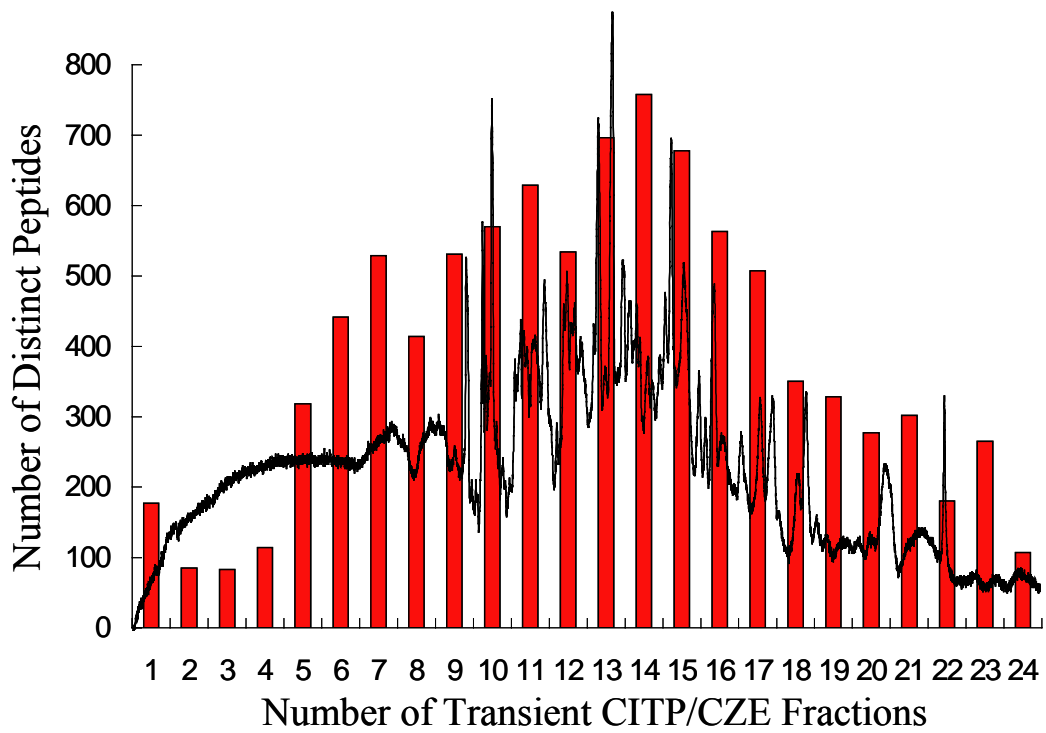
Figure 2-1 Overlaid plots containing the transient CITP/CZE-UV trace monitored at

214 nm and the number of distinct peptides identified in each of the 24

CITP/CZE fractions.
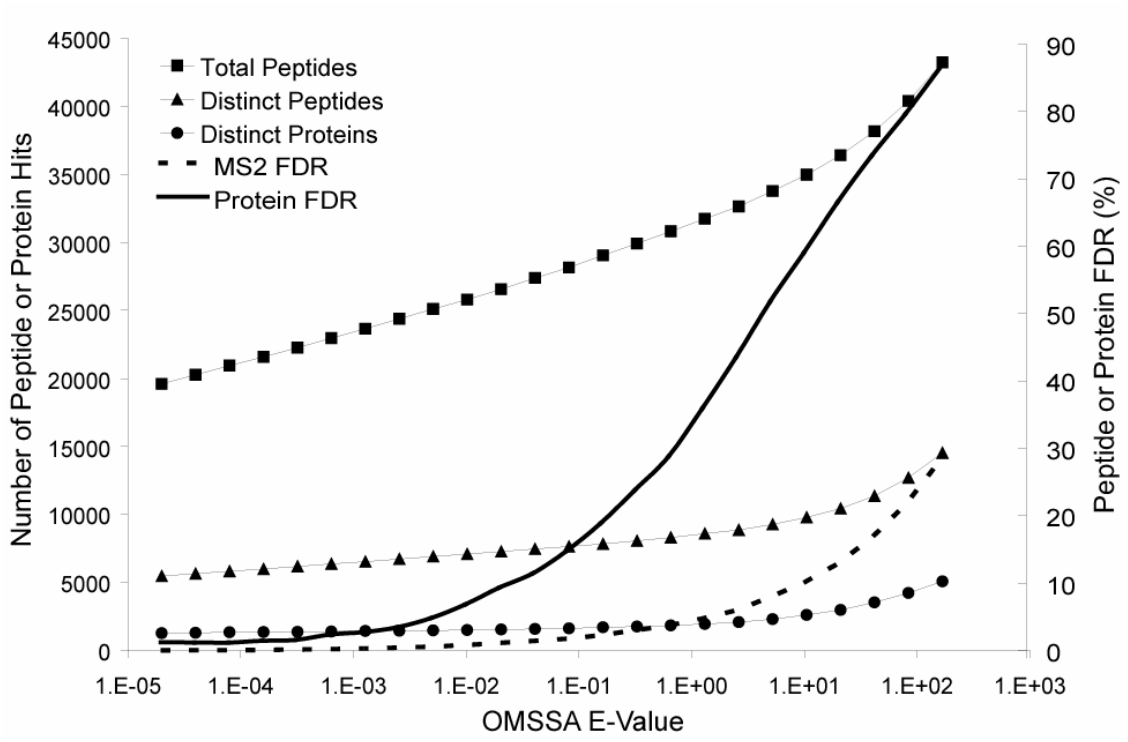
Figure 2-2 Plots of the FDRs and the numbers of total peptide, distinct peptide, and distinct protein identifications versus the E-value obtained from the search of the peak list files against a decoyed SwissProt human database using OMSSA.

2.3.3 Comparison of CITP- and CIEF-Based Proteome Technology

By comparing with the CIEF-based proteome technology[14, 17, 18], transient CITP/CZE separations offer a broader field of application. Peptides with extreme pI values may be outside the working pH range of CIEF due to limited availability of commercial ampholytes for the creation of a pH gradient inside the capillary. Furthermore, analytes focused in CIEF reside at their pIs where they have an increased tendency to precipitate. In contrast, the stacked peptides in CITP are less prone to precipitation due to their charged nature.

As shown in Figure 2-3, the proteome results obtained from combined transient CITP/nano-RPLC separations are further compared with those achieved using 54 CIEF fractions. A total of 1,121 human saliva proteins were identified by both the CITP- and CIEF-based multidimensional separation platforms coupled with ESI-tandem MS, corresponding to greater than 81% protein overlapping. Small proteins with a molecular mass of less than 20 kDa constitute a characteristic portion of the human saliva proteome. A total of 361 small proteins were identified by combining the results attained from both the CITP and CIEF proteome platforms. A number of these small proteins are PRPs which are involved in creating a protective film over teeth and promoting remineralization of the enamel. Additionally, several antimicrobial proteins are also identified including β-defensin 125 precursor, β-defensin 1 precursor, FALL-39 precursor, mucins 5B and 2, 6 different cystatin precursors, and lysozyme.

Distinct Proteins

CITP 358 1121 253 CIEF

Small Proteins

CITP 58 267 36 CIEF
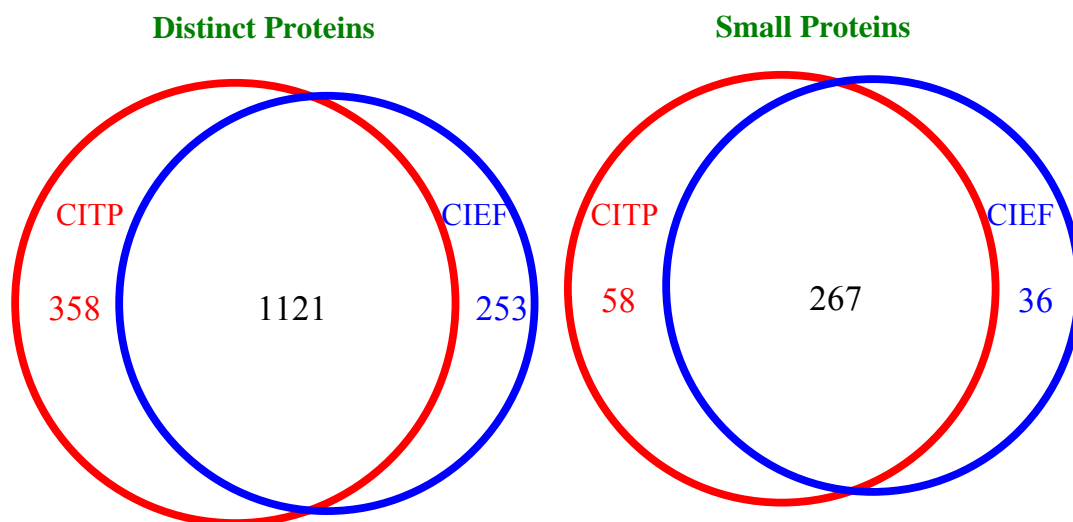
Figure 2-3 Venn diagrams comparing the saliva proteome results obtained from the (A) CITP- versus (B) CIEF-based multidimensional separation platforms.

The ultrahigh resolving power of transient CITP/CZE is evidenced by the large number of distinct peptide identifications measured from each CITP fraction (Figure 2-1) with even lower peptide fraction overlapping than that typically observed in the CIEF separation (Figure 2-4). In this case, approximately 85% of distinct peptides were identified in only a single CITP fraction. Of the remaining peptides, 8% were identified in two fractions and 7% were measured in three and more fractions. In comparison with the multidimensional liquid chromatography system[55, 64], a high degree of peptide overlapping in the first dimension of strong cation exchange chromatography 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 chromatography 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 the ESI-MS analysis.

When evaluating the protein sequence coverage by the number of distinct peptides mapping to each protein identification, CITP similarly achieves the best performance with 674 proteins (46%) having 3 or more distinct peptides, 288 (19%) having 2 distinct peptides, and 517 (35%) having a single distinct peptide (Figure 2-5). The CIEF-based multidimensional separation returns 38% of proteins with 3 or more distinct peptides, 20% with 2 distinct peptides, and 42% with single peptide hits. Another measure of proteome data quality includes the ratio of total peptide to distinct protein identifications which are 17.19 (25,428 total peptides/1,479 distinct proteins) and 9.85 (13,534 total peptides/1,374 distinct proteins) for the CITP- and

CIEF-based proteome platforms, respectively. This ratio becomes increasingly important when implementing the spectral counting-based protein quantification approach[76], as the expression levels of each protein are determined by the number of tandem MS events.

As demonstrated by the common observation in the literature, most false peptide identification tend to be ones in which the corresponding proteins is only identified by a single peptides. Thus, many proteomic research laboratories routinely discard single-peptide identifications to significantly reduce the FDR of distinct protein identifications. However, as pointed out by the work of Elias and Gygi [77], these single-protein identifications after filtering are mostly correct and typically represent 30-50% of a proteome data set. The removal of these protein identifications greatly and negatively impacts the coverage and the sensitivity of overall proteome analysis. As advocated by Elias and Gygi [77] and also implemented in this study, the target-decoy search strategy is employed as the routine practice to design more stringent criteria for single-peptide identifications.

Due to selective analyte enrichment of transient CITP/CZE separations, there is also a higher proportion of signaling proteins, including kinases, identified by the CITP-based proteome technology. Representative kinases including cAMP-dependent protein kinase type II-α regulatory subunit (P13861), heat-shock protein β-8 (Q9UJY1), and hexokinase type III (P52790) are presented in Figure 2-6 together with corresponding tandem mass spectra of distinct peptides leading to their identifications.
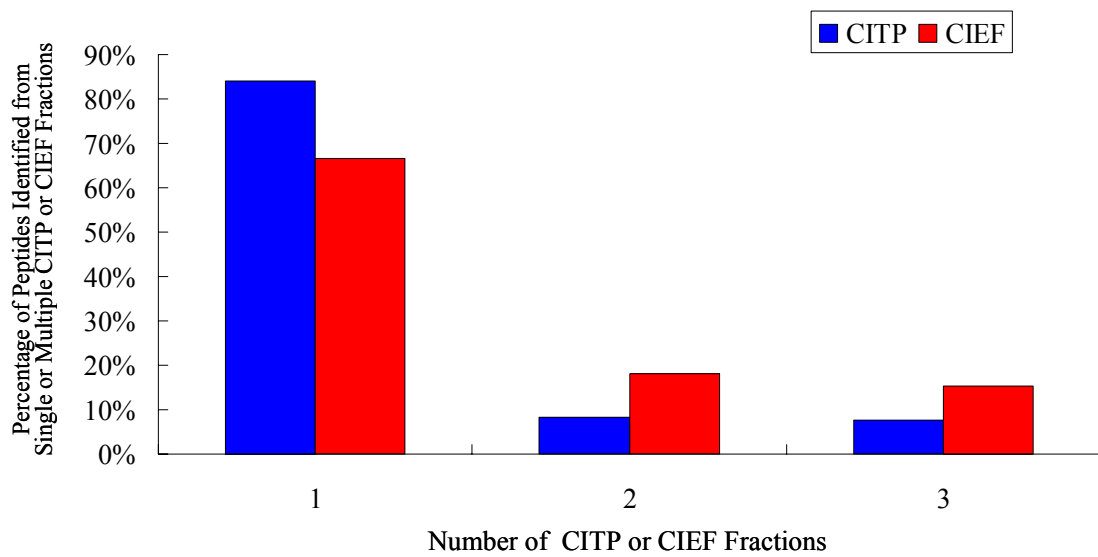
Figure 2-4 The percentage of distinct peptides identified from single or multiple CITP

or CIEF fractions.
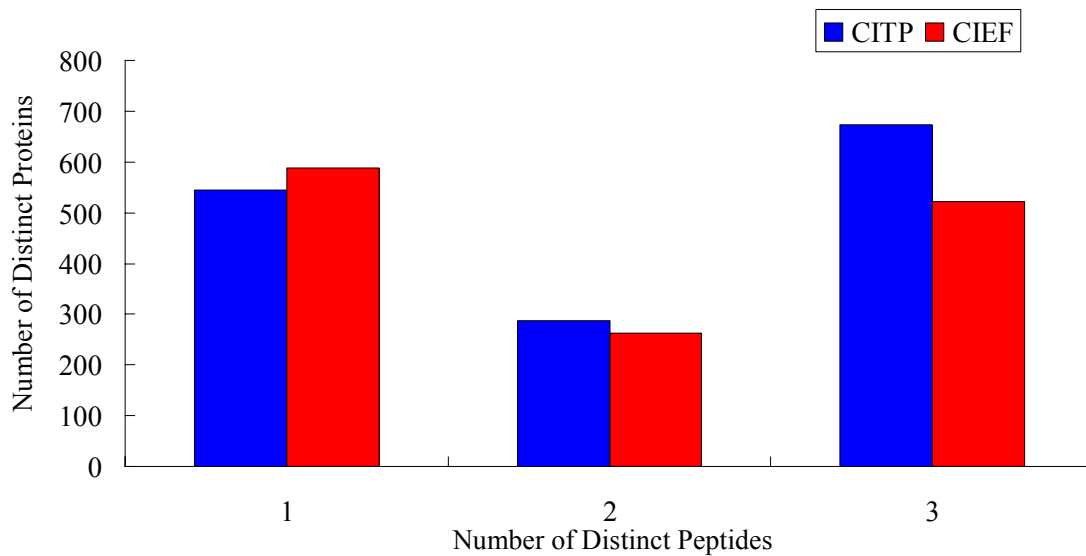
Figure 2-5 The number of distinct proteins identified by single or multiple distinct

peptides measured from the CITP and CIEF-based proteome platforms.
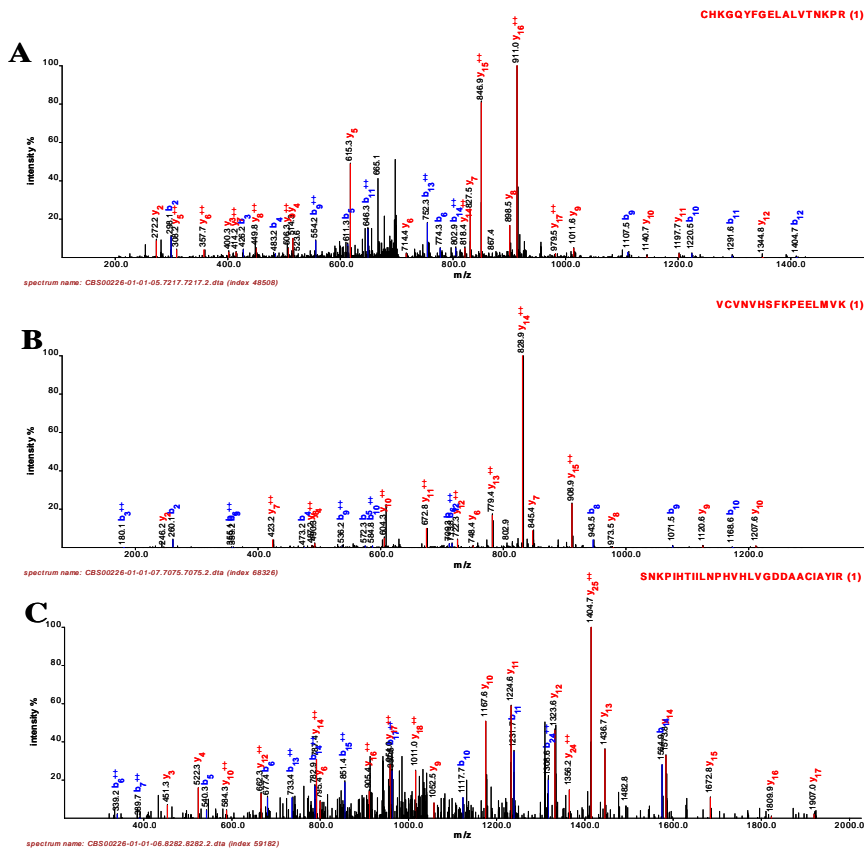
Figure 2-6 Tandem MS spectra of distinct peptide hits contributing to the identification of representative kinases including (A) cAMP-dependent protein kinase type II-*α* regulatory subunit, (B) heat-shock protein *β*-8, and (C) calcium/calmodulin-dependent protein kinase type II.

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. 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. This work presented our continuous efforts toward the identification and cataloging of salivary secretory components using combined CITP/nano-RPLC separations coupled with ESI-tandem MS. Compared with CIEF utilized as the first separation dimension in our previous studies[17, 18, 65], transient CITP/CZE not only contributed to selective analyte enrichment toward low abundance proteins, but also provided high resolving power for addressing the complexity of the saliva proteome. The ultrahigh resolving power of transient CITP/CZE is evidenced by the large number of distinct peptide identifications measured from each CITP fraction together with the low peptide fraction overlapping among identified peptides.

A total of 6,112 fully tryptic peptides were detected at a 1% FDR, leading to the identification of 1,479 distinct SwissProt protein entries from a single human saliva sample. A hit rate (total peptide identifications over tandem MS queries) of 0.11 was achieved over approximately 232,600 MS2 queries in a single CITP-based proteome measurement. The remaining 89% of queries included noise, contaminants, non-peptide biological components, post-translational modifications that were not searched, peptides with cleavage events that were not searched, and, frequently, fully tryptic peptides that were below the accepted threshold for identification. It has been shown that improved mass accuracy can lower the threshold at which a peptide hit is

considered correct[78]. Furthermore, it has been suggested that the application of ultrahigh mass accuracy and resolution from a linear ion trap-Orbitrap hybrid mass spectrometer may allow the elimination of virtually any false positive peptide identifications[79]. An as yet largely unexplored option for improving peptide search sensitivity is the use of spectral matching[80, 81], in which the tandem MS spectral signature of a high quality peptide hit is used as the reference comparison during the search, rather than a predicted fragmentation pattern. This spectral matching technique has long been utilized in the field of gas chromatography-MS and is widely accepted for the classification of small molecules.

2.5 Acknowledgement

Chapter 3: Application of Capillary Isotachophoresis-Based Multidimensional Separations Coupled with Electrospray Ionization-Tandem Mass Spectrometry for Characterization of Mouse Brain Mitochondrial Proteome

3.1 Introduction

Mitochondria have a variety of essential responsibilities in cellular metabolism, e.g. the production of ATP through oxidative phosphorylation in aerobic cells as well as the initiation of the signal cascade leading to apoptosis [82]. Alteration of the mitochondrial proteome and altered mitochondrial function has been implicated in a variety of degenerative diseases, heart disease, aging, and cancer [83-85]. Because the mitochondria are a major source of endogenously generated reactive oxygen species (ROS), it is not surprising that they have been proposed to play a major role in aging [86, 87]. The free radical theory of aging predicts that oxidative damage to the mitochondria can lead to an amplifying effect whereby damaged mitochondria release more ROS, further increasing oxidative damage [88]. The accumulation of damaged mitochondria results in a decrease in the capacity to produce ATP. This can, in turn, cause cellular dysfunction if oxidative

damage exceeds the capacity of the cell to handle removal of both ROS and oxidatively damaged macromolecules [89].

In addition to their biological and medical significance, mitochondria are ideal targets for global proteome analysis because they have a manageable level of complexity as a consequence of their apparent prokaryotic ancestry. Their endosymbiotic origins have been preserved in their double membrane structure, and they possess their own circular genome with mitochondria-specific transcription, translation, and protein assembly systems. Mammalian mitochondria DNA encodes for 13 essential polypeptide components of the oxidative phosphorylation system, small and large rRNAs, and 22 tRNAs. The remaining mitochondrial proteins of the oxidative phosphorylation system, metabolic enzymes, membrane channel proteins, and any other proteins regulating mitochondrial function are derived from the nuclear genome. Import of these proteins into the mitochondria is highly regulated through the function of protein chaperones and the inner and outer transmembrane peptide import complexes [90, 91]. However, the coordination of protein synthesis from two separate genomes within a cell and their correct submitochondrial assembly is not fully understood [92].

While two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) has been employed toward an inventory of the mitochondrial proteome [89, 93-98], mitochondria contain many membrane-bound and very basic proteins that are quite difficult to be analyzed by 2-D PAGE [92, 94]. Consequently, considerable efforts have been devoted to the application of various liquid chromatography (LC) techniques in single or multidimensional separation format prior to mass spectrometry (MS) detection for the analysis of mitochondrial proteins [96-110]. In particular, the peptide-based shotgun proteomic

studies fully exploit the resolution and sensitivity achievable with multidimensional LC-MS or sodium dodecyl sulfate (SDS)-PAGE/LC-MS approaches, allowing many additional mitochondrial proteins to be identified.

Still, proteins cataloged from reported proteomic studies comprise only small part of the estimated 1,500 [111, 112] and 697-4,532 [113] total human mitochondrial proteins. The actual number of mitochondrial proteins is difficult to determine as these estimations can have a false discovery rate (FDR) of up to 68% [113]. Even with reduced protein complexity in the subcellular proteome, the goal of achieving comprehensive proteome analysis is still challenged by the large variation of protein relative abundances. A capillary isotachophoresis (CITP)-based multidimensional separation platform, capable of providing selective analyte enrichment [7, 66-68, 114] and high resolving power toward complex protein/peptide mixtures, is therefore employed in this study for the analysis of synaptic mitochondria isolated from the brains of mice.

3.2 Experimental Section

3.2.1 Material 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, bovine serum albumin (BSA), dithiothreitol (DTT), ethylene glycol tetraacetic acid (EGTA), N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), iodoacetamide (IAM), mannitol, and sucrose were obtained from Sigma (St. Louis, MO). Acetonitrile, hydroxypropyl cellulose (average MW 100,000), SDS, Tris(hydroxymethyl)aminomethane (Tris), and urea were purchased from Fisher Scientific

(Pittsburgh, PA). Pharmalyte 3-10 and Percoll were acquired from Amersham Biosciences (Piscataway, NJ). 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).

3.2.2 Isolation of Synaptic Mouse Brain Mitochondria

Animal experiments were performed in accordance with the Guide for the Care and Use of Laboratory Animals and approved by the University of Maryland Institutional Animal Care and Use Committee. Male (20 g) wild-type C57BL/6 mice were used. Mice were decapitated, and brains were removed and homogenized in ice-cold isolation medium containing 225 mM mannitol, 75 mM sucrose, 5 mM HEPES, and 1 mM EGTA, pH 7.4, at 4 ^{0}C. As described previously [115], a slightly modified procedure by Dunkely et al. [116] was employed to separate synaptosomes and nonsynaptic mitochondria. Briefly, after low-speed spin (1300 g for 3 min) of the brain homogenate, the supernatant was centrifuged at 21,000 g for 10 min. The pellet was suspended in 3% Percoll and layered on top of a preformed gradient of 24% (3.5 mL), 15% (2 mL), 10% (1.5 mL) of Percoll and centrifuged at 32,000 g for 8 min. Nonsynaptic mitochondria sedimented at the bottom of the tube and the synaptosomes accumulated at the interface of 24 and 15% Percoll and also at the interface of 15 and 10% Percoll.

The synaptosomes were collected and the mitochondria were released from synaptosomes using a nitrogen cavitation technique [117, 118]. The suspension was layered onto a preformed gradient of 40% (1.5 mL) and 24% (3.5 mL) Percoll. The mitochondria accumulated at the interface of 40 and 24% Percoll gradient. After the mitochondria fraction was collected and diluted with isolation medium, it was centrifuged to pellet purified synaptic

mitochondria. Synaptic mitochondria were suspended in isolation medium containing 1 mg/mL BSA to remove free fatty acid from mitochondrial membranes. The purity of synaptosome and mitochondria isolated from the Percoll gradient was examined with electron microscopy and demonstrated in our previous studies [115-118].

3.2.3 Mitochondrial Sample Preparation

After centrifugation at 7,500 g, the mitochondrial pellet was suspended and treated with a solution containing 1% SDS, 50 mM Tris (pH 8.0), and 1 mM DTT. After centrifugation at 20,000 g for 30 min, the supernatant was collected as the SDS-solubilized mitochondrial proteins. The SDS-solubilized mitochondrial proteins were placed in a dialysis cup and dialyzed overnight at 4 ^{0}C against 100 mM Tris at pH 8.0. The dialyzed proteins were denatured, reduced, and alkylated by sequentially adding urea, DTT, and IAM with final concentrations of 8 M, 1 mg/mL, and 2 mg/mL, respectively. The solution was incubated at 37 ^{0}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 ^{0}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 ^{0}C. Approximately 100 - 300 μg of mitochondrial proteins were obtained from each of the whole mouse forebrain (brain without the cerebellum).

3.2.4 Transient CITP/Capillary Zone Electrophoresis (CZE)-Based Multidimensional Separations

The CITP apparatus was constructed in-house using a CZE 1000R high-voltage power supply (Spellman High-Voltage Electronics, Plainview, NY). A 80-cm long CITP capillary (100 μm i.d./365 μm o.d.) coated with hydroxypropyl cellulose was initially filled a background electrophoresis buffer of 0.1 M acetic acid at pH 2.8. The sample containing mitochondrial protein digests was prepared in a 2% pharmalyte solution and 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.

UV absorbance detection at 214 nm was placed at 14 cm from the cathodic end of the capillary. The cathodic end of the capillary was housed inside a stainless steel needle using a coaxial liquid sheath flow configuration [69]. A sheath liquid composed of 0.1 M acetic acid was delivered by 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.

To couple transient CITP/CZE with nano-reversed phase liquid chromatography (nano-RPLC), peptides collected in individual wells were sequentially injected into individual trap columns (3 cm x 200 μm i.d. x 365 μm o.d.) packed with 5 μm porous C$_{18}$ 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) 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.

3.2.5 Data Analysis

Raw LTQ data were converted to peak list files by msn_extract.exe (ThermoFinnigan). OMSSA [19] was used to search the peak list files against a decoyed *Mus musculus* subset of the UniProt sequence library (April 20, 2006). This database was constructed by reversing all 41,178 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).

FDRs were determined using the method of Elias and co-workers [70]. 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. Similar to those described and discussed in our previous work [119], an E-value threshold corresponding to a 1% FDR for total peptide identifications was used as the cutoff in this study.

3.3 Result and Discussion

While CITP has been widely used for analyte preconcentration prior to electrophoretic separation [66-68], the application of CITP to selectively enrich trace amounts of proteins/peptides in complex mixtures presents an untapped strategy for maximizing proteome coverage. As demonstrated in our recent studies [7, 114], such selective enhancement toward low abundance proteins can drastically reduce the range of relative protein abundances and greatly improve the sequence analysis of low abundance proteins in complex proteomes such as human saliva proteome [114]. Furthermore, on-column transition of CITP to CZE was employed in this study in order to optimize selective analyte enrichment while achieving excellent separation resolution.

To induce transient CITP/CZE separation [7, 66], the mobility of the co-ion of the background electrolyte (0.1 M acetic acid at pH 2.8) was lower than the sample components (tryptic digest of mitochondrial proteins) during the transient CITP step so that it served as the terminating ion. The mitochondrial sample contained an additional co-

ion (2% pharmalyte) [74, 75] with high electrophoretic mobility as the leading electrolyte. Under the influence of applied electric field, the higher mobility pharmalyte ions rushed to the frontal boundary of the sample zone and sample stacked between the leading and terminating electrolytes. Transient CITP was followed by CZE of concentrated peptides in the background electrolyte. Stacked peptides were resolved in CZE based on their differences in electrophoretic mobility.

A representative transient CITP/CZE-UV trace is shown in Figure 3-1. In all, 20 peptide fractions were sampled from the first dimension of the transient CITP/CZE separation with a total peptide loading of only 15 μg (injected sample volume of 4.0 μL x peptide concentration of 3.75 μg/μL). Each peptide fraction was further resolved by nano-RPLC as the second separation dimension and the chromatography eluants were analyzed using ESI-tandem MS. By using a target-decoy search strategy [70, 119], an E-value threshold of 0.0013, corresponding to 1% FDR of total peptide identifications, was chosen as the cutoff in this study and led to the identification of 12,110 fully tryptic peptides covering 1,795 distinct mouse SwissProt protein entries. In this study, tandem MS spectra were searched against a decoyed *Mus musculus* subset of the UniProt sequence library (April 20, 2006). The UniProt sequence library consists of entries from both SwissProt and TrEMBL. Only protein hits mapping to the SwissProt subset are reported here in order to reduce the redundancy of the protein search results. To highlight the relatively low degree of protein redundancy in the SwissProt database employed in this study, these 1,795 distinct mouse protein entries consisted of 1,705 non-redundant proteins.
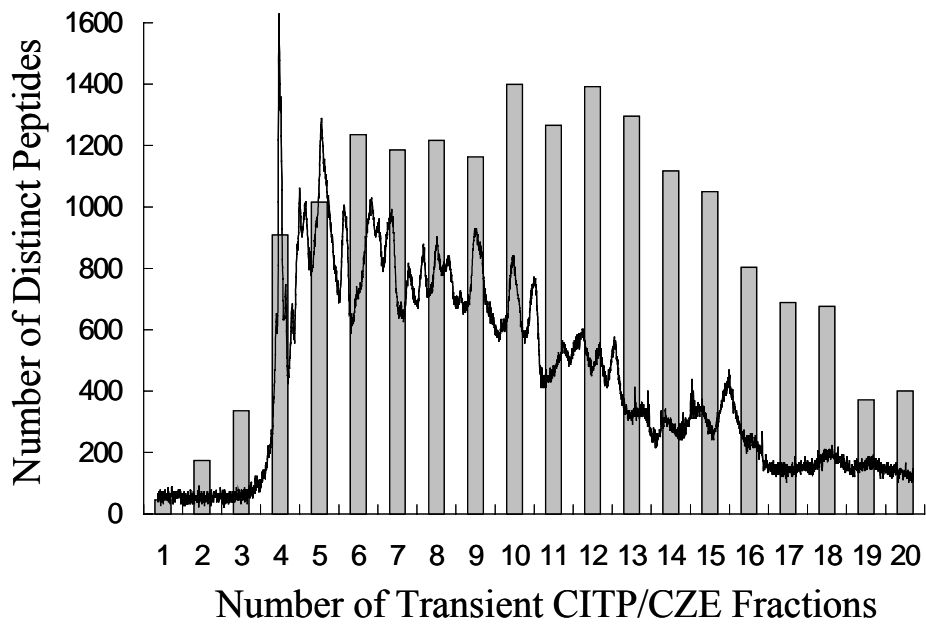
Figure 3-1 Overlaid plots containing the transient CITP/CZE-UV trace monitored at

214 nm and the number of distinct peptides identified in each of the 20

CITP/CZE fractions.

In addition to a representative transient CITP/CZE-UV trace, Figure 3-1 also summarizes the number of distinct peptides identified from each of the 20 CITP/CZE fractions. The ultrahigh resolving power of transient CITP/CZE is clearly demonstrated by the large number of distinct peptide identifications measured from each CITP fraction together with significantly low peptide fraction overlapping shown in Figure 3-2. For the analysis of tryptic peptides of mitochondrial proteins, approximately 78% of distinct peptides were identified in only a single CITP fraction. Of the remaining peptides, 15% were identified in two fractions and 7% were measured in three and more fractions.

As shown in Figure 3-2, the degree of peptide overlapping among CITP fractions is even lower than that achieved using combined capillary isoelectric focusing (CIEF)/nano-RPLC separations [14, 17, 18, 65] for the analysis of the same mitochondrial sample. As observed in our previous studies [14, 17, 18, 65], the isoelectric point (pI) of basic peptides may not be well defined and the basic peptides do not focus as tightly as those of acidic counterparts during the CIEF separation. In comparison with the shotgun-based multidimensional LC system [55, 64], a high degree of peptide overlapping was typically observed with 40-80% of carry over peptides that were identified in previous salt steps in the first dimension of strong cation exchange chromatography.
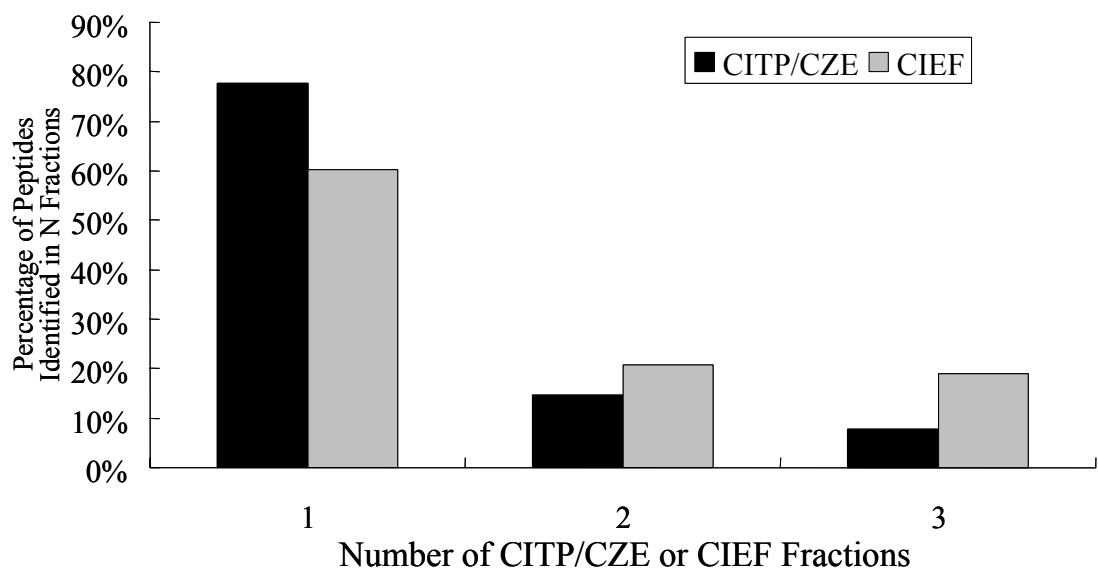
Figure 3-2 The degree of peptide overlapping among CITP/CZE or CIEF fractions evaluated as the percentage of peptides identified in one, two, and three fractions.

When evaluating the protein sequence coverage by the number of distinct peptides mapping to each mitochondrial protein identification, CITP/CZE similarly achieved superior performance with 1,041 proteins (58%) having 3 or more distinct peptides, 233 (13%) having 2 distinct peptides, and 521 (29%) having a single distinct peptide. The CIEF-based mitochondrial proteome analysis returned 53% of proteins with 3 or more distinct peptides, 16% with 2 distinct peptides, and 31% with single peptide hits. Another measure of proteome data quality includes the ratio of total peptide hits to distinct protein identifications which were 22.4 (40,116 total peptide hits/1,795 distinct protein) for the CITP/CZE-based proteome platform. This ratio becomes increasingly important when implementing the spectral counting-based protein quantification approach [76], as the expression levels of each protein are determined by the number of tandem MS events.

As reported by the common observation in the literature, most false peptide identifications tend to be ones in which the corresponding protein is only identified by a single peptide. Thus, many proteomic research laboratories routinely discard single peptide identifications to significantly reduce the FDR of distinct protein identifications. However, as pointed out by the work of Elias and Gygi's [77] and our laboratories [119], these single-protein identifications after filtering are mostly correct and often represent 30-50% of a proteome data set. The removal of these protein identifications greatly and negatively impacts the coverage and the sensitivity of overall proteome analysis. As advocated by Elias and Gygi [77] and also implemented in this study, the target-decoy search strategy is employed as the routine practice to design more stringent criteria for single-peptide identifications.

The same protein digest obtained from synaptic mouse brain mitochondria was repeatedly analyzed by both the CITP/CZE and CIEF-based multidimensional separation platforms equipped with ESI-tandem MS. The percentage of overlapping proteins obtained from repeated CITP/CZE-nano-RPLC runs was around 86% as illustrated by the Venn diagram shown in Figure 3-3A. A total of 2,095 distinct mouse SwissProt protein entries (or 1,992 non-redundant proteins) were identified from combined CITP/CZE-nano-RPLC runs of a single mouse brain mitochondria sample. On the other hand, the repeated CIEF-nano-RPLC proteomic studies returned a total of 1,543 distinct protein identifications (or 1,465 non-redundant proteins) with greater than 87% of protein overlapping among replicates (Figure 3-3B). As shown in Figure 3-3C, more than 93% of CIEF-cataloged mitochondrial proteins were inclusive in the results of the CITP/CZE proteomic studies. There were 96 and 648 proteins uniquely identified by the CIEF and CITP/CZE-based proteomic platforms, respectively.

Due to the use of electrokinetic migration/mobilization of charged analytes in transient CITP/CZE, the collected CITP fractions for subsequent nano-RPLC separations are less prone to contamination such as non-ionic surfactants or polymers employed during various sample storage and preparation processes. Significant enhancement in the overall peak capacity of the CITP-based proteomic platform can be realized by increasing the number of CITP fractions. The intrinsic high resolution of transient CITP-CZE allows the number of fractions sampled in the first separation dimension to be further increased by simply increasing the number of deposits collected in a microtiter plate.
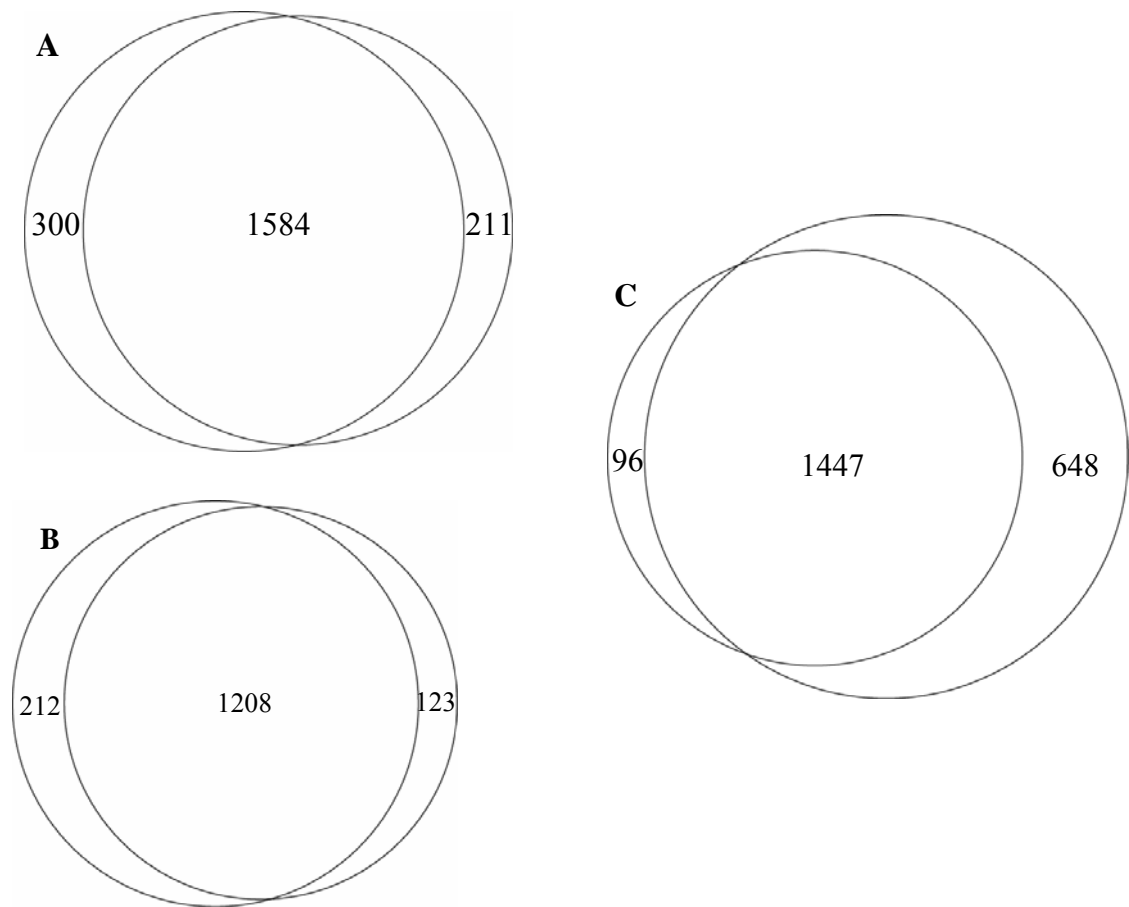
Figure 3-3 Venn diagrams comparing the mouse mitochondrial proteome results obtained from repeated (A) CITP/CZE and (B) CIEF analyses. (C) Comparing the combined proteomic results obtained from repeated CITP/CZE runs with those achieved by CIEF replicates.

The excellent reproducibility of transient CITP-CZE together with straightforward CITP fractionation using a robust sheath liquid interface [69] enables user discretion as to the number of fractions to be gathered and analyzed, and which interesting fractions may be revisited (reanalyzed) or repeatedly accumulated for identification of extremely low abundance peptides and proteins.

Combined proteomic results obtained from two CITP/CZE-nano-RPLC runs of a single mouse brain mitochondria sample (Figure 3-3A) were compared to the mitochondrial protein reference set predicted by Maestro [113]. By using a threshold score of 4.69 (corresponding to a 10% FDR), the Maestro reference set contained a list of 1,207 mouse mitochondrial proteins. Among these 1,207 proteins, a total of 198 proteins were no longer listed in the updated UniProt sequence library (April 20, 2006) and were subtracted from the reference dataset. Combined CITP/CZE analyses covered 59% (598 proteins) and 41% (413 proteins) of the updated Maestro mitochondrial reference set on the basis of proteins identified by a single and at least two distinct peptides, respectively.

In contrast, only 178 out 1,075 mouse brain mitochondrial proteins identified by Kislinger et al. [108] overlapped with the updated Maestro mitochondrial reference set, corresponding to less than 18% coverage. Besides significant increase in the coverage of the Maestro reference set, the ratios of mitochondrial proteins (included in Maestro) to total protein identifications in our proteomic results ranged from 0.26 (413/1,571) to 0.29 (598/2,095) on the basis of proteins identified by a single and at least two distinct peptides, respectively. These ratios were again higher than 0.16 (178/1075) achieved by Kislinger and co-workers [108].

The combined CITP/CZE and CIEF dataset was compared to the mitochondrial protein reference set of the MitoP2-database [120] containing a list of 731 mouse mitochondrial proteins. 553 out 2,191 distinct proteins (or 2,082 non-redundant proteins) presented in this study overlapped with the MitoP2-database, corresponding to 76% mouse mitochondrial proteome coverage. On the other hand, approximately 45% of the proteins identified by Forner and colleagues [109] (a total of 887 proteins collected from rat tissues of muscle, heart, and liver) had an apparent orthologue against the 2004 mouse MitoP database [121]. The proteomic results of mouse brain mitochondria achieved by Kislinger et al. [108] were also compared to the mitochondrial protein reference set of the MitoP2-database. Only 146 out 1,075 cataloged proteins overlapped with the MitoP2-database, corresponding to less than 20% coverage. It is interesting to note that the entirety of these 146 proteins and 90% of proteins collected from the work of Kislinger et al. [108] were included in the combined CITP/CZE and CIEF dataset.

By employing the mitochondrial proteome dataset obtained from combined CITP/CZE and CIEF analyses, the top 17 biosynthetic and metabolic pathways were mapped out by the Ingenuity Pathway Analysis^{TM} and are shown in Figure 3-4. The ratio is calculated as the number of proteins in a given pathway which meet cutoff criteria divided by total number of proteins which make up that pathway. The significance value associated with the functional analysis for a dataset is a measure for how likely the proteins from the dataset under investigation participate in that function. The significance is expressed as a p-value, which is calculated using the right-tailed Fisher's Exact Test. The p-value is calculated by comparing the number

of user-specified proteins of interest that participate in a given function or pathway, relative to the total number of occurrences of these proteins in all functional/pathway annotations stored in the Ingenuity Pathway Knowledge Base.

The most significant metabolic pathway revealed by our mitochondrial dataset is clearly the oxidative phosphorylation pathway. As energy producers to sustain the living cell activities, mitochondria constantly convert the glycolysis products, pyruvate and NADH, into energy in the form of ATP through the cellular respiration process. In order to generate ATP in the efficient way, mitochondria require numerous transport proteins and enzymes to perform the oxidation reactions of the respiratory chain, the ATP synthesis, and the regulations of metabolites and substance transportations.

A total of 100 proteins engaged in the oxidative phosphorylation pathway, corresponding to approximately 62% pathway coverage, were identified in this study. Five complexes involved in the oxidative phosphorylation pathway were shown in Figure 3-5, including NADH dehydrogenase, fumarate reductase, cytochrome bc1 complex, cytochrome c oxidase, and ATP synthase families. The coverage of different complexes were analyzed as 87% (40/46) for complex I, 80% (4/5) for complex II, 73% (8/11) for complex III, 85% (11/13) for complex IV, and 94% (15/16) for complex V. The overall complex coverage was around 86% (78/91) and higher than 62% pathway coverage. As shown in Figure 3-6, tandem MS spectra of single unique peptide hits leading to the identifications of NDUA1, NU3M, and NU4LM as representative subunits of complex I exhibited excellent E-values of 5.8 x 10^{-10}, 6.6 x 10^{-12}, and 8.3 x 10^{-16}, respectively, from a target-decoy OMSSA search.
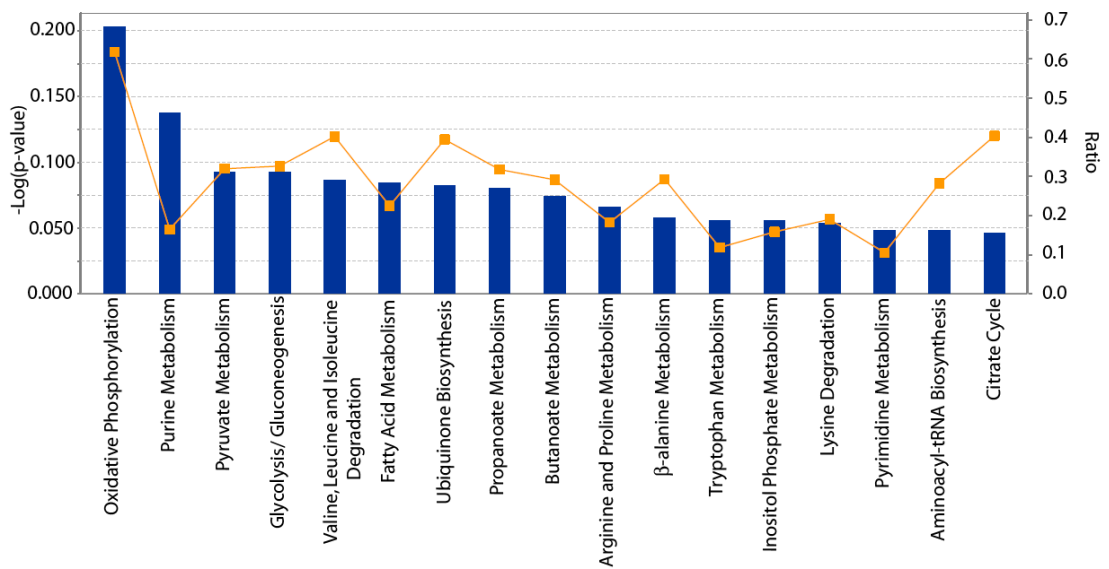
Figure3- 4 Top 17 biosynthetic and metabolic pathways constructed from combined mitochondrial proteome dataset using the Ingenuity Pathway Analysis[TM]. The significance and the ratio of the proteome dataset associated with individual pathways are represented by the bar and the line, respectively.
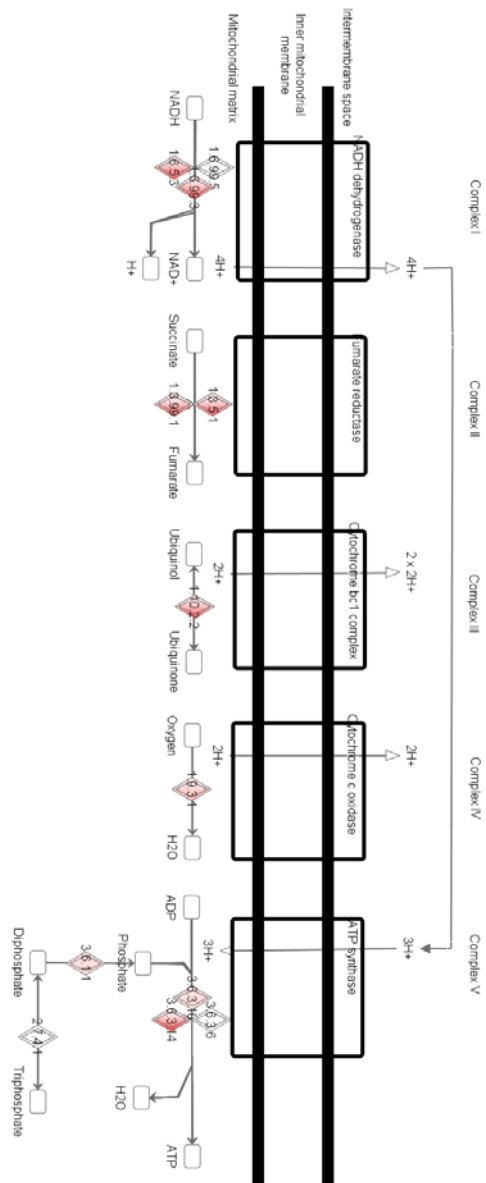
Figure 3-5 Simplified illustration of the oxidative phosphorylation pathway containing five complexes.
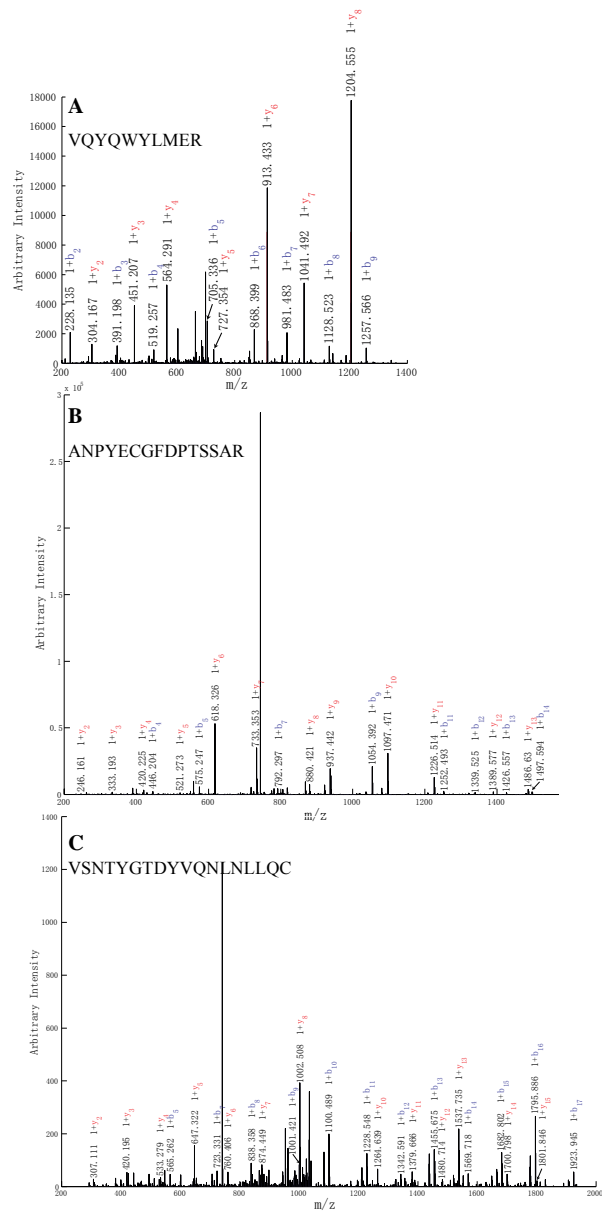
Figure 3-6 Tandem MS spectra of single unique peptide hits leading to the identifications

of (A) NDUA1, (B) NU3M, and (C) NU4LM.

3.4 Conclusion

Mitochondria are a major source and target of free radicals [83-85, 94, 122, 123], and the collapse of the mitochondrial transmembrane potential can initiate the signaling cascades involved in programmed cell death or apoptosis [82, 124-126]. Mitochondria play a crucial role in the regeneration of antioxidants through the production of reducing equivalents [127], and are responsible for the vast majority of ATP production within most cells and higher organisms. Mitochondria also infer a major role in cell signaling, as well as biosynthesis (e.g. heme) and degradation (e.g. urea cycle) [128]. Mitochondrial mutations and dysfunction have been implicated in numerous diseases including aging, cancer, heart disease, neurological diseases such as Parkinson, Alzheimer, and Huntington, and various neuromuscular syndromes [122, 129-131]. The mitochondrial connection to these major degenerative diseases has consequently driven significant research efforts to define the mitochondrial proteome and to discover new molecular targets for drug development and therapeutic intervention.

The work presented in this study highlights our continuous efforts toward the development and characterization of electrokinetic-based proteome technologies for mining deeper into complex proteomes such as the mouse mitochondria proteome. Transient CITP/CZE not only contributed to selective enrichment of low abundance peptides/proteins [7, 114], but also resolved peptides/proteins based on their differences in electrophoretic mobility. The ultrahigh resolving power of transient CITP/CZE was evidenced by the large number of distinct peptide identifications measured from each CITP/CZE fraction and the low peptide fraction overlapping among identified peptides. In addition to selective analyte enrichment, total peak capacity improvements in combined CITP/CZE-nano-RPLC

separations greatly increased the number of detected peptides and proteins identified due to better use of the MS dynamic range and reduced discrimination during ionization.

A total of 12,110 fully tryptic peptides were detected at a 1% FDR, leading to the identification of 1,795 distinct mouse SwissProt protein entries (or 1,705 non-redundant proteins) from a single CITP/CZE-based proteomic study of synaptic mouse brain mitochondria. The reproducibility of protein identifications was evaluated and was found to be around 86% by comparing proteins identified from repeated runs of the same mitochondrial sample. The analysis of the mouse mitochondrial proteome by two CITP/CZE runs led to the detection of 2,095 distinct protein entries (or 1,992 non-redundant proteins), corresponding to 59% coverage of the updated Maestro mitochondrial reference set. By combining the results obtained from the CITP/CZE and CIEF-based proteomic studies, the collective analysis yielded the identification of 2,191 distinct protein entries (or 2,082 non-redundant proteins), corresponding to 76% coverage of the MitoP2-database reference set. The increased coverage of mouse mitochondrial proteome further revealed the top 17 biosynthetic and metabolic pathways in synaptic mitochondria isolated from mouse brain using the Ingenuity Pathway Analysis[TM].

3.5 Acknowledgement

Chapter 4: Comparison of Multidimensional Shotgun Technologies Targeting Tissue Proteomics

Reproduced with permission from Fang, X., Park, D.M., Lee, C. S., Balgley, B. M., Submitted to Nat. Methods.

4.1 Introduction

Although blood is the preferred assay specimen for a final diagnostic test, the biological materials analyzed for biomarker discovery need not be plasma. In addition to the analytical challenges posed by the complexity and depth of the plasma proteome, many biomarkers elaborated from diseased tissues as the result of leakage and secretion are massively diluted into circulating blood with ultimate concentrations which are five to seven orders of magnitude lower in abundance than the most highly concentrated plasma proteins. If biomarkers specific for a particular disease arise locally from the affected tissue, it is highly plausible that a gradient of concentration diminishes with increasing distance from the site of disease. Thus, it has been proposed that a better way to discover biomarkers might be to study tissues involved in disease or the fluids bathing theses tissues[132, 133].

However, in the absence of protein amplification, proteomic analysis of tumor 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[23]. While surface-enhanced laser desorption/ionization-mass spectrometry (SELDI-MS) has been shown to produce observable MS profiles with

as few as 25-50 cells[4, 134], the transition from spectral patterns produced by SELDI-MS to protein identity is generally quite difficult[135]. 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 the overall SELDI-MS approach when working with limited clinical samples.

While two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) analyses of tissue samples have been attempted[24, 136-143], including the analysis of minute tumor tissues[25, 144-147], these studies required significant manual effort and time to extract sufficient levels of protein for analysis, while providing little information on protein expression beyond a relatively small number of high abundance proteins. On the other hand, the reported tissue proteomic studies employing multidimensional liquid chromatography separations[148-151] have been mainly based on the analysis of entire tissue sections instead of targeted cell populations. The minimal quantity of available protein samples from limited cell populations has restricted recent analyses[152-160] to the use of only a single chromatography separation prior to tandem MS analysis and greatly limited the ability for mining deeper into the tissue proteome.

Combined capillary isoelectric focusing (CIEF)/nano-reversed phase liquid chromatography (nano-RPLC) separations coupled with electrospray ionization-mass spectrometry (ESI-MS) have been demonstrated to enable ultrasensitive analysis of minute proteins extracted from fresh-frozen[14, 16, 17] and formalin-fixed and paraffin-embedded (FFPE) tissues[21, 22]. As evaluated in our recent analyses of human saliva and mouse brain mitochondrial samples[7, 114, 161], a capillary isotachophoresis (CITP)-

based proteomic platform not only enables selective enrichment of low abundance proteins, but also provides superior resolving power than that achieved using CIEF/nano-RPLC separations. In this work, comparisons among shotgun proteome technologies, including multidimensional CITP and liquid chromatography separations[54, 55, 162], are performed regarding their abilities to address the challenges of protein complexity and relative abundance inherent in cancer stem cells as a model tissue proteome system.

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), iodoacetamide (IAM), and trifluoroacetic acid (TFA) were purchased from Sigma (St. Louis, MO). Acetonitrile, ammonium acetate, ammonium formate, hydroxypropyl cellulose (average MW 100,000), sodium dodecyl sulfate (SDS), tris(hydroxymethyl)aminomethane (Tris), and urea were obtained from Fisher Scientific (Pittsburgh, PA). Pharmalyte 3-10 was purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). Sequencing grade trypsin was acquired from Promega (Madison, WI). All solutions were prepared using water purified by a Neu-Ion system (Baltimore, MD) equipped with a UV sterilizing lamp and a 0.05 μm membrane final filter.

4.2.2 Preparation of Glioblastoma Derived Cancer Stem Cells

Neural stem cells were isolated from glioblastoma multiforme (GBM) tissues and cultured[163] in Dr. Deric M. Park's laboratory at the University of Pittsburgh Cancer Institute. Pelleted cells were suspended in a buffer that consisted of 50 mM Tris (pH 8.0), 65 mM DTT, and 10% glycerol. The cells were disrupted by homogenization with glass beads in a Mini-BeadBeater (BioSpec Products, Bartlesville, OK). The soluble proteins were collected in the supernatant by centrifugation at 20,000 g for 30 min.

Proteins collected in the supernatant were denatured, reduced, and alkylated by sequentially adding urea, DTT, and IAM with final concentrations of 8 M, 1 mg/mL, and 2 mg/mL, respectively. The solution was incubated at 37 ^{0}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 ^{0}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 ^{0}C.

4.2.3 Transient CITP/Capillary Zone Electrophoresis (CZE)-Based Multidimensional Separations

Transient CITP/Capillary Zone Electrophoresis (CZE)-Based Multidimensional Separations. Similar to the procedures described in our previous studies[114, 161], a 80-cm long CITP capillary (100 μm i.d./365 μm o.d.) coated with hydroxypropyl cellulose was initially filled with a background electrophoresis buffer of 0.1 M acetic

acid at pH 2.8. The samples containing protein digests processed from neural stem cells were 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 filled with a 0.1 M acetic acid solution.

The cathodic end of the capillary was housed inside a stainless steel needle using a coaxial liquid sheath flow configuration[164, 165]. A sheath liquid composed of 0.1 M acetic acid was delivered by 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.

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 Strong Cation-Exchange (SCX) Coupled with Nano-RPLC-ESI-MS

A SCX column (5 cm length x 300 μm i.d.) packed with 5-μm poly(2-sulfoethyl aspartamide) particles were purchased from PolyLC (Columbia, MD). The salt gradient for SCX was the same as that employed by Liu and co-workers for the characterization of the human plasma proteome[166]. Briefly, peptides were eluted at a flow rate of 10 μL/min using a 10-255 mM ammonium formate gradient (containing 25% acetonitrile) over 50 min with the remaining 20 min for column regeneration and equilibration. Peptides collected in individual wells of a microtiter plate were lyophilized to dryness and reconstituted in a 0.1% TFA solution containing 2% acetonitrile. Peptide fractions resolved from SCX were subsequently and individually analyzed using nano-RPLC-ESI-MS by following the same procedures described previously.

4.2.5 MS Data Analysis

Raw LTQ data were converted to peak list files by msn_extract.exe (ThermoFinnigan). Open Mass Spectrometry Search Algorithm (OMSSA)[19] 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[70] and employed in our previous study for a comparative evaluation among commonly used tandem MS identification search algorithms[119]. 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 are reported here.

4.3 Result and Discussion

4.3.1 Evaluation of overall proteome performance

Increasing evidence support the hypothesis that a variety of cancers, including GBM, the most common and malignant brain tumor, arise from malignant transformation of undifferentiated stem-progenitor cells. Similar to normal neural stem cell (multipotent cells in the central nervous system capable of differentiating into neurons, astrocytes, and oligodendrocytes), GBM derived stem cells express primitive markers, self-renew, and are able to differentiate to multiple lineages upon activation of appropriate pathways[167, 168]. The presence of malignant stem-like cells within the heterogeneous cell populations in GBM is thought to be responsible for recurrence, infiltration into surrounding white matter, and radiation-chemoresistance. Therefore, treatment strategies directed against the malignant stem cell compartment of GBM has the potential to limit proliferation and invasion to allow a greater tumor resection, and more durable response to radiation and chemotherapy.

Neural stem cells were isolated from GBM tissues and cultured[163] in Dr. Park's laboratory as part of ongoing research collaboration toward the identification of stem and differentiation markers in determining cellular behavior, including growth, motility, proliferative index, and survival. Similar to primary cell cultures, glioblastoma derived cancer stem cells in culture have a very limited lifespan before they senesce. Thus, these small cancer stem cell populations isolated from tumor tissues and enriched in culture represent an excellent opportunity for conducting comparisons among shotgun proteome techniques targeting tissue proteomics,

including transient CITP/CZE-based multidimensional separations and multidimensional protein identification technology (MudPIT)[54, 55, 162].

Comparisons were achieved using a single processed tryptic digest of cancer stem cells, equal loading of 15 μg digest into the CITP capillary and the SCX column, same nano-RPLC-ESI-MS conditions, and identical search parameters together with a consistent cutoff as established by the target-decoy determined FDR[70, 119]. By enhancing the overall separation peak capacity and the MS duty cycle, the numbers of total spectral counts (or total peptide identifications), distinct peptides, and distinct proteins summarized in Figure 4-1 all increased with increasing the number of SCX fractions from 4, 8, and 30. A 30 CITP/CZE-fraction measurement, however, resulted in further enhancements in total peptide, distinct peptide, and distinct protein identifications over a 30-fraction SCX run by 119%, 192%, and 79%, respectively.

There were very good agreements between proteomic results obtained from 30 fractions SCX- and CITP-based shotgun techniques on a core set of proteins identified by at least one (Figure 4-2A) or two (Figure 4-2B) distinct peptides. The Venn diagrams illustrated that the CITP-based proteomic platform uniquely detected 2,614 proteins by one or more distinct peptides or 1,524 proteins by two or more distinct peptides. Proteins identified by two or more distinct peptides using MudPIT were completely included in the CITP measurements. In addition to increasing the number of protein identifications, the percentage of proteins identified by two or more distinct peptides also increased from 60% (19% + 41%) in 4- and 8-fraction runs to 65% (16% + 49%) in a 30-fraction MudPIT experiment (Figure 4-3). This

percentage was further increased to 77% using transient CITP/CZE-based multidimensional separations.

Other measures of proteome data quality are the average numbers of spectral counts and distinct peptide identifications per protein. Increase in the number of distinct peptide identifications per protein is interpreted as an improvement in the confidence and the sequence coverage of a protein identification[169]. The number of spectral counts becomes very important when implementing spectral counting-based label-free quantification[76, 170, 171], as the expression levels of each protein are determined by the number of tandem MS events. Although CITP delivered 79% more protein identifications (Figure 4-1), CITP still outperformed MudPIT with respect to both measures: an average of 18.8 spectral counts and 6.9 distinct peptides per protein in CITP versus an average 15.4 spectral counts and 4.3 distinct peptides of each protein in MudPIT.

4.3.2 Comparison of separation performance

The high resolving power of transient CITP/CZE as the first separation dimension is clearly demonstrated by significantly low peptide fraction overlapping shown in Figure 4-4. Approximately 89% of distinct peptides were identified in only a single CITP fraction. Of the remaining peptides, 8% were identified in two fractions and 3% were measured in three and more fractions. Only a 4-fraction MudPIT run exhibited lower fraction overlapping than that observed in CITP, however, at the expense of the overall resolving power.
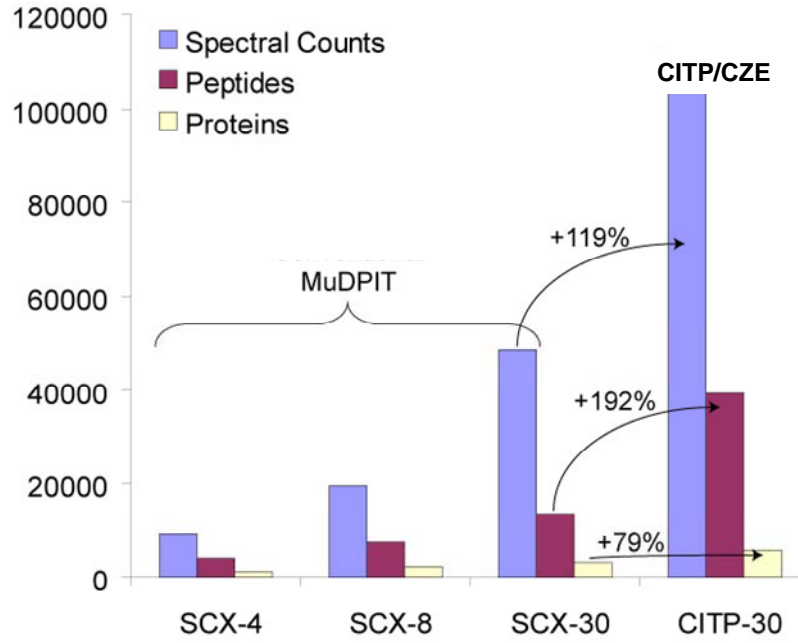
Figure 4-1 Summary of proteomic results of GBM derived neural stem cells using

MuDPIT and CITP/CZE-based multidimensional separations.
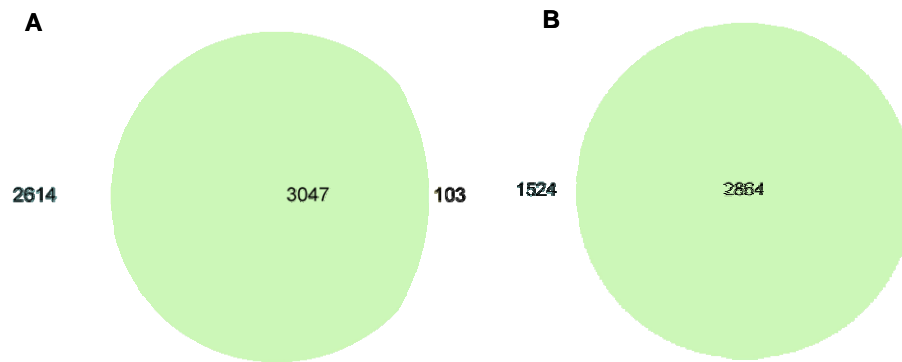
Figure 4-2 The Venn diagrams comparing proteins detected using MudPIT (small

cycle) and CITP/CZE-based multidimensional separations (large cycle).

Proteins are identified by at least one (A) or two (B) distinct peptides.
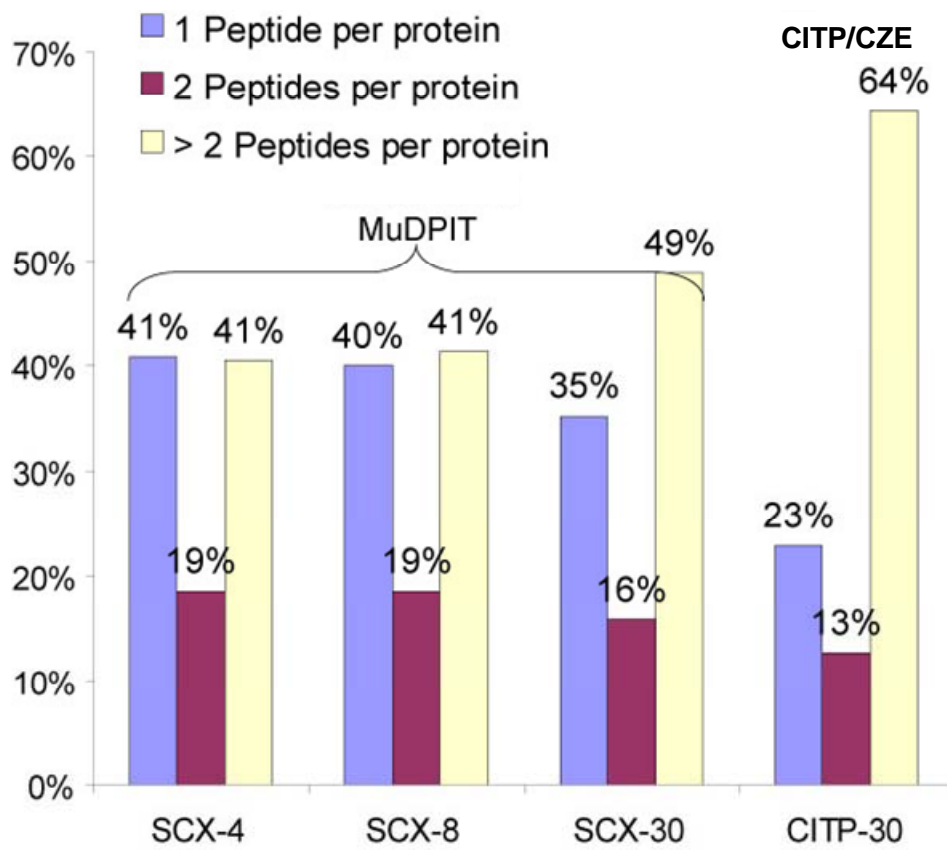
Figure 4-3 Summary of the percentages of proteins identified by one, two, or more

distinct peptides using MudPIT and CITP/CZE-based shotgun proteome

techniques.
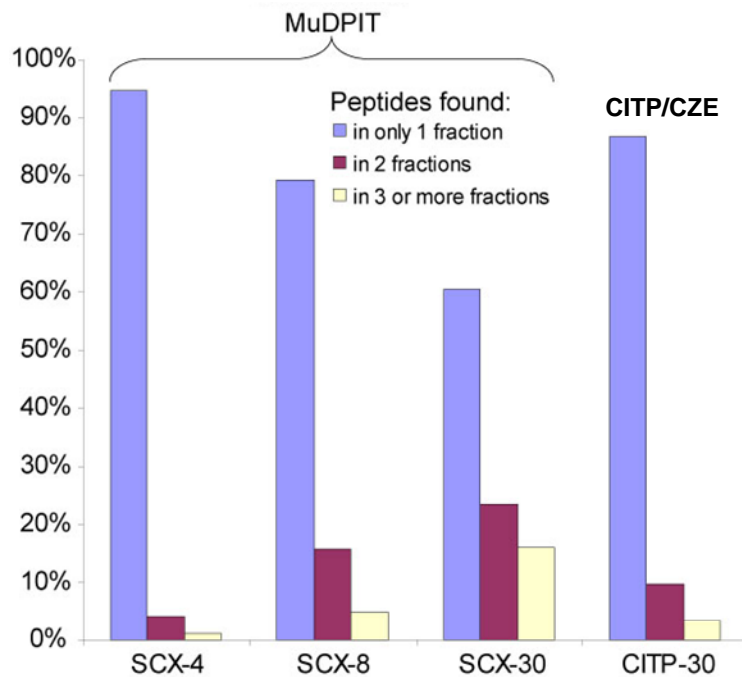
Figure 4-4 Summary of the percentages of peptides identified from single or multiple

CITP and SCX fractions using MudPIT and CITP/CZE-based

multidimensional separations, respectively.

The percentage of distinct peptides identified in only a single SCX fraction decreased from 94% in 4 fractions to 80% in 8 fractions and 61% in 30 fractions. In contrast, this percentage determined from the CITP measurements was only reduced from 95% in 15 fractions to 89% in 30 fractions. Thus, the intrinsically high resolution of transient CITP/CZE allows the number of fractions sampled in the first separation dimension to be increased for enhancing the overall peak capacity of multidimensional separations and the resulting proteome coverage, particularly for the characterization of small cell populations and limited tissue samples.

While increasing the number of CITP fractions, the amount of individual (not total) peptides collected in each of CITP fractions and subsequently loaded into each of nano-RPLC columns still remained the same as long as there was no over-sampling. By maintaining the same amount of individual peptides introduced and resolved in nano-RPLC-MS, the enhancement in the overall peak capacity further reduced matrix effects by simplifying the complexity of co-eluting peptides prior to MS detection and sequencing. Matrix effects negatively affect the dynamic range and detection sensitivity of MS measurements in three critical ways. First, high abundance peptides compete for ionization and leave less current available for low abundance peptides. Second, the presence of highly abundant ions limits the opportunity of low abundance ions for entering the ion trap and being selected for tandem MS sequencing. Last, the MS detector itself has a finite dynamic range and can be saturated by the presence of higher abundance ions.

In addition to analyzing fraction overlapping, the ratio of new distinct peptides identified in each SCX or CITP fraction to distinct peptides collectively determined in

all prior fractions was also evaluated and summarized in Figure 4-5. Extensive fraction overlapping in SCX clearly limits its ability to identify new peptides. Comparisons among 4-, 8-, and 30-fraction SCX runs indicated that the new peptide ratio reached a maximum around the salt concentration of 110-130 mM ammonium formate, corresponding to the elution of mainly triply charged peptides. The association of this maximum new peptide ratio with 3^+ charged peptides was also in good agreement with the highest number of peptide identifications achieved from triply charged peptides in the work of Peng and co-workers [162].

In contrast, the CITP-based multidimensional separations delivered high new peptide ratios evenly across all 30 fractions. Transient CITP/CZE separations resolve peptides due to their differences in electrophoretic mobility which is influenced by the charge and size of the peptide ions, and provides better selectivity over charge-based SCX. It is known that capillary electrophoresis, in general, offers superior separation efficiency than that typically achieved in LC, further contributing to enhancement in the overall resolving power.

4.3.3 Assessment of proteome reproducibility

The CITP-based proteomic platform was coupled with the spectral counting quantification approach[76, 170, 171] for determining protein expression profile within GBM derived cancer stem cells. The spectral counts of individual proteins were normalized against the average number of spectral counts across all runs and termed expression values. Several measures, including the coefficient of variation (CV) and the Pearson correlation coefficient, were employed to assess proteomic data quality and reproducibility.
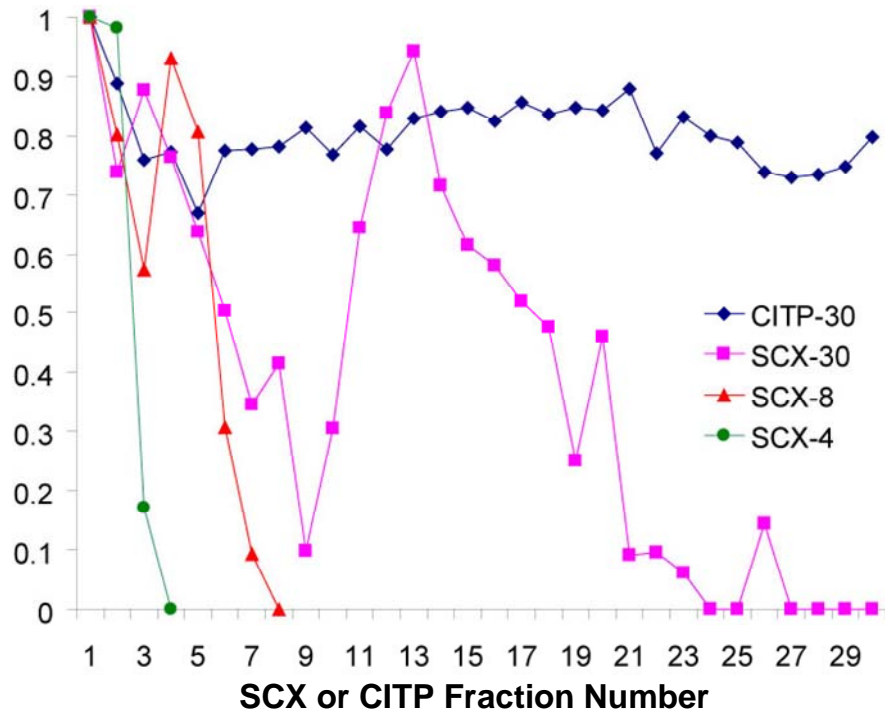
Figure 4-5 The ratio of new distinct peptides found in each SCX or CITP fraction versus distinct peptides found previously in all prior fractions.

Correlation of two runs of the same tryptic digest of neural stem cells is shown in Figure 4-6. All separations were performed on different CITP capillaries and nano-RPLC columns. All proteins with one or more spectral counts were used to calculate the correlation yielding a Pearson R^2 value of 0.98 over a dynamic range of greater than 10^3. This value compares favorably with a recent report of a Pearson R^2 correlation of 0.88 among technical replicates using MuDPIT[172]. Furthermore, the CV was less than 10% for over half of proteins quantified and averaged 15% across all proteins.

4.3.4 Protein functional subclass and pathway coverage

The Venn diagrams shown in Figure 4-7 specifically compare the proteome coverage among various functional categories, including tyrosine kinases, kinases, transmembrane, proto-oncogenes, apoptosis, and transcription regulators, achieved using the CITP-based proteomic platform and MudPIT. The CITP measurements routinely detected about two fold more proteins in any given subclass, even for very low abundance proteins such as tyrosine kinases and proto-oncogenes. Similar to comparative proteome results shown in Figure 4-2, very few proteins in these subclasses were uniquely identified by MudPIT.

The approach of grouping individual expression changes into pathways of actions is not only useful in trying to interpret genomic and proteomic datasets, but also allows common biological themes to emerge from the comparative proteomic analyses[173].
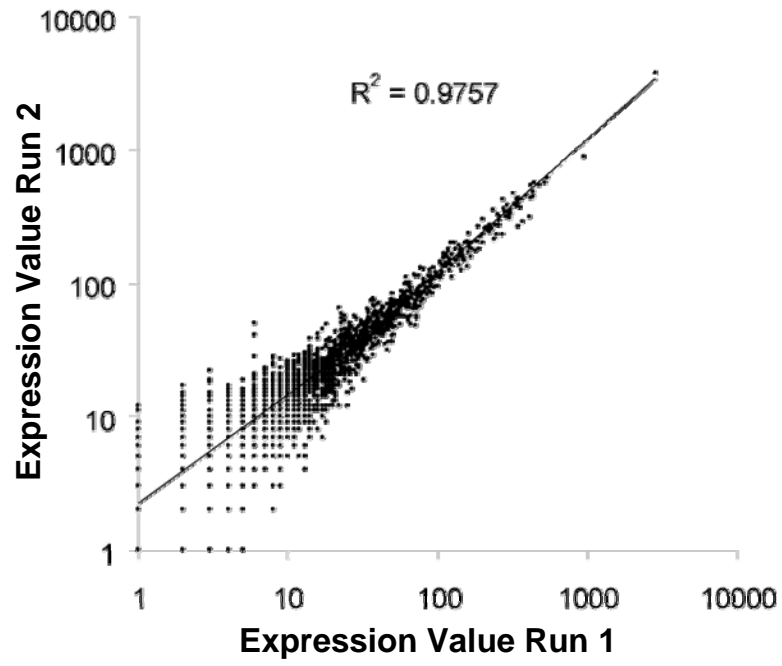
Figure 4-6 Pearson correlation plot of all proteins quantified in two runs of the same

tryptic digest obtained from GBM derived cancer stem cells.
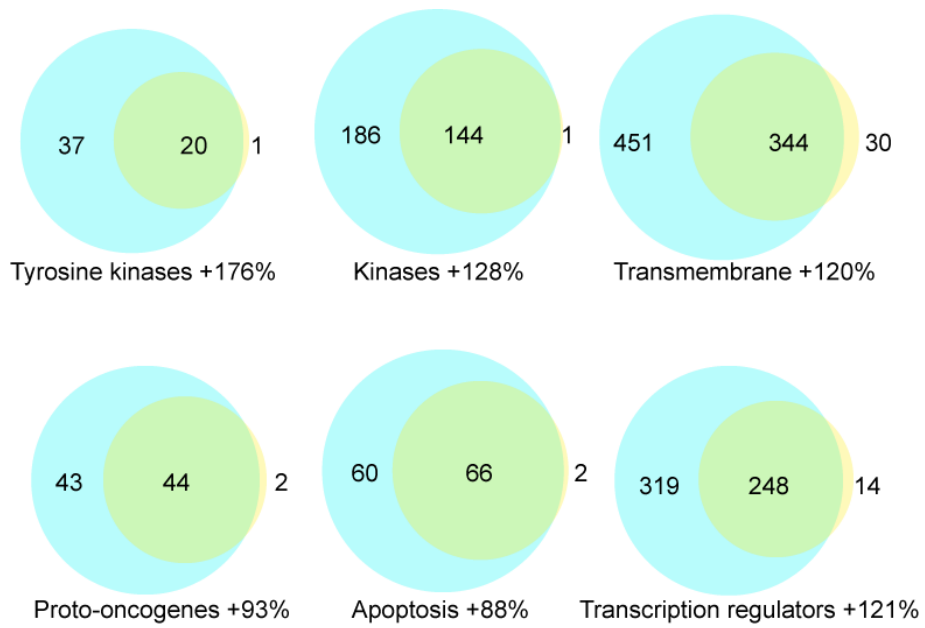
Figure4-7 The Venn diagrams comparing proteins identified within different functional categories using MudPIT (small cycle) and CITP/CZE-based multidimensional separations (large cycle).

Comparative proteomics involving measurements in changes of biological pathways or functions are expected to provide relevant therapy-associated biomarkers and networks, and the initial steps and control points in signaling cascades and compensatory processes that drive the progression of therapeutic response and/or drug resistance. From a practical perspective, the evaluation of comparative proteomic dataset within a biological context is essential for high-throughput data validation, prioritization of follow-on biomarker selection and validation experiments.

The coverage of key protein networks related to neural cancer stem cells such as the extracellular signal regulated kinases (ERK)/mitogen-activated protein kinases (MAPK) pathway (Figure 4-8) was analyzed using the Ingenuity System^{TM} and compared among the results achieved by the CITP proteomic platform and MudPIT. Clearly, combined CITP/nano-RPLC separations, due to its excellent resolving power, accomplished significantly greater coverage in this key pathway than that obtained from the multidimensional liquid chromatography system. Many biologically relevant proteins, including MKP, Rafs, and Src in the ERK/MAPK pathway, were only identified by the CITP-based shotgun proteomic measurements.
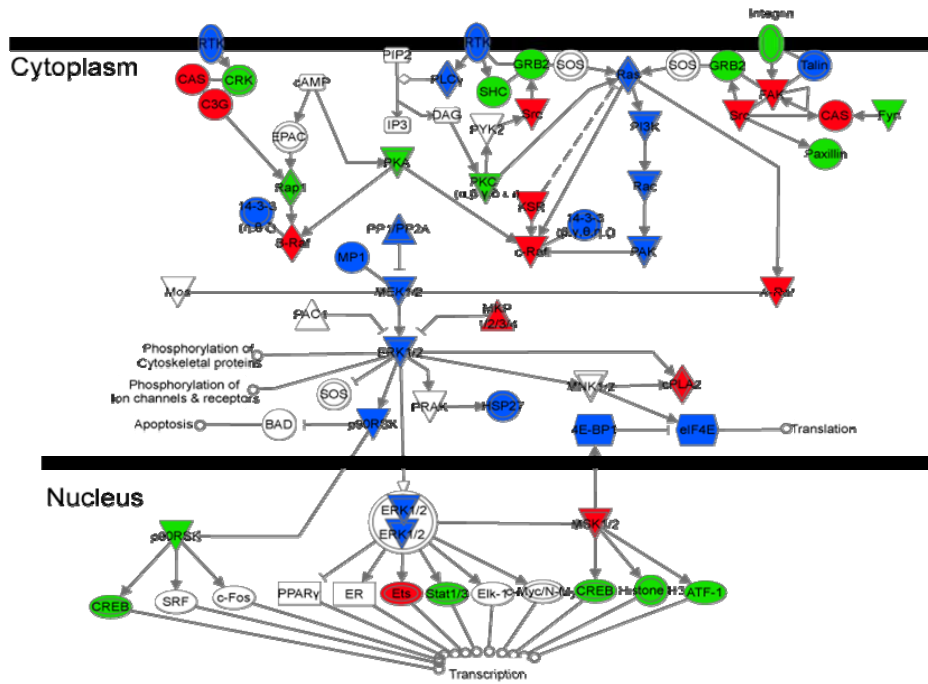
Figure 4-8 Comparison of the coverage in the ERK/MAPK pathway achieved by MudPIT and CITP/CZE-based shotgun proteome techniques. Red: proteins only identified by CITP; blue: proteins identified by CITP with higher confidence (larger numbers of spectral counts and distinct peptide identifications per protein) than that achieved by MudPIT; green: proteins identified by both CITP and MudPIT with approximately equal confidence.

4.4 Conclusion

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 proteomic technology is critically needed to enable comprehensive and sensitive studies of protein profiles that will have diagnostic and therapeutic relevance. By employing small cancer stem cell populations isolated from GBM tissues and enriched in culture, comparisons among shotgun proteome techniques, including transient CITP/CZE-based multidimensional separations and MudPIT, were performed and reported in this work. Comparisons were conducted using a single processed protein digest with equal sample loading, identical second dimension separation and MS conditions, and consistent search parameters and cutoff established by the target-decoy determined FDR.

For the evaluation of overall proteome performance, a CITP/CZE measurement demonstrated significant enhancements in total peptide, distinct peptide, and distinct protein identifications over a corresponding SCX run by 119%, 192%, and 79%, respectively. On the basis of at least two or more distinct peptides leading to each protein identification, the Venn diagram indicated the complete inclusion of proteins identified by MudPIT within the CITP-based proteomic results. CITP further outperformed MudPIT with respect to the average numbers of distinct peptide identifications and spectral counts per protein which lead to improvements in the sequence coverage and the confidence of protein quantification.

As evaluated by the measurements of CV and the Pearson correlation coefficient, the CITP-based proteome platform coupled with the spectral counting approach was

demonstrated to be highly sensitive and reproducible in determining protein relative abundance in complex tissue proteome. In contrast, extensive fraction overlapping in SCX greatly limited the ability of MudPIT for mining deeper into the tissue proteome. As evidenced by the coverage among various protein functional categories, CITP routinely detected about two fold more proteins in any given subclass, even for very low abundance proteins such as tyrosine kinases and proto-oncogenes. Combined CITP/nano-RPLC separations, due to its excellent resolving power, further accomplished superior coverage in key pathways than that of MudPIT. Many biologically relevant proteins, including MKP, Rafs, and Src in the ERK/MAPK pathway, were only identified by the CITP-based shotgun proteomic measurements.

4.5 Acknowledgment

Chapter 5: Conclusion

The ability to perform comparative proteome analysis, including the changes in protein expression and post-translational modifications, would have profound impact on our understanding of how disease occurs and develops, and provide a molecular basis for future drug intervention to treat human diseases. Besides protein complexity, the greatest bioanalytical challenge facing comprehensive proteome analysis of clinical specimens, particularly in the identification of low abundance proteins, is related to the large variation of protein relative abundances (>6 orders of magnitude). One obvious approach to increase the detection capability of low abundant proteins is to significantly raise the sample loading along with many additional protein/peptide fractionation and separation procedures, and therefore is impractical for proteomic studies of small cell populations and limited tissue samples.

Thus, this project aims to demonstrate the application of a CITP-based multidimensional separation platform, capable of providing selective analyte enrichment and extremely high resolving power toward complex protein/peptide mixtures, to achieve comprehensive proteome analysis including the characterization of low abundance proteins. In contrast to universally enriching all analytes by a similar degree, the result of the CITP stacking process is that major components may be diluted, but trace compounds are concentrated. Such selective enhancement toward low abundance proteins drastically reduces the range of relative protein abundances within complex proteomes, and greatly enhances proteomic coverage using the CITP-based proteomic technology. Furthermore, CITP offers the benefits of speed and straightforward manipulation/switching between the stacking and separation modes in

transient CITP/CZE. Transient CITP/CZE further provides seamless combination with nano-RPLC as two highly resolving and completely orthogonal separation techniques in an integrated platform.

5.1 Summary of Accomplishments

Chapter 2 presents our continuous efforts toward the development and application of electrokinetics-based proteome technologies for mining deeper into the human saliva proteome. Compared with CIEF utilized as the first separation dimension in our previous studies, transient CITP/CZE not only contributes to selective analyte enrichment toward low abundance proteins, but also provides high resolving power for addressing the complexity of the saliva proteome. The ultrahigh resolving power of transient CITP/CZE is evidenced by the large number of distinct peptide identifications measured from each CITP fraction together with the low peptide fraction overlapping among identified peptides. A total of 6,112 fully tryptic peptides are detected at a 1% FDR, leading to the identification of 1,479 distinct human SwissProt protein entries from a single human saliva sample.

Mitochondria are a major source and target of free radicals, and the collapse of the mitochondrial transmembrane potential can initiate the signaling cascades involved in programmed cell death or apoptosis. Mitochondria play a crucial role in the regeneration of antioxidants through the production of reducing equivalents, and are responsible for the vast majority of ATP production within most cells and higher organisms. Mitochondrial mutations and dysfunction have been implicated in numerous diseases including aging, cancer, heart disease, neurological diseases such as Parkinson, Alzheimer, and Huntington, and various neuromuscular syndromes. The mitochondrial connection to these major

degenerative diseases has consequently driven significant research efforts to define the mitochondrial proteome and to discover new molecular targets for drug development and therapeutic intervention.

As presented in Chapter 3, selective analyte enrichment together with total peak capacity improvements in combined CITP/CZE-nano-RPLC separations greatly increase the number of detected peptides and proteins identified from synaptic mitochondria isolated from mouse brain due to better use of the MS dynamic range and reduced discrimination during ionization. A total of 12,110 fully tryptic peptides are detected at a 1% FDR, leading to the identification of 1,795 distinct mouse SwissProt protein entries from a single CITP/CZE-based proteomic study of synaptic mouse brain mitochondria. The reproducibility of protein identifications is found to be around 86% by comparing proteins identified from repeated runs of the same mitochondrial sample. The analysis of the mouse mitochondrial proteome by two CITP/CZE runs leads to the detection of 2,095 distinct protein entries, corresponding to 59% coverage of the updated Maestro mitochondrial reference set. By combining the results obtained from the CITP/CZE and CIEF-based proteomic studies, the collective analysis yields the identification of 2,191 distinct protein entries, corresponding to 76% coverage of the MitoP2-database reference set. The increased coverage of mouse mitochondrial proteome further reveals the top 17 biosynthetic and metabolic pathways in synaptic mitochondria isolated from mouse brain using the Ingenuity Pathway Analysis^{TM}.

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 proteomic technology is critically needed to enable

comprehensive and sensitive studies of protein profiles that will have diagnostic and therapeutic relevance. By employing small cancer stem cell populations isolated from GBM tissues and enriched in culture, comparisons among shotgun proteome techniques, including transient CITP/CZE-based multidimensional separations and MuDPIT, are performed and reported in Chapter 4. Comparisons are conducted using a single processed protein digest with equal sample loading, identical second dimension separation and MS conditions, and consistent search parameters and cutoff established by the target-decoy determined FDR.

For the evaluation of overall proteome performance, a CITP/CZE measurement demonstrates significant enhancements in total peptide, distinct peptide, and distinct protein identifications over a corresponding SCX run by 119%, 192%, and 79%, respectively. On the basis of at least two or more distinct peptides leading to each protein identification, the Venn diagram indicates the complete inclusion of proteins identified by MudPIT within the CITP-based proteomic results. CITP further outperforms MudPIT with respect to the average numbers of distinct peptide identifications and spectral counts per protein which lead to improvements in the sequence coverage and the confidence of protein quantification.

As evaluated by the measurements of CV and the Pearson correlation coefficient, the CITP-based proteome platform coupled with the spectral counting approach is demonstrated to be highly sensitive and reproducible in determining protein relative abundance in complex tissue proteome. In contrast, extensive fraction overlapping in SCX greatly limited the ability of MudPIT for mining deeper into the tissue proteome. As evidenced by the coverage among various protein functional categories, CITP routinely

detected about two fold more proteins in any given subclass, even for very low abundance proteins such as tyrosine kinases and proto-oncogenes.

5.2 Contributions to Current Proteomic Technologies Development

By comparing with CIEF, CITP offers a broader field of application. Peptides with extreme pI values may be outside the working pH range of CIEF due to limited availability of commercial ampholytes for the creation of a pH gradient inside the capillary. Furthermore, analytes focused in CIEF reside at their pIs where they have an increased tendency to precipitate. In contrast, the stacked peptides in CITP are less prone to precipitation due to their charged nature. Most importantly, the ultrahigh resolving power of CITP/CZE is evidenced by the large number of distinct peptide identifications measured from each CITP fraction together with the low peptide fraction overlapping among identified peptides.

The intrinsic high resolution of transient CITP/CZE allows the number of fractions sampled in the first separation dimension to be further increased by simply increasing the number of deposits collected in a microtiter plate. The excellent reproducibility together with straightforward CITP fractionation enables user discretion as to the number of fractions to be gathered and analyzed, and which interesting fractions may be revisited (reanalyzed) or repeatedly accumulated for identification of extremely low abundance peptides. In comparison with a multidimensional liquid chromatography system such as MudPIT, an increase in the number of the SCX fractions only contributes to minor performance improvements due to extensive peptide fraction overlapping in SCX.

5.3 Future Work

The proteome measurement system's dynamic range is defined as the ratio of the quantity of a protein of high abundance to the quantity of a protein of low abundance. The maximum dynamic range for a single mass spectrum without the use of spectrum averaging or summation is typically constrained to about 1,000. The use of multidimensional chromatography platforms in shotgun proteomics has resulted in better utilization of the MS dynamic range and reduced discrimination among analytes during ionization, thus achieving a system's dynamic range of 10,000:1. One part of my proposed future studies will involve the determination of the overall dynamic range of the CITP-based proteomic technology using complex protein digests spiked with isotopically labeled peptides of known concentrations. These peptides will be synthesized and labeled with stable isotopes using Sigma's custom services. Additionally, the results of quantitative immunoassays, provided by Rules-Based Medicine (www.rulesbasedmedicine.com), will also be employed to evaluate the dynamic range of proteins identified by combined CITP-CZE/nano-RPLC separations equipped with ESI-LTQ-MS/MS.

Besides the use of Western Blot and immunohistochemistry techniques for validating individual protein biomarkers, another part of my proposed future studies will be centered on the application of an absolute quantification (AQUA)-LC-single reaction monitoring (SRM) approach. Peptides which are diagnostic for the candidate marker proteins based on discovery results obtained from the CITP-based proteomic studies will be synthesized and labeled with stable isotopes using Sigma's custom services. The labeled peptides are spiked into additional disease samples at known quantities

and serve as internal standards. AQUA peptides will be synthesized to contain at least a 6 Da shift utilizing ^{13}C to avoid isotopic interference with the native peptide and to ensure identical elution time with the native peptide in nano-RPLC.

Optimal collision energy levels will be determined independently for each AQUA peptide by nano-electrospray infusion of the peptide. The ratio of peak areas from a pair of labeled and non-labeled peptides of the same sequence permits a determination of the absolute quantity of the unlabeled peptide based on the known quantity of the AQUA peptide. Both labeled and label-free peptides will fragment in exactly the same manner, providing further verification of the peptide sequence. AQUA-LC-SRM permits the analysis of hundreds of candidate peptide ions in a single nano-RPLC run, allowing for broad-based, sensitive validation of the initial biomarker results.

References

(1)   Kaiser, T.; Hermann, A.; Kielstein, J. T.; Wittke, S.; Bartel, S.; Krebs, R.; Hausadel, F.; Hillmann, M.; Golovko, I.; Koester, P.; Haller, H.; Weissinger, E. M.; Fliser, D.; Mischak, H. *J Chromatogr A* **2003**, *1013*, 157-171.

(2)   Wittke, S.; Fliser, D.; Haubitz, M.; Bartel, S.; Krebs, R.; Hausadel, F.; Hillmann, M.; Golovko, I.; Koester, P.; Haller, H.; Kaiser, T.; Mischak, H.; Weissinger, E. M. *J Chromatogr A* **2003**, *1013*, 173-181.

(3)   Neuhoff, N. v.; Kaiser, T.; Wittke, S.; Krebs, R.; Pitt, A.; Burchard, A.; Sundmacher, A.; Schlegelberger, B.; Kolch, W.; Mischak, H. *Rapid Commun Mass Spectrom* **2004**, *18*, 149-156.

(4)   Cottingham, K. *Anal Chem* **2003**, *75*, 476.

(5)   Theodorescu, D.; Wittke, S.; Ross, M. M.; Walden, M.; Conaway, M.; Just, I.; Mischak, H.; Frierson, H. F. *Lancet Oncol* **2006**, *7*, 230-240.

(6)   Sassi, A. P.; Andel, F.; Bitter, H.-M. L.; Brown, M. P. S.; Chapman, R. G.; Espiritu, J.; Greenquist, A. C.; Guyon, I.; Horchi-Alegre, M.; Stults, K. L.; Wainright, A.; Heller, J. C.; Stults, J. T. *Electrophoresis* **2005**, *26*, 1500-1512.

(7)   An, Y.; Cooper, J. W.; Balgley, B. M.; Lee, C. S. *Electrophoresis* **2006**, *27*, 3599-3608.

(8)   Janini, G. M.; Chan, K. C.; Conrads, T. P.; Issaq, H. J.; Veenstra, T. D. *Electrophoresis* **2004**, *25*, 1973-1980.

(9)   Mao, Y.; Li, Y.; Zhang, X. *Proteomics* **2006**, *6*, 420-426.

(10)   Yu, W.; Li, Y.; Deng, C.; Zhang, X. *Electrophoresis* **2006**, *27*, 2100-2110.

(11)   Michels, D. A.; Hu, S.; Schoenherr, R. M.; Eggertson, M. J.; Dovichi, N. J. *Mol Cell Proteomics* **2002**, *1*, 69-74.

(12)   Hu, S.; Michels, D. A.; Fazal, M. A.; Ratisoontorn, C.; Cunningham, M. L.; Dovichi, N. J. *Anal Chem* **2004**, *76*, 4044-4049.

(13)   Zhang, M.; El Rassi, Z. *J Proteome Res* **2006**, *5*, 2001-2008.

(14)   Wang, Y.; Rudnick, P. A.; Evans, E. L.; Li, J.; Zhuang, Z.; Devoe, D. L.; Lee, C. S.; Balgley, B. M. *Anal Chem* **2005**, *77*, 6549-6556.

(15)   Wang, Y.; Balgley, B. M.; Lee, C. S. *Expert Rev Proteomics* **2005**, *2*, 659-667.

(16)   Balgley, B. M.; Wang, W.; Devoe, D. L.; Lee, C. S. *Personalized Medicine* **2007**, *4*, 45-58.

(17)   Wang, W.; Guo, T.; Rudnick, P. A.; Song, T.; Li, J.; Zhuang, Z.; Zheng, W.; Devoe, D. L.; Lee, C. S.; Balgley, B. M. *Anal Chem* **2007**, *79*, 1002-1009.

(18)   Chen, J.; Balgley, B. M.; DeVoe, D. L.; Lee, C. S. *Anal Chem* **2003**, *75*, 3145-3152.

(19)   Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *J Proteome Res* **2004**, *3*, 958-964.

(20)   Shi, S. R.; Key, M. E.; Kalra, K. L. *J Histochem Cytochem* **1991**, *39*, 741-748.

(21)   Shi, S.-R.; Liu, C.; Balgley, B. M.; Lee, C.; Taylor, C. R. *J Histochem Cytochem* **2006**, *54*, 739-743.

(22)   Guo, T.; Wang, W.; Rudnick, P. A.; Song, T.; Li, J.; Zhuang, Z.; Weil, R. J.; DeVoe, D. L.; Lee, C. S.; Balgley, B. M. *J Histochem Cytochem* **2007**, *55*, 763-772.

(23)   Bonner, R. F.; Emmert-Buck, M.; Cole, K.; Pohida, T.; Chuaqui, R.; Goldstein, S.; Liotta, L. A. *Science* **1997**, *278*, 1481-1481.

(24)   De Souza, A. I.; McGregor, E.; Dunn, M. J.; Rose, M. L. *Proteomics* **2004**, *4*, 578-586.

(25) Zhuang, Z.; Lee, Y. S.; Zeng, W.; Furuta, M.; Valyi-Nagy, T.; Johnson, M. D.; Vnencak-Jones, C. L.; Woltjer, R. L.; Weil, R. J. *Neurology* **2004**, *62*, 2316-2319.

(26) Amerongen, A. V. N.; Veerman, E. C. I. *Oral Dis* **2002**, *8*, 12-22.

(27) Fischer, H. P.; Eich, W.; Russell, I. J. *Semin Arthritis Rheum* **1998**, *27*, 348-359.

(28) Kaufman, E.; Lamster, I. B. *J Clin Periodontol* **2000**, *27*, 453-465.

(29) Kaufman, E.; Lamster, I. B. *Crit Rev Oral Biol Med* **2002**, *13*, 197-212.

(30) Nagler, R. M.; Hershkovich, O.; Lischinsky, S.; Diamond, E.; Reznick, A. Z. *J Investig Med* **2002**, *50*, 214-225.

(31) Streckfus, C. F.; Bigler, L. R. *Oral Dis* **2002**, *8*, 69-76.

(32) Wong, D. T. *J Am Dent Assoc* **2006**, *137*, 313-321.

(33) Brinkman, B.; Wong, D. *Curr Opin Oncol* **2006**, *18*, 228-233.

(34) Amado, F. M. L.; Vitorino, R. M. P.; Domingues, P. M. D. N.; Lobo, M. J. C.; Duarte, J. A. R. *Expert Rev Proteomics* **2005**, *2*, 521-539.

(35) Banks, R. E.; Dunn, M. J.; Hochstrasser, D. F.; Sanchez, J. C.; Blackstock, W.; Pappin, D. J.; Selby, P. J. *Lancet* **2000**, *356*, 1749-1756.

(36) Drake, R. R.; Cazare, L. H.; Semmes, O. J.; Wadsworth, J. T. *Expert Rev Mol Diagn* **2005**, *5*, 93-9100.

(37) Ayad, M.; Van Wuyckhuyse, B. C.; Minaguchi, K.; Raubertas, R. F.; Bedi, G. S.; Billings, R. J.; Bowen, W. H.; Tabak, L. A. *J Dent Res* **2000**, *79*, 976-982.

(38) Chen, Y.-C.; Li, T.-Y.; Tsai, M.-F. *Rapid Commun Mass Spectrom* **2002**, *16*, 364-369.

(39) Dale, B. A.; Krisanaprakornkit, S. *J Oral Pathol Med* **2001**, *30*, 321-327.

(40) Ganz, T. *Nat Rev Immunol* **2003**, *3*, 710-720.

(41) Mathews, M.; Jia, H. P.; Guthmiller, J. M.; Losh, G.; Graham, S.; Johnson, G. K.; Tack, B. F.; McCray, P. B. *Infect Immun* **1999**, *67*, 2740-2745.

(42) Mizukawa, N.; Sugiyama, K.; Fukunaga, J.; Ueno, T.; Mishima, K.; Takagi, S.; Sugahara, T. *Anticancer Res* **1998**, *18*, 4645-4649.

(43) Mizukawa, N.; Sugiyama, K.; Ueno, T.; Mishima, K.; Takagi, S.; Sugahara, T. *Oral Dis* **1999**, *5*, 139-142.

(44) Mahoney, E. J.; Spiegel, J. H. *Otolaryngol Clin North Am* **2003**, *36*, 733-745.

(45) Peeters, F. P.; deVries, M. W.; Vissink, A. *Gen Hosp Psychiatry* **1998**, *20*, 150-154.

(46) Vissink, A.; Burlage, F. R.; Spijkervet, F. K. L.; Jansma, J.; Coppes, R. P. *Crit Rev Oral Biol Med* **2003**, *14*, 213-225.

(47) Tenovuo, J. *Oral Dis* **2002**, *8*, 23-29.

(48) Ghafouri, B.; Tagesson, C.; Lindahl, M. *Proteomics* **2003**, *3*, 1003-1015.

(49) Hardt, M.; Thomas, L. R.; Dixon, S. E.; Newport, G.; Agabian, N.; Prakobphol, A.; Hall, S. C.; Witkowska, H. E.; Fisher, S. J. *Biochemistry* **2005**, *44*, 2885-2899.

(50) Hu, S.; Xie, Y.; Ramachandran, P.; Ogorzalek Loo, R. R.; Li, Y.; Loo, J. A.; Wong, D. T. *Proteomics* **2005**, *5*, 1714-1728.

(51) Huang, C.-M. *Arch Oral Biol* **2004**, *49*, 951-962.

(52) Vitorino, R.; Lobo, M. J. C.; Ferrer-Correira, A. J.; Dubin, J. R.; Tomer, K. B.; Domingues, P. M.; Amado, F. M. L. *Proteomics* **2004**, *4*, 1109-1115.

(53) Yao, Y.; Berg, E. A.; Costello, C. E.; Troxler, R. F.; Oppenheim, F. G. *J Biol Chem* **2003**, *278*, 5300-5308.

(54) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat Biotechnol* **2001**, *19*, 242-247.

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

(56) Wilmarth, P. A.; Riviere, M. A.; Rustvold, D. L.; Lauten, J. D.; Madden, T. E.; David, L. L. *J Proteome Res* **2004**, *3*, 1017-1023.

(57) Xie, H.; Rhodus, N. L.; Griffin, R. J.; Carlis, J. V.; Griffin, T. J. *Mol Cell Proteomics* **2005**, *4*, 1826-1830.

(58) Hu, S.; Loo, J. A.; Wong, D. T. *Proteomics* **2006**, *6*, 6326-6353.

(59)   Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. *Electrophoresis* **2000**, *21*, 1104-1115.
(60)   Lollo, B. A.; Harvey, S.; Liao, J.; Stevens, A. C.; Wagenknecht, R.; Sayen, R.; Whaley, J.; Sajjadi, F. G. *Electrophoresis* **1999**, *20*, 854-859.
(61)   Sato, A. K.; Sexton, D. J.; Morganelli, L. A.; Cohen, E. H.; Wu, Q. L.; Conley, G. P.; Streltsova, Z.; Lee, S. W.; Devlin, M.; DeOliveira, D. B.; Enright, J.; Kent, R. B.; Wescott, C. R.; Ransohoff, T. C.; Ley, A. C.; Ladner, R. C. *Biotechnol Prog* **2002**, *18*, 182-192.
(62)   Schirmer, E. C.; Florens, L.; Guan, T.; Yates, J. R.; Gerace, L. *Science* **2003**, *301*, 1380-1382.
(63)   Ramachandran, P.; Boontheung, P.; Xie, Y.; Sondej, M.; Wong, D. T.; Loo, J. A. *J Proteome Res* **2006**, *5*, 1493-1503.
(64)   Motoyama, A.; Venable, J. D.; Ruse, C. I.; Yates, J. R. *Anal Chem* **2006**, *78*, 5109-5118.
(65)   Guo, T.; Rudnick, P. A.; Wang, W.; Lee, C. S.; Devoe, D. L.; Balgley, B. M. *J Proteome Res* **2006**, *5*, 1469-1478.
(66)   Foret, F.; Szoko, E.; Karger, B. L. *Electrophoresis* **1993**, *14*, 417-428.
(67)   Gebauer, P.; Bocek, P. *Electrophoresis* **2000**, *21*, 3898-3904.
(68)   Stegehuis, D. S.; Irth, H.; Tjaden, U. R.; Van der Greef, J. *J Chromatogr* **1991**, *538*, 393-402.
(69)   Yang, L.; Lee, C. S.; Hofstadler, S. A.; Pasa-Tolic, L.; Smith, R. D. *Anal Chem* **1998**, *70*, 3235-3241.
(70)   Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. *Nat Methods* **2005**, *2*, 667-675.
(71)   Shen, Y.; Berger, S. J.; Anderson, G. A.; Smith, R. D. *Anal Chem* **2000**, *72*, 2154-2159.
(72)   Whitelegge, J. P.; Zhang, H.; Aguilera, R.; Taylor, R. M.; Cramer, W. A. *Mol Cell Proteomics* **2002**, *1*, 816-827.
(73)   Whitelegge, J. P.; Gundersen, C. B.; Faull, K. F. *Protein Sci* **1998**, *7*, 1423-1430.
(74)   Mohan, D.; Lee, C. S. *Electrophoresis* **2002**, *23*, 3160-3167.
(75)   Mohan, D.; Pasa-Tolic, L.; Masselon, C. D.; Tolic, N.; Bogdanov, B.; Hixson, K. K.; Smith, R. D.; Lee, C. S. *Anal Chem* **2003**, *75*, 4432-4440.
(76)   Liu, H.; Sadygov, R. G.; Yates, J. R. *Anal Chem* **2004**, *76*, 4193-4201.
(77)   Elias, J. E.; Gygi, S. P. *Nat Methods* **2007**, *4*, 207-214.
(78)   Olsen, J. V.; Ong, S.-E.; Mann, M. *Mol Cell Proteomics* **2004**, *3*, 608-614.
(79)   Yates, J. R.; Cociorva, D.; Liao, L.; Zabrouskov, V. *Anal Chem* **2006**, *78*, 493-500.
(80)   Craig, R.; Cortens, J. C.; Fenyo, D.; Beavis, R. C. *J Proteome Res* **2006**, *5*, 1843-1849.
(81)   Frewen, B. E.; Merrihew, G. E.; Wu, C. C.; Noble, W. S.; MacCoss, M. J. *Anal Chem* **2006**, *78*, 5678-5684.
(82)   Green, D. R.; Reed, J. C. *Science* **1998**, *281*, 1309-1312.
(83)   Leonard, J. V.; Schapira, A. H. *Lancet* **2000**, *355*, 299-304.
(84)   Lenaz, G. *Biochim Biophys Acta* **1998**, *1366*, 53-67.
(85)   Schapira, A. H. *Biochim Biophys Acta* **1999**, *1410*, 99-9102.
(86)   Boveris, A.; Cadenas, E.; Stoppani, A. O. *Biochem J* **1976**, *156*, 435-444.
(87)   Raha, S.; McEachern, G. E.; Myint, A. T.; Robinson, B. H. *Free Radic Biol Med* **2000**, *29*, 170-180.
(88)   Harman, D. *J Am Geriatr Soc* **1972**, *20*, 145-147.
(89)   Chang, J.; Van Remmen, H.; Cornell, J.; Richardson, A.; Ward, W. F. *Mech Ageing Dev* **2003**, *124*, 33-41.
(90)   Lithgow, T. *FEBS Lett* **2000**, *476*, 22-26.
(91)   Koehler, C. M. *FEBS Lett* **2000**, *476*, 27-31.

(92)   McDonald, T. G.; Van Eyk, J. E. *Basic Res Cardiol* **2003**, *98*, 219-227.
(93)   Lopez, M. F.; Kristal, B. S.; Chernokalskaya, E.; Lazarev, A.; Shestopalov, A. I.; Bogdanova, A.; Robinson, M. *Electrophoresis* **2000**, *21*, 3427-3440.
(94)   Lopez, M. F.; Melov, S. *Circ Res* **2002**, *90*, 380-389.
(95)   Hanson, B. J.; Schulenberg, B.; Patton, W. F.; Capaldi, R. A. *Electrophoresis* **2001**, *22*, 950-959.
(96)   Sickmann, A.; Reinders, J.; Wagner, Y.; Joppich, C.; Zahedi, R.; Meyer, H. E.; Schonfisch, B.; Perschil, I.; Chacinska, A.; Guiard, B.; Rehling, P.; Pfanner, N.; Meisinger, C. *Proc Natl Acad Sci U S A* **2003**, *100*, 13207-13212.
(97)   McDonald, T.; Sheng, S.; Stanley, B.; Chen, D.; Ko, Y.; Cole, R. N.; Pedersen, P.; Van Eyk, J. E. *Mol Cell Proteomics* **2006**, *5*, 2392-2411.
(98)   Reinders, J.; Zahedi, R. P.; Pfanner, N.; Meisinger, C.; Sickmann, A. *J Proteome Res* **2006**, *5*, 1543-1554.
(99)   Pflieger, D.; Le Caer, J.-P.; Lemaire, C.; Bernard, B. A.; Dujardin, G.; Rossier, J. *Anal Chem* **2002**, *74*, 2400-2406.
(100)   Mootha, V. K.; Bunkenborg, J.; Olsen, J. V.; Hjerrild, M.; Wisniewski, J. R.; Stahl, E.; Bolouri, M. S.; Ray, H. N.; Sihag, S.; Kamal, M.; Patterson, N.; Lander, E. S.; Mann, M. *Cell* **2003**, *115*, 629-640.
(101)   Taylor, S. W.; Fahy, E.; Zhang, B.; Glenn, G. M.; Warnock, D. E.; Wiley, S.; Murphy, A. N.; Gaucher, S. P.; Capaldi, R. A.; Gibson, B. W.; Ghosh, S. S. *Nat Biotechnol* **2003**, *21*, 281-286.
(102)   Taylor, S. W.; Warnock, D. E.; Glenn, G. M.; Zhang, B.; Fahy, E.; Gaucher, S. P.; Capaldi, R. A.; Gibson, B. W.; Ghosh, S. S. *J Proteome Res* **2002**, *1*, 451-458.
(103)   Gaucher, S. P.; Taylor, S. W.; Fahy, E.; Zhang, B.; Warnock, D. E.; Ghosh, S. S.; Gibson, B. W. *J Proteome Res* **2004**, *3*, 495-505.
(104)   Prokisch, H.; Scharfe, C.; Camp, D. G.; Xiao, W.; David, L.; Andreoli, C.; Monroe, M. E.; Moore, R. J.; Gritsenko, M. A.; Kozany, C.; Hixson, K. K.; Mottaz, H. M.; Zischka, H.; Ueffing, M.; Herman, Z. S.; Davis, R. W.; Meitinger, T.; Oefner, P. J.; Smith, R. D.; Steinmetz, L. M. *PLoS Biol* **2004**, *2*.
(105)   Arnold, R. J.; Hrncirova, P.; Annaiah, K.; Novotny, M. V. *J Proteome Res* **2004**, *3*, 653-657.
(106)   Rezaul, K.; Wu, L.; Mayya, V.; Hwang, S.-I.; Han, D. *Mol Cell Proteomics* **2005**, *4*, 169-181.
(107)   Jiang, X.-S.; Dai, J.; Sheng, Q.-H.; Zhang, L.; Xia, Q.-C.; Wu, J.-R.; Zeng, R. *Mol Cell Proteomics* **2005**, *4*, 12-34.
(108)   Kislinger, T.; Cox, B.; Kannan, A.; Chung, C.; Hu, P.; Ignatchenko, A.; Scott, M. S.; Gramolini, A. O.; Morris, Q.; Hallett, M. T.; Rossant, J.; Hughes, T. R.; Frey, B.; Emili, A. *Cell* **2006**, *125*, 173-186.
(109)   Forner, F.; Foster, L. J.; Campanaro, S.; Valle, G.; Mann, M. *Mol Cell Proteomics* **2006**, *5*, 608-619.
(110)   Johnson, D. T.; Harris, R. A.; French, S.; Blair, P. V.; You, J.; Bemis, K. G.; Wang, M.; Balaban, R. S. *Am J Physiol Cell Physiol* **2007**, *292*, 689-697.
(111)   DiMauro, S.; Schon, E. A. *Nat Genet* **1998**, *19*, 214-215.
(112)   Taylor, S. W.; Fahy, E.; Ghosh, S. S. *Trends Biotechnol* **2003**, *21*, 82-88.
(113)   Calvo, S.; Jain, M.; Xie, X.; Sheth, S. A.; Chang, B.; Goldberger, O. A.; Spinazzola, A.; Zeviani, M.; Carr, S. A.; Mootha, V. K. *Nat Genet* **2006**, *38*, 576-582.
(114)   Fang, X.; Yang, L.; Wang, W.; Song, T.; Lee, C.; Devoe, D.; Balgley, B. *Anal Chem* **2007**, *79*, 5785-5792.
(115)   Chandrasekaran, K.; Hazelton, J. L.; Wang, Y.; Fiskum, G.; Kristian, T. *J Neurosci* **2006**, *26*, 13123-13127.

(116) Dunkley, P. R.; Heath, J. W.; Harrison, S. M.; Jarvie, P. E.; Glenfield, P. J.; Rostas, J. A. *Brain Res* **1988**, *441*, 59-71.

(117) Brown, M. R.; Sullivan, P. G.; Dorenbos, K. A.; Modafferi, E. A.; Geddes, J. W.; Steward, O. *J Neurosci Methods* **2004**, *137*, 299-303.

(118) Kristian, T.; Hopkins, I. B.; McKenna, M. C.; Fiskum, G. *J Neurosci Methods* **2006**, *152*, 136-143.

(119) Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. *Mol Cell Proteomics* **2007**, *6*, 1599-1608.

(120) Prokisch, H.; Andreoli, C.; Ahting, U.; Heiss, K.; Ruepp, A.; Scharfe, C.; Meitinger, T. *Nucleic Acids Res* **2006**, *34*, 705-711.

(121) Andreoli, C.; Prokisch, H.; Hortnagel, K.; Mueller, J. C.; Munsterkotter, M.; Scharfe, C.; Meitinger, T. *Nucleic Acids Res* **2004**, *32*, 459-462.

(122) Hirano, M.; Davidson, M.; DiMauro, S. *Curr Opin Cardiol* **2001**, *16*, 201-210.

(123) Melov, S. *Ann N Y Acad Sci* **2000**, *908*, 219-225.

(124) Susin, S. A.; Lorenzo, H. K.; Zamzami, N.; Marzo, I.; Snow, B. E.; Brothers, G. M.; Mangion, J.; Jacotot, E.; Costantini, P.; Loeffler, M.; Larochette, N.; Goodlett, D. R.; Aebersold, R.; Siderovski, D. P.; Penninger, J. M.; Kroemer, G. *Nature* **1999**, *397*, 441-446.

(125) Nicholls, D. G.; Budd, S. L. *Physiol Rev* **2000**, *80*, 315-360.

(126) Kuwana, T.; Smith, J. J.; Muzio, M.; Dixit, V.; Newmeyer, D. D.; Kornbluth, S. *J Biol Chem* **1998**, *273*, 16589-16594.

(127) Kagan, V. E.; Tyurina, Y. Y. *Ann N Y Acad Sci* **1998**, *854*, 425-434.

(128) Gunter, T. E.; Buntinas, L.; Sparagna, G. C.; Gunter, K. K. *Biochim Biophys Acta* **1998**, *1366*, 5-15.

(129) Beckman, K. B.; Ames, B. N. *Physiol Rev* **1998**, *78*, 547-581.

(130) Kovacic, P.; Jacintho, J. D. *Curr Med Chem* **2001**, *8*, 863-892.

(131) Verma, M.; Kagan, J.; Sidransky, D.; Srivastava, S. *Nat Rev Cancer* **2003**, *3*, 789-795.

(132) Rifai, N.; Gillette, M. A.; Carr, S. A. *Nat Biotechnol* **2006**, *24*, 971-983.

(133) Cottingham, K. *J Proteome Res* **2007**, *6*, 2052-2052.

(134) Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. *Lancet* **2002**, *359*, 572-577.

(135) Gillette, M. A.; Mani, D. R.; Carr, S. A. *J Proteome Res* **2005**, *4*, 1143-1154.

(136) Ruepp, S. U.; Tonge, R. P.; Shaw, J.; Wallis, N.; Pognan, F. *Toxicol Sci* **2002**, *65*, 135-150.

(137) Banks, R. E.; Dunn, M. J.; Forbes, M. A.; Stanley, A.; Pappin, D.; Naven, T.; Gough, M.; Harnden, P.; Selby, P. J. *Electrophoresis* **1999**, *20*, 689-700.

(138) Ornstein, D. K.; Gillespie, J. W.; Paweletz, C. P.; Duray, P. H.; Herring, J.; Vocke, C. D.; Topalian, S. L.; Bostwick, D. G.; Linehan, W. M.; Petricoin, E. F.; Emmert-Buck, M. R. *Electrophoresis* **2000**, *21*, 2235-2242.

(139) Lawrie, L. C.; Curran, S.; McLeod, H. L.; Fothergill, J. E.; Murray, G. I. *Mol Pathol* **2001**, *54*, 253-258.

(140) Craven, R. A.; Totty, N.; Harnden, P.; Selby, P. J.; Banks, R. E. *Am J Pathol* **2002**, *160*, 815-822.

(141) Somiari, R. I.; Sullivan, A.; Russell, S.; Somiari, S.; Hu, H.; Jordan, R.; George, A.; Katenhusen, R.; Buchowiecka, A.; Arciero, C.; Brzeski, H.; Hooke, J.; Shriver, C. *Proteomics* **2003**, *3*, 1863-1873.

(142) Heijne, W. H. M.; Stierum, R. H.; Slijper, M.; van Bladeren, P. J.; van Ommen, B. *Biochem Pharmacol* **2003**, *65*, 857-875.

(143) Somiari, R. I.; Somiari, S.; Russell, S.; Shriver, C. D. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, *815*, 215-225.

(144) Furuta, M.; Weil, R. J.; Vortmeyer, A. O.; Huang, S.; Lei, J.; Huang, T.-N.; Lee, Y.-S.; Bhowmick, D. A.; Lubensky, I. A.; Oldfield, E. H.; Zhuang, Z. *Oncogene* **2004**, *23*, 6806-6814.

(145) Vogel, T. W.; Zhuang, Z.; Li, J.; Okamoto, H.; Furuta, M.; Lee, Y.-S.; Zeng, W.; Oldfield, E. H.; Vortmeyer, A. O.; Weil, R. J. *Clin Cancer Res* **2005**, *11*, 3624-3632.

(146) Li, J.; Zhuang, Z.; Okamoto, H.; Vortmeyer, A. O.; Park, D. M.; Furuta, M.; Lee, Y. S.; Oldfield, E. H.; Zeng, W.; Weil, R. J. *Neurology* **2006**, *66*, 733-736.

(147) Zhuang, Z.; Huang, S.; Kowalak, J. A.; Shi, Y.; Lei, J.; Furuta, M.; Lee, Y.-S.; Lubensky, I. A.; Rodgers, G. P.; Cornelius, A. S.; Weil, R. J.; Teh, B. T.; Vortmeyer, A. O. *Int J Oncol* **2006**, *28*, 103-110.

(148) Li, C.; Hong, Y.; Tan, Y.-X.; Zhou, H.; Ai, J.-H.; Li, S.-J.; Zhang, L.; Xia, Q.-C.; Wu, J.-R.; Wang, H.-Y.; Zeng, R. *Mol Cell Proteomics* **2004**, *3*, 399-409.

(149) Zhang, J.; Hu, H.; Gao, M.; Yang, P.; Zhang, X. *Electrophoresis* **2004**, *25*, 2374-2383.

(150) DeSouza, L.; Diehl, G.; Rodrigues, M. J.; Guo, J.; Romaschin, A. D.; Colgan, T. J.; Siu, K. W. M. *J Proteome Res* **2005**, *4*, 377-386.

(151) Cagney, G.; Park, S.; Chung, C.; Tong, B.; O'Dushlaine, C.; Shields, D. C.; Emili, A. *J Proteome Res* **2005**, *4*, 1757-1767.

(152) Wu, S.-L.; Hancock, W. S.; Goodrich, G. G.; Kunitake, S. T. *Proteomics* **2003**, *3*, 1037-1046.

(153) Zang, L.; Palmer Toy, D.; Hancock, W. S.; Sgroi, D. C.; Karger, B. L. *J Proteome Res* **2004**, *3*, 604-612.

(154) Baker, H.; Patel, V.; Molinolo, A. A.; Shillitoe, E. J.; Ensley, J. F.; Yoo, G. H.; Meneses-Garcia, A.; Myers, J. N.; El-Naggar, A. K.; Gutkind, J. S.; Hancock, W. S. *Oral Oncol* **2005**, *41*, 183-199.

(155) Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Furusato, B.; Lucas, D. A.; Ringeisen, B. R.; Sesterhenn, I. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B. *Mol Cell Proteomics* **2005**, *4*, 1741-1753.

(156) Crockett, D. K.; Lin, Z.; Vaughn, C. P.; Lim, M. S.; Elenitoba-Johnson, K. S. J. *Lab Invest* **2005**, *85*, 1405-1415.

(157) Palmer-Toy, D. E.; Krastins, B.; Sarracino, D. A.; Nadol, J. B.; Merchant, S. N. *J Proteome Res* **2005**, *4*, 2404-2411.

(158) Rahimi, F.; Shepherd, C. E.; Halliday, G. M.; Geczy, C. L.; Raftery, M. J. *Anal Chem* **2006**, *78*, 7216-7221.

(159) Hwang, S. I.; Thumar, J.; Lundgren, D. H.; Rezaul, K.; Mayya, V.; Wu, L.; Eng, J.; Wright, M. E.; Han, D. K. *Oncogene* **2007**, *26*, 65-76.

(160) Bagnato, C.; Thumar, J.; Mayya, V.; Hwang, S.-I.; Zebroski, H.; Claffey, K. P.; Haudenschild, C.; Eng, J. K.; Lundgren, D. H.; Han, D. K. *Mol Cell Proteomics* **2007**, *6*, 1088-1102.

(161) Fang, X.; Wang, W.; Yang, L.; Chandrasekaran, K.; Kristian, T.; Balgley, B. M.; Lee, C. S. *Electrophoresis* **2008**.

(162) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *J Proteome Res* **2003**, *2*, 43-50.

(163) Park, D. M.; Li, J.; Okamoto, H.; Akeju, O.; Kim, S. H.; Lubensky, I.; Vortmeyer, A.; Dambrosia, J.; Weil, R. J.; Oldfield, E. H.; Park, J. K.; Zhuang, Z. *Cell Cycle* **2007**, *6*, 467-470.

(164) Tang, W.; Harrata, A. K.; Lee, C. S. *Anal Chem* **1997**, *69*, 3177-3182.

(165) Yang, L.; Lee, C. S.; Hofstadler, S. A.; Smith, R. D. *Anal Chem* **1998**, *70*, 4945-4950.

(166) Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Xiao, W.; Moldawer, L. L.; Kaushal, A.; Monroe, M. E.; Varnum, S. M.; Moore, R. J.; Purvine, S. O.; Maier, R. V.; Davis, R.

W.; Tompkins, R. G.; Camp, D. G.; Smith, R. D. *Mol Cell Proteomics* **2006**, *5*, 1899-1913.

(167) Bao, S.; Wu, Q.; McLendon, R. E.; Hao, Y.; Shi, Q.; Hjelmeland, A. B.; Dewhirst, M. W.; Bigner, D. D.; Rich, J. N. *Nature* **2006**, *444*, 756-760.

(168) Piccirillo, S. G. M.; Reynolds, B. A.; Zanetti, N.; Lamorte, G.; Binda, E.; Broggi, G.; Brem, H.; Olivi, A.; Dimeco, F.; Vescovi, A. L. *Nature* **2006**, *444*, 761-765.

(169) MacCoss, M. J.; Wu, C. C.; Yates, J. R. *Anal Chem* **2002**, *74*, 5593-5599.

(170) Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M. *Genome Res* **2002**, *12*, 1231-1245.

(171) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. *Mol Cell Proteomics* **2005**, *4*, 1265-1272.

(172) Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M. *Nat Biotechnol* **2007**, *25*, 117-124.

(173) Lin, B.; White, J. T.; Lu, W.; Xie, T.; Utleg, A. G.; Yan, X.; Yi, E. C.; Shannon, P.; Khrebtukova, I.; Lange, P. H.; Goodlett, D. R.; Zhou, D.; Vasicek, T. J.; Hood, L. *Cancer Res* **2005**, *65*, 3081-3091.