DEVELOPMENT OF MULTIPLEXED TECHNIQUES USING 2D-HPLC, PROTEIN MICROARRAYS AND MASS SPECTROMETRY FOR INVESTIGATIONS IN PROTEIN POSTTRANSLATIONAL MODIFICATIONS AND DISEASE PROGRESSION PATHWAYS

by

Manoj Pal

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2007

Doctoral Committee:

Professor David M. Lubman, Chair
Professor Ari Gafni
Associate Professor Ayyalusamy Ramamoorthy
Assistant Professor Kristina I. Hakansson

Good science is the ability to look at things in a new way and achieve an understanding that you didn't have before... It is opening windows on the world... you perceive a little tiny glimpse of the way the Universe hangs together, which is a wonderful feeling

- Hans Kornberg

To my family

ACKNOWLEDGEMENTS

I take this opportunity to express my sincere gratitude and deep regards to my research advisor, Professor David M. Lubman for his valuable guidance, support and constant encouragement throughout the course of my Ph.D. study. I would also like to thank my committee members, Late Professor Richard Sacks, Professor Ari Gafni, Professor Ayyalusamy Ramamoorthy and Professor Kristina Hakansson for their precious time spent serving on my committee providing with helpful suggestions and advise.

My sincere thanks also to my collaborators Professor Steven Ethier and Dr. Arun Sreekumar for their help at providing with samples and materials, and specially Dr. David Misek and Rork Kuick for their suggestions, constructive criticism and invaluable help with my research.

The completion of my graduate career would not have been possible without the love and support I have received from my family, friends and colleagues. Words cannot express my gratitude for the encouragement and unconditional support I have received from my companion Jin Young and my family and relatives who made my dreams their own, my friends who took pride in my achievements and my colleagues for their companionship in making my laboratory research experience a graceful one.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

°C	degrees Celcius
1D	one-dimensional
2D	two-dimensional
2-DE	two-dimensional electrophoresis
2D-GE	two-dimensional gel electrophoresis
2D-LC	two-dimensional liquid chromatography
ACN	acetonitrile
BPH	benign prostatic hyperplasia
DMSO	dimethyl sulfoxide
DTT	dithiothreitol or (2S,3S)-1,4-Bis-sulfanylbutane-2,3-diol
Bis-Tris	2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol
BSA	bovine serum albumin
cDNA	complementary DNA
CAP	calf intestinal alkaline phosphatase
CE	capillary electrophoresis
CF	chromatofocussing
CHCA	α-cyanohydroxycinnamic acid
Cy3	cyanine-3 dye
Cy5	cyanine-5 dye
Da	Dalton
DAC	diammonium citrate
DNA	de-oxyribonucleic acid
EGFR	epithelial growth factor receptor
ESI	electrospray ionization
FA	formic acid
FGFR	fibroblast gorwth factor receptor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFBA	heptafluorobutyric acid
HPLC	high performance liquid chromatigraphy
ICAT	isotope coded affinity tags
IDA	imminodiacetic acid
IEF	isoelectric focussing
IPA	2-propanol
IT	ion trap

kDa	kilo Dalton
LC	liquid chromatography
LC-MS/MS	liquid chromatography tandem mass spectrometry
LOOCV	leave-one-out cross validation
MALDI	matrix-assisted laser desorption/ionization
MALDI-MS	matrix-assisted laser desorption/ionization mass spectrometry
MCM	molecular concept mapping
$MgCl_2$	magnesium chloride
mRNA	messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MudPIT	multidimensional analysis of protein identification technology
MW	molecular weight
m/z	mass to charge ratio
NaCl	sodium chloride
NaF	sodium fluoride
Na_3VO_4	sodium orthovanadate
NH_4HCO_3	ammonium bicarbonate
NH_4OH	ammonium hydroxide
NPS	nonporous
ODS	octadecyl silane(C18)
OG	n-octyl-β-D-glucopyrannoside
PBS	phosphate buffered sline
PBS-T	phosphate buffered saline with 0.1% Tween 20
PDAC	pancreatic adenocarcinoma
pI	isoelectric point
PCa	prostate cancer
PCA	principal component analysis
PMF	peptide mass fingerprinting
PMSF	phenylmethylsulphonyl fluoride
ppm	parts per million
PSA	prostate specific antigen
PTM	posttranslational modification
Q	quadrupole
QTOF	quadrupole time-of-flight
RNA	ribonucleic acid
RP	reversed phase
RP-HPLC	reversed chromatography high performance liquid chromatigraphy
rpm	revolutions per minute
SVM	support vector machine
TCEP	tris(2-carboxyethyl)phosphine

THAP	trihydroxy acetophenone
TIC	total ion chromatogram
TOF	time-of-flight
TPCK	L-1-tosylamido-2-phenyethyl chloromethyl ketone
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
TFA	trifluoroacetic acid
UV	ultraviolet
Xcorr	cross-correlation

CHAPTER I

Introduction

1.1 Proteomics in the Post-Genomic Era

The availability of the human genome map [1–3] has greatly enhanced our understanding of the underlying biology of disease progression and response and galvanized research in the rapidly advancing field of proteomics and biomarker discovery. Identifying, quantitating and characterizing all expressed proteins in the proteome is the ultimate goal for a deeper understanding of disease response at a molecular level. All genetic mutations which gives rise to disease are ultimately manifested at the protein level. These are characterized by derangements in protein function and information flow within diseased cells and the interconnected tissue micro-environment. Thus, the study of proteins altered in the course of a disease holds great importance toward realizing this goal. These changes in protein expression can not only reveal biomarkers for the diagnosis of diseases but also provide novel therapeutic targets for more effective personalized cure. Moreover, such studies can reveal valuable information about the underlying biological processes, such as perturbations in protein signaling pathways.

Genomics or transcriptomics is a common way to study diseases such as cancer, but

recent studies point out that the correlation between mRNA and protein levels in cells of most organisms is remarkably and unexpectedly low [4–7]. It implies therefore, that mRNA studies are less predictive of complex traits than protein studies. Though it is too early to completely discount mRNA studies and favor protein expression profiling [8] nonetheless, proteomics offers an alternative diagnostic platform for analysis of excreted proteins and body fluids like blood serum which is essential for pathological applications. As such, studies in determining the protein content of the cell are important in attempting to understand cellular processes in cancer. Though current multidimensional separations and mass spectrometry platforms can rapidly generate a high resolution map of the proteome, we are still far from deciphering cellular functions that are maintained by proteins. The study of the proteome poses great challenges due to its complexity and dynamic range. Though estimated to originate from around 40,000 genes [9], there are close to an estimated 1 million proteins in the human proteome and the dynamic range in their expression levels exceeds 10 orders of magnitude [10, 11].

The low correlation in mRNA and protein concentration in cells is generally hypothesized to result from post-translational modifications in proteins [12] which seems to be more prevalent than previously assumed. Additionally, numerous isoforms [13] add up to the complexity of the proteome. Proteins undergo post-translational modifications, cleavage and degradation in response to various cell signals both for maintaining normal cellular functions as well as in response toward diseases [14, 15]. In this respect, phosphorylations are most important and are directly responsible for regulating cellular signaling pathways in cells, where any alterations in which can lead to cancer [16]. Glycosylations also are common and are a heterogeneous class

of PTMs playing a key role in cellular recognition which is responsible for normal functioning in cells. Defects in glycosylation mediated signaling can also result in disease [17]. Alternative splicing of mRNA which is the proposed mechanism by which higher order diversity is created among proteins [18] for instance, can produce many different proteins from a single gene. As many as 30% of the genes in humans and other eukaryotes are thought to be alternatively spliced. Splice variant proteins are known to display the same, opposite or completely different and unrelated physiological activity. This in turn affects key non biological factors such as stability, clearance rate, cellular localization, temporal pattern of expression, up-regulation or down-regulation mechanisms and response to agonists or antagonists which are critical aspects in the studies of disease progression.

1.2 Applications of Proteomics in Cancer Research

Cancer has been one of the most widely studied diseases using proteomics. Cancer is not a single disease but an accumulation of several events, genetic and epigenetic, arising in a single cell over a long time interval. A high priority has been attached to the identification of these events. This can be achieved by characterizing cancer-associated genes and their protein products. Identifying the molecular alterations that distinguish a cancer cell from a normal cell will ultimately help in defining the nature [19] and predict the pathologic behavior [20] of a cancer cell. It will also indicate the responsiveness to treatment of that particular tumor. Understanding the profile of molecular changes in cancer is extremely useful to be able to correlate the phenotype of cancer with molecular events. Achieving these goals will provide an opportunity for discovering new biomarkers for early detection of cancer and de-

veloping approaches for prevention. Early detection is a difficult challenge for proper diagnosis and prevention, since in many cases, cancer is not diagnosed and treated until cancer cells have become invasive or metastatic [21]. Early detection could then enable effective interventions and therapies contributing to reduction in mortality and morbidity. Through the course of progression of cancer, biomarkers serve as molecular signposts of the physiologic state of the cell [22] and are therefore truly dynamic unlike the genome. Biomarkers could prove to be vital for the identification of early cancer and subjects at risk of developing cancer, though presently, biomarkers that allow precise monitoring or classification of disease are very limited. The discovery of new highly sensitive and specific biomarkers for early detection of disease and development of personalized therapies holds the key to effective treatment of diseases. Apart from cells and tissues, the circulatory proteome contains a rich source of information that is helpful both in the early detection of disease state and risk assessment [23–27]. Being easily obtainable through non invasive techniques, biofluids are well suited for pathological applications. It is important then, that any analytical method for the analysis of biofluids must be robust enough to deal with the associated complexities for effective pathological applications.

1.3 The Mass Spectrometry Advantage

Current multidimensional separation and mass spectrometry based platforms for proteomics can rapidly generate a high resolution map of the proteome. Though developments in separation technologies had been crucial toward attaining that capability, the ability to progressively detect lower concentrations of proteins in biological samples has largely been brought about by the recent advances made in the field of mass

spectrometry. The applicability of mass spectrometry for the analysis of peptides as well as large biomolecules was greatly improved through the introduction of two soft ionization techniques, MALDI [28,29] and ESI [30,31]. Though the exact mechanism of generation of charged analytes in MALDI is not clear, nevertheless, MALDI was successfully introduced for the analysis of peptides and proteins commonly using a TOF based analyzer which is suited for pulsed techniques. Peptide mass fingerprinting (PMF) obtained using MALDI-TOF MS is the most common method for rapid identification of proteins in which the pattern of peptide m/z values obtained through MS analysis of a proteolytically digested protein is compared against a database of theoretical fragmentations to identify the original protein. Though MALDI-MS is a fast and efficient method capable of identifying proteins from very small amounts of sample, it is not suitable for identifications of PTMs for which MS/MS techniques are more suited.

Electrospray ionization (ESI) produces gas phase ions from analytes in the liquid phase using an electric field. It also generates multiply-charged ions that not only allow determination of accurate molecular weights but also provide the ability to detect large molecules using an analyzer with limited mass range. Moreover it has excellent capabilities of online interfacing with chromatographic and other various liquid separation techniques. Online LC-ESI-MS can add another dimension to liquid separations thereby increasing the peak capacities of this hyphenated technique. Both MALDI and ESI had been effectively used for peptide sequencing where the peptide chains are fragmented commonly using ion-neutral collisions and the m/z of the fragments are measured and queried against a theoretically generated database for protein identification. Recently, using nano-electrospray combined with ion traps

(IT), it had been possible to detect proteins at sub-picomolar levels [32]. Moreover, the ability to conduct mass spectrometric analysis at MS^3 level make it valuable for detection of PTMs. The use of hybrid MALDI-IT-TOF instruments on the other hand have demonstrated the ability to obtain information on glycoprotein structures and attain sensitivities of low femtomolar levels.

1.4 Multidimensional Separation Technologies

The advances in analytical techniques that were driven by the needs of the post-genomic era have provided us with the ability to analyze biological samples for biomarkers in ways never before possible. With a huge array of potential methodologies with very unique capabilities, it is not often clear however, which of these analytical technologies or a combination thereof, will yield the most comprehensive results. 2D-PAGE has been one of the principal tools for proteomics since its inception [33]. It enabled high resolution separation of proteomes where the spot patterns between two or more samples could be compared for differences and analyzed thereafter using mass spectrometry [34, 35]. However, despite its resolving power, 2D-PAGE has significant limitations with respect to throughput, reproducibility, mass resolution and dynamic range, making it far from an ideal tool for biomarker discovery. Also, large amounts of manual labor involved in this method make it difficult to automate. Alternative liquid based multidimensional separation technology has been used effectively as a means for fractionating and purifying protein fractions. This technique is easily adaptable to mass spectrometry and several preparative as well as analytical scale separation methodologies have been used to study various biological samples. Besides the applicability in top-down approaches, this technique

makes possible bottom-up analysis where whole cell lysates are digested and the peptide mixture is separated and analyzed using mass spectrometry. Commonly used bottom-up approaches or MudPIT use strong cation exchange followed by reversed phase separation to fractionate the peptide mixture [36]. Alternative 1-D separation of complex peptide mixtures using a long reversed phase capillary HPLC column can also be performed [37], but in general, protein identifications obtained through bottom-up methods have high false positive rates since they are identified using too few peptides [37]. The most important aspect though, is the loss of valuable information at the protein level unlike in top-down approaches where intact proteins are digested separately so that in-depth mass spectrometry based analysis is possible.

1.4.1 Liquid Chromatography

The most widely used fractionation technique applied for top-down proteomics uses a 2-D separation method where chromatofocusing (CF) [38] is used in the first dimension and nonporous reversed phase (NPS-RP) HPLC is applied for separation in the second dimension. CF is performed on a silica based weak anion exchange column where proteins are loaded on the column at a higher pH and then eluted gradually using a low pH buffer. The mixing of the two buffers during elution creates a pH gradient inside the column eluting out proteins in the order of decreasing pI. This 2-D LC separation scheme has been successfully applied for the study of human cancer and bacterial cells where proteins were identified using ESI-TOF and MALDI-TOF MS based techniques [39–41]. The intact MW of the proteins can also be used to create virtual 2-D maps resembling gels which can then be used for interlysate comparisons. Unlike in gels, the experimental conditions for 2-D LC can be maintained easily so as to produce highly reproducible maps. The use of non porous silica in the

second dimension reversed phase column eliminates irreversible protein binding to the stationary phase thereby providing higher recovery. A short column packed with such nonporous silica C18 material enables fast separations and provides sufficient peak capacity to reliably fractionate pH fractions obtained using chromatofocusing. One disadvantage of CF with respect to gels is the pH limitation associated with the usage of silica based stationary phase. CF performed using columns with polymeric stationary phases on the other hand are free from such limitations. Since the 2D-LC fractionated proteins are obtained in intact form, they can also be collected off-line for characterization of PTMs. Typically collected protein fractions are digested and subjected to MALDI-TOF MS. Due to a number of factors associated with sample preparation that affect MALDI ionization and sensitivity, ESI-MS/MS is commonly performed to verify protein identifications. Also proteins are commonly subjected to ESI-MS/MS analysis to obtain information about sequence variations and modifications because the peptide sequence coverage obtainable from MALDI-MS analysis is usually low. The sensitivity levels offered by ESI-MS/MS can be greatly enhanced when interfaced with liquid chromatography at very low flow rates.

1.4.2 Capillary Electrophoresis

Capillary Electrophoresis (CE) separation is based on differential migration of analytes that arises from the differences in electrophoretic mobility [42] determined by charge, size and shape of the ions in the liquid phase [43]. CE demonstrates a very high separation efficiency where electroosmotic flow (EOF) caused by an electrical double layer formed at the stationary/solution interface inside a silica capillary, generates a uniform flow profile across the cross-section and length of the capillary minimizing band broadening. Capillary electrophoresis separations can have several

hundred thousand theoretical plates. The efficiency of CE separation is only limited by diffusion and is proportional to the strength of the electric field. Applying a high voltage during CE separation also helps achieve high speeds in separation of complex peptide mixtures so that a typical separation can be completed in a few minutes. The problem though with fast separations is the introduction of Joule heating from high voltage which can adversely affect peak resolution and separation efficiency. For proteomic applications, CE is normally applied on samples that had been separated in one or more dimensions using isoelectric focusing, capillary gel electrophoresis and other methods [43]. The slow flow rate of CE had been successfully applied in interfacing it with mass spectrometry using electrospray ionization as well as MALDI [44]. CE-MS had been applied for both identification [45] and characterization of important modifications [46–48] in biological samples.

Despite the excellent resolution and speed of CE, it has exhibited several problems when interfaced with ESI-MS. When CE is used for MS analysis in positive ion mode, applying a low pH condition to generate the ions, the inner capillary wall must be modified to minimize interaction between negatively charged silanol groups and the positively charged analytes [49]. A dynamic coating procedure based on adsorption as described in Chapter V does not provide sufficiently long lifetime. The unstable EOF at low pH gives poor reproducibility in analyte migration times and poses a serious challenge when comparison of several runs becomes necessary. The interfacing of CE to MS is also complicated. Sheathless flow has been utilized in the work in Chapter V to obtain maximum sensitivity by preventing sample dilution and associated reduced sensitivity which occurs when make-up flow is used in sheath-flow interfaces [50]. The sheathless interface created by attaching an emitter at the end

of the separation column is prone to detachment on contact with liquid phase column effluent making it difficult to maintain a closed circuit in the CE-ESI setup. Moreover, analyte samples must be highly purified to avoid interruption of voltage gradients from column clogging.

1.4.3 Monolithic Capillary HPLC

HPLC separations using monolithic capillary columns prepared with either silica or other polymers [51–54] provides with an alternative method for separation of peptides in protein digests with a very high efficiency comparable to that of CE. Monolithic LC is an emerging separation method which has been successfully applied to the analysis of biomolecules including nucleic acids, RNA, proteins and peptides [55–57]. Monolithic capillary LC columns provide unique characteristics of high speed, high resolution, high efficiency and high recovery rates because of fast mass transfer owing to lack of interstitial space [58]. These separation qualities are comparable to that of CE. Monolithic capillary HPLC moreover, allows for higher loading capacity and is a much simpler technique to interface with ESI-MS, thus providing a highly robust, rugged and reproducible analysis tool. Polymer-based monolithic columns have higher stability at extreme conditions [59] and have also been used for analysis of phosphorylations using alkaline solvent systems for detection with MS in negative ion mode [60]. Monolithic capillary LC has also been applied for separations utilizing affinity chromatography techniques [61,62] and as support material for enzymatic digestions [63] apart from more common applications in quantitative analysis of human serum proteins [64] and in peptide mapping [65]. Monolithic capillary based separation is an ideal technique when analysis on a small amount of sample is desired. The high recovery rate can provide better sensitivity for very low sample amounts.

1.5 Microarray Technology

A concerted effort by scientists and engineers from many different fields helped develop DNA microarrays, a key technology in the field of genomics that made possible the monitoring of expression levels of all genes in an organism simultaneously [66–70]. The key element in this new technology was the development of surface based assays in which numerous probes are immobilized in a spatially addressable manner [71]. Such array formats were suitable for miniaturization and multiplexing. Though the concept of microarrays was first introduced in 1989 [72], the term 'microarray' was not used widely till much later [73, 74].

The principle of miniaturized ligand binding assay was first described by Ekins et al. almost two decades ago [72] who argued that a miniaturized assay with 'microspots' of immobilized capture molecules on solid phase would be more sensitive than conventional macroscale methods. Although the amount of capture molecules present in a 'microspot' is low, a high density of molecules can be obtained. Due to a limited number of capture molecules, only a small number of analyte molecules can be captured during an assay procedure so the concentration of free analyte in the solution is not changed much by the binding reaction. This phenomenon termed 'ambient analyte condition' ensures high sensitivity [72]. Since the analyte molecules are confined to a very small area, microspot assays result in a much higher sensitivity compared to other 96-well plate based macro assays. As a result, femto-molar concentrations of antigens could be easily detected [75]. Miniaturization also allowed for parallelization which in combination with higher sensitivity provided microarrays

with an enormous potential in diagnostic applications.

Technologies established for DNA chips were adapted for microarray based research. Although DNA microarray provided the leads in the development of protein microarrays, the methodology for the latter were significantly different. The main reason for this is the tendency of proteins to undergo denaturation and exhibit nonspecific binding. Miniaturization made these issues more complex, since the surface to volume ratio increases dramatically when the volume of a sample spot is scaled down. Despite these difficulties, important advances in methods and technology enabled the use of protein microarrays for various applications.

Initial reports demonstrated the feasibility of antibody microarrays using a variety of methods that had included spotting on membranes [76–79], derivatized glass slides [78, 80–82] and hydrogels [83, 84] and detection of bound antibodies using radioactive isotopes [76], fluorescence [80–84] and chemiluminiscense [78, 79, 84, 85]. These kinds of experiments utilizing the protein microarray format could be broadly categorized into two classes - direct labeling experiments, and dual antibody sandwich assays. In the direct labeling method, all proteins in a complex mixture are covalently labeled with a fluorescent tag. After incubation on an antibody microarray, the tag provides a means of detecting the bound proteins. The signal from the bound proteins can also be amplified using suitable tag chemistry. In the sandwich assay, proteins captured on an antibody microarray are detected by a cocktail of detection antibodies. The detection antibodies are in turn detected through fluorophore-labeled secondary antibodies. The disadvantage of direct labeling experiments is the potential for high background since all proteins are labeled so the

sensitivity achievable is not very high.

1.5.1 Applications in Diagnostics

Protein microarray immunoassays offer an attractive alternative when several parameters of a single sample have to be analyzed in parallel, such as in allergy [86,87] or autoimmune diagnostics in which patient sera needs to be screened for a number of different auto antibodies [78,88]. Auto antibodies toward immobilized auto antigens used as diagnostic markers for autoimmune conditions can be accurately determined from less than 1 μL of patient serum. This reflects the enormous potential of protein microarrays employed to study the humoral response against a large number of antigens. Microarrays have also been successfully used to detect the presence of specific IgG and IgM antibodies directed against parasitic and viral antigens [89–91]. The analytical sensitivity of these assays were similar to those obtained using standard ELISA technology [89]. Sandwich immunoassays were therefore adapted to the microarray format [85,92,93] creating highly specific and sensitive protein microarrays which were capable of quantifying many different cytokines from patient sera. However the parallelization of sandwich immunoassays has limitations from cross reactivity of certain detection antibodies making the routine use of highly multiplexed sandwich immunoassays difficult.

Despite the problem with antibody cross reactivity which limits the scope of certain experiments when compatible antibodies are unavailable, the potential of array based proteomic approaches is enormous [94,95]. Protein microarray technology had been used to simultaneously determine levels of large numbers of target proteins using comparative methods [96] where the array bound proteins are usually detected

using biotin-based signal amplification [97]. Protein microarrays were also applied to discovery of diseases and tumor markers [98]. These approaches reflect the power of antibody microarrays in determining changes in protein expression in a single experiment.

Protein microarrays also offer the opportunity to study protein-protein interactions by immobilizing purified recombinant proteins. Purified proteins, enriched protein fractions or complete cell lysates had been used for interaction assays [99]. These experiments demonstrated the stability of microarrays to screen for protein-protein interactions at a proteome wide level [100–103] and similar approaches could be applied to study protein-drug and protein-lipid interactions which were difficult to study using other approaches.

1.5.2 Reversed Phase Microarrays

In contrast to the techniques described above, cellular lysates prepared from cultured cells or tissues can also be immobilized on a microspot and screened with specific antibodies for the presence of defined target proteins. Characteristic features of reversed phase microarrays include high linearity and excellent sensitivity. The biggest advantage is that the samples need not be labeled [104]. Reversed phase arrays can use denatured lysates so that the retrieval of antigens do not pose problems. Non denatured lysates can also be used to identify the target protein of interest as well to elucidate protein-protein, protein-DNA and/or protein-RNA interactions. The samples arrayed in dilution series can provide an internal standard and direct quantitative assessments can also be made by including several positive and negative controls and internal calibration standards. Because measurements lie within the

linear dynamic range of the antibody-analyte interaction at any given point in the dilution curve [105], direct quantitative measurements can be made using reverse phase microarrays. Rather than arraying cell lysates, where the identity of the binding protein to an antibody probe is difficult to ascertain, fractionated lysates offer a remarkable methodology where multidimensional LC fractionated cellular or tissue lysates provide purified proteins for spotting on to the array [106]. Though numerous methods had been used for obtaining proteins for the purpose of spotting [92,99,107], 2-D LC fractionation provides the most robust technique thus far. This method has the advantage associated with obtaining proteins with biologically relevant PTMs that is difficult to obtain using other means.

The use of protein microarrays allows the measurement of several parameters in one reaction. For systems like autoimmune or humoral response assay, the degree of multiplex achievable is limited to the number of antigens available. The sensitivity of such systems is high and so the amount of serum required can be <1 μL. Autoimmune assays work with serum dilutions of 1:200 to 1:1000. Nevertheless, we should keep in mind that the results obtained with antibody microarrays must be verified and confirmed since some antibodies exhibit strong cross reactivity. In addition, proteins are often assembled in multi-protein complexes. Thus a strong signal on a microspot can result not only from the presence of a large number of target molecules but also from the nonspecific capture of a labeled dye molecule or antibody. Standard methods such as immuno-histochemical staining and immunobloting are commonly used for validation.

1.5.3 Analysis of Post-translational Modifications

As described earlier is Section 1.1 protein post translational modifications specifically phosphorylation is critical in maintaining cellular functions, and methodologies for sensitive and accurate phosphoprotein analysis are very essential. Though the undeniable strengths of large scale mass spectrometry based approaches cannot be ignored, protein microarrays offer a complementary but faster and more sensitive method for detection of protein modifications. Sandwich assays have been applied successfully for large scale phosphoprotein analysis. Techniques like probing antibody spotted arrays with cellular lysates followed by detection using fluorophore conjugated anti-phosphotyrosine antibodies [108] had not been very popular owing to the difficulties in generating specific antibodies but holds great promise for in-depth analysis. Using such methods and using only nanogram quantities of total protein extracts, a ratio metric study could be conducted to obtain differential phosphorylation patterns. Alternative reversed phase protein microarray methods have relied on anti-phosphoprotein antibodies, but the drawback is the 'ligand problem' which revolves around the unavailability or inability at synthesizing high quality antibodies so as to eliminate non-specific interactions [109]. Recent developments in alternative dye based modification detection arrays [110] have received much attention and work presented in this thesis describes methodologies using the dye-based approach for differential detection of phosphorylations across cellular proteomes. Other PTM detection methodologies eg. glycoprotein microarray strategies employing lectins [111,112] had also been applied for biomarker discovery using 2-D LC fractionated cell lysates. These applications serve to present protein microarrays as an efficient platform for proteome wide PTM discovery and quantitative detection.

Since traditional PTM detection relies on immunoassay based techniques the limits of detection are dictated by antibody recognition chemistry and nonspecific complexation with fluorophores may give rise to false positives. As a result, alternative mass spectrometry based detection strategies [109] using isotope-labeling techniques had been proposed [113], however, it cannot be applied to protein-detecting arrays. In-vivo protein labeling followed by direct MS can be a better alternative which needs to be explored further.

1.6 Molecular Concept Modeling

The power of microarray based approaches lie in the ability to provide genome or proteome wide expression patterns otherwise impossible to obtain. Experiments using this ability had largely focused on the differential expression of disease related biomarkers but recently, construction of disease response pathways had been demonstrated [114]. This methodology uses an analytical framework for exploring the network of relationships among a growing collection of 'molecular concepts', or biologically related gene sets [115]. As a result, visualization of disease specific signaling networks has become possible. This not only opens up a new dimension in disease classification and diagnosis but also demonstrates the versatility of microarray based parallel techniques. Work described in this thesis demonstrates the use of the MCM technique for generating protein signaling networks and associated applications in cancer diagnosis. The inclusion of protein modification information in this analytical model would provide us with a truly versatile technique.

1.7 Statement of Research

This thesis discusses research in multiplexed high-throughput proteomic approaches to identify molecular signatures of cancer and post translational modifications in cell-line and tissue proteomes. 2-D LC fractionated human breast cancer SUM-52PE cell-line, pancreatic cancer Panc-1 cell-line and prostate cancer tissue have been used in the work described in Chapters II, III and IV respectively to generate protein microarrays which were then used for phosphorylation detection or autoantibody response mapping. Chapter II demonstrates the use of a small molecule phosphoprotein dye for global phosphoprotein detection. Around 100 differential phosphorylations were detected and differentially phosphorylated proteins were identified using ESI-TOF MS and MALDI-TOF MS, the later employing a modified MALDI matrix.

Chapters III and IV discuss autoantibody response based methods for biomarker discovery and disease response pathway discovery respectively. Panc-1 cell-lines were used in the biomarker discovery experiment since pancreatic cancer poses a challenge for early detection and diagnosis. nESI-LC-MS/MS was used for protein identification and glycosylation patterns were obtained using lectin microarrays to study possible correlations with autoantibody response. Chapter IV presents the use of bioinformatics using a molecular concept modeling based method for obtaining disease progression pathways in prostate cancer.

Chapters V and VI presents monolithic capillary HPLC based techniques for analyzing peptides and proteins providing high sequence coverage and sensitivity. High sequence coverage is essential for detecting reliable identification of proteins and iden-

tifying PTMs. Applying the monolithic LC based method, several modifications in MFC10A derived human breast cancer cell lysates were detected and characterized using ESI-TOF MS and MALDI-TOF MS.

Lastly, Chapter VI describes a hyphenated technology where monolithic LC separation of intact proteins from human esophageal tissue samples was combined with on-plate digestion and MALDI-MS based protein identification. This work describes a method which retains the advantages of top-down proteomics and at the same time uses automation to increase throughput.

References

[1] Venter J. C. et. al. *Science*, 291:1304, 2001.

[2] Lander E. S. et. al. *Nature*, 409:860, 2001.

[3] Pandey A.; Mann M. *Nature*, 405:837–846, 2000.

[4] Anderson L.; Seilhamer J. *Electrophoresis*, 18:533–537, 1997.

[5] Gygi S. P.; Rochon Y.; Franza B. R.; Aebersold R. *Mol. Cell. Biol.*, 19:1720, 1999.

[6] Ideker T. et al. *Science*, 292:929–934, 2001.

[7] Ghaemmaghami S.; Huh W. K.; Bower K; Houson R. W.; Belle A.;Dephoure N.; O'Shea E. K.; Weissman J. S. *Nature*, 425, 2003.

[8] Jansen R. C. and Nap J. P. *Nat. Biotechnol.*, 20(1):19, 2002.

[9] Rappsilber J.; Mann M. *Trends Biochem. Sci.*, 27:74–78, 2002.

[10] Jacobs J. M.; Adkins J. N.; Qian

[11] Issaq H. J.; Chen K. C.; Janini G. M.; Conrads T. P.; Veenstra T. D. *J. chromatogr. B*, 817:35, 2005.

[12] Jansen R. C.; Nap J. P. and Mlynarova L. *Nat. Biotech.*, 20(1):19, 2002.

[13] Harry J. L.; Wilkins M. R.; Herbert B. R.; Packer N. H.; Gooley A. A.; Williams K. L. *Electrophoresis*, 21:1071, 2000.

[14] Nakal K *J. Str. Biol.*, 134:103–116,2001.

[15] Harrison P. M.; Kumar A.; Lang N.; Snyder M.; Gerstein M. *Nucleic Acids Res.*, 30:1083, 2002.

[16] Gschwind A.; Fischer O. M.; Ullrich A. *Nat. Rev.*, 4:361–370, 2004.

[17] Kobata A.; Amano J. *Immunol. Cell Biol.*, 83:429–439, 2005.

[18] Brett D.; Pospisil H.; Valcarcel J.; Reich J. and Bork P. *Nat. Genet.*, 30:29–30, 2001.

[19] Vlahou A.; Fountoulakis M. *J. Chromatogr. B*, 814:11, 2005.

[20] Wu W.; Hu W.; Kavanagh J. J. *Int. J. Gyenocol. Cancer*, 12:409, 2002.

[21] Wulfkuhle J. D.; Liotta L. A.; Petricoin E. F. *Nature Rev.*, 3:267, 2003.

[22] Srivastava S.; Srivastava R. G. *J. Proteome Res.*, 4:1098, 2005.

[23] Laulu S. L. and Roberts W. L. *Am J Clin Pathol.*, 127(3):436–40, 2007.

[24] Vitali A.; Ardoino S.; Durante P.; Ferro M. A.; Li C. F.; Parodi C.; Sanguineti G.; Gaffuri M.; Paerachino M.; Salvadori R. P. *Anticancer Res.*, 14(4A):1503–1507, 1994.

[25] Catalona W. J.; Partin A. W.; Slawin K. M.; Brawer M. K.; Flanigan R. C.; Patel A.; Richie J. P.; deKernion J. B.; Walsh P. C.; Scardino P. T.; Lange P. H.; Subong E. N.; Parson R. E.; Gasior G. H.; Loveland K. G. and Southwick P. C. *JAMA*, 279(19):1542–1547, 1998.

[26] Billis A.; Meirelles L. R.; Maqna L. A.; Baracat J.; Prando A.; Ferreira U. *Urology*, 69(5):927–30, 2007.

[27] Wang X.; Yu J.; Sreekumar A.; Varambally S.; Shen R.; Giacherio D.; Mehra R.; Montie J. E.; Pienta K. J.; Sanda M. G.; Kantoff P. W.; Rubin M. A.; Wei J. T.; Ghosh D. and Chinnaiyan A. M. *N. Engl. J. Med.*, 353(12):1224–1235, 2005.

[28] Tanaka K.; Waki H.; Ido Y.; Akita S.; Yoshida Y.; Yoshida T. *Rapid Commun. Mass Spectrom.*, 2:151–153, 1988.

[29] Karas M.; Hillenkamp F. *Anal. Chem.*, 60:2299–2301, 1988.

[30] Fenn J. B.; Mann M.; Meng C. K.; Wong S. F.; Whitehouse C. M. *Science*, 246:64–71, 1989.

[31] Gaskell S. J. *J. Mass Spectrom.*, 32:677, 1997.

[32] Kenyon G. L.; DeMarini D. M.; Fuchs E.; Galas D. J.; Kirsch J. F.; Leyh T. S.; Moos W. H. et al. *Mol. Cell. Proteomics*, 1:763780, 2002.

[33] O'Farrell P. H. *J. Biol. Chem.*, 250:4007-4021, 1975.

[34] Shevchenko A.; Wilm M.; Vorm O.; Mann M. *Anal. Chem*, 68:850, 1996.

[35] Gygi S. P.; Corthals G. L.; Zhang Y.; Rochon Y.; Aebersold R. *Proc. Natl. Acad. Sci. USA*, 97:9390, 2000.

[36] Washburn M. P.; Wolters D.; Yates J. R. *Nat. Biotechnol.*, 19:242–247, 2001.

[37] Shen Y.; Zhang R.; Moore R. J.; Kim J.; Metz T. O.; Hixson K. K.; Zhao R.; Livesay E. A.; Udseth H. R.; Smith R. D. *Anal. Chem.*, 77:3090, 2005.

[38] Liu Y. S.; Anderson D. J. *J. Chromatogr. A*, 762:47–54, 1997.

[39] Hamler R. L.; Zhu K.; Buchanan N. S.; Kreunin P.; Kachman M. T.; Miller F. R.; Lubman D. M. *Proteomics*, 4:562–577, 2004.

[40] Wang H.; Kachman M. T,; Schwartz K. R.; Cho K. R.; Lubman D. M. *Proteomics*, 4:2476, 2004.

[41] Zheng S.; O'Neil K. A.; Barder T. J.; Lubman D. M. *Biotechniques*, 35:1202, 2003.

[42] Jorgenson J. W. and Lukacs K. D. *Anal. Chem.*, 53:31, 1981.

[43] Monning C. A. and Kennedy R. T. *Anal. Chem.*, 66:280R, 1994.

[44] Preisler J.; Hu P.; Rejtar T.; Karger B. L. *Anal. Chem.*, 72(20):4785-95, 2000.

[45] Manabe T. *Electrophoresis*, 20:3116, 1999.

[46] Jin X.; Kim J.; Parus S.; Lubman D. M.; and Zand R. *Anal. Chem.*, 71:3591, 1999.

[47] Kim J.; Zand R. and Lubman D. M. *Electrophoresis*, 23:782, 2003.

[48] Zhu K.; Kim J.; Yoo C.; Miller F. R. and Lubman D. M. *Anal. Chem.*, 75:6209, 2003.

[49] Bateman K. P.; White R. L.; Thibault P. *Rapid Comn. Mass Spectrom.*, 11:307, 1997.

[50] Issaq H. J.; Janini G. M.; Chan K. C. and Veenstra T. D. *J. Chomatogr. A*, 1053:37, 2004.

[51] Hjerten S.; Liao J. L.; Zhang R. *J. Chroamtogr.*, 473:273, 1989.

[52] Barroso B.; Lubda D.; Bischoff R. *Proteome Res.*, 2:633, 2003.

[53] Walcher W,; Toll H.; Ingendoh A.; Huber C. G. *J. Chromatogr. A*, 1053:107, 2004.

[54] Svec F. *J. Sep. Sci.*, 27:747, 2004.

[55] Premstaller A.; Oberacher H; Walcher W.; Timperio A. M.; Zolla L.; Chervet J. P.; Cavusoglu N.; Dorsselaer A. V. and Huber C. G. *Anal. Chem.*, 73:2390, 2001.

[56] Oberacher H.; Huber C. G. *Trends Anal. Chem.*, 21:2002, 2002.

[57] Holzl G.; Oberacher H.; Pitsch S.; Stutz A,; Huber C. G. *Anal. Chem.*, 77:673, 2005.

[58] Gritti F.; Piatkowski W.; Guiochon G. *J. Chromatogr. A*, 983:51, 2003.

[59] Oberacher H.; Premstaller A.; Huber C. G. *J. Chromatogr. A*, 1030:201, 2004.

[60] Tholey A.; Toll H.; Huber C. G. *Anal. Chem.*, 77:4618, 2005.

[61] Josic D.; Buchacher A. *J. Biochem. Biophys. Methods*, 49:153, 2001.

[62] Bedair M.; Rassi Z. E. *J. Chromatogr. A*, 1044:177, 2004.

[63] Peterson D. S.; Rohr T.; Scec F.; Frechet J. M. J. *J. Proteome Res.*, 1:563, 2002.

[64] Mayr B. M.; Kohlbacher O.; Reinert K.; Sturm M.; Gropl C.; Lange E.; Clein C.; Huber C. G. *J. Proteome Res.*, 5:414, 2006.

[65] Hennessy T. P.; Boysen R. I.; Huber M. I.; Unger K. K.; Hearn M. T. W. *J. Chromatogr. A*, 1009:15, 2003.

[66] DeRisi J. L.; Iyer V. R.; Brown P. O. *Science*, 278:680–686, 1997.

[67] The Chipping Forecast *Nat. Genet.*, 21 Suppl:1–60, 1999.

[68] The Chipping Forecast *Nat. Genet.*, 32 Suppl:461–452, 2002.

[69] Kapranov P.; Cawley S. E.; Drenkow J,; Bekiranov S.; Strausberg R. L.; Fodor S. P. and Gingeras T. R. *Science*, 296:916–919, 2002.

[70] Cawley S.; Bekiranov S.; Ng H. H.; Kapranov P.; Sekinger E. A.; Kampa D.; Piccolboni A.; Sememtchenko V.; Cheng J.; Williams A. J.; Wheeler R.; Wong B.; Drenkow J.; Yamanaka M.; Patel S.; Brubaker S.; Tammana H.; Helt G.; Struhl K. and Gingeras T. R. *Cell*, 116:499–509, 2004.

[71] Wilson D. S.; Nock S. *Curr. Opin. Chem. Biol.*, 6:81–85, 2001.

[72] Ekins. R. P.; et. al. *J. Pharm. Biomed. Anal.*, 7:155–168, 1989.

[73] Ferguson J. A.; Boles T. C.; Adams C. P. and Walt D. R. *Nat. Biotech.*, 14:1681–1684, 1996.

[74] Schena M.; Shalon D.; Davis R. W. and Brown P. O. *Science*, 270:467–470, 1995.

[75] Finckh P; Berger H.; Karl J.; Eichenlaub U.; Weindel K.; Hornauer H.; Lenz H.; Sluka P.; Weinreich G. E.; Chu F. And Ekins R. *Proc. U.K. Natl. Ext. Qual. Assess. Serv. Meeting*, 3:155–165, 1998.

[76] Ge H. *Nucleic Acids Res.*, 28:e3, 2000.

[77] DeWildt R. M. T.; Mundy C. R.; Gorick B. D.; Tomlinson I. M. *Nat. Biotechnol.*, 18:989–994, 2000.

[78] Joos T. O.; Schrenk M.; Hopfl P.; Kroger K. et al. *Electrophoresis*, 21:2641–2650, 2000.

[79] Knezevic V.; Leethanakul C.; Bichsel V. E.; Worth J. M. et al. *Proteomics*, 1:1271–1278, 2001.

[80] MacBeath G.; Schriber S. L. *Science*, 289:1760–1763, 2000.

[81] Haab B. B.; Dunham M. J.; Brown P. O. *Genome Biol.*, 2:1–13, 2001.

[82] Rowe C. A.; Scruggs S. B.; Feldstein M. J.; Golden J. P. et al. *Anal. Chem.*, 71:433–439, 1999.

[83] Guschin D.; Yershov G.; Zaslavsky A.; Gemmell A. et al. *Anal. Biochem.*, 250:203–211, 1997.

[84] Arenkov P.; Kukhtin A.; Gemmell A.; Voloschuk S. et al. *Anal. Biochem.*, 278:123–131, 2000.

[85] Huang R. P. *Clin. Chem. Lab. Med.*, 39:209–214, 2001.

[86] Wiltshire S.; O'Malley S.; Lambert J.; Kukanskis K.; Edgar D.; Kingsmore S. F.; Schweitzer B. *Clin. Chem.*, 46:1990–1993, 2000.

[87] Bacarese-Hamilton T.; Mezzasoma L.; Ingham C.; Ardizzoni A.; Rossi R.; Bistoni F.; Crisanti A. *Clin. Chem.*, 48:1367–1370, 2002.

[88] Robinson W. H.; DiGennaro C.; Hueber W.; Haab B. B.; Kamachi M.; Dean E. J.; Fournel S.; Fong D.; Genovese M. C. et al. *Nat. Med.*, 8:295–301, 2002.

[89] Mezzasoma L.; Bacarese-Hamilton T.; DiChristina M.; Rossi R.; Bistoni F.; Chrisanti A. *Clin. Chem.*, 48:121–130, 2002.

[90] Opalka D.; Lachman C. E.; MacMullen S. A.; Jansen K. U.; Smith J. F.; Chimule N.; Esser M. T. *Clin. Diagn. Lab. Immunol.*, 10:108–115, 2003.

[91] Bacarese-Hamilton T.; Ardizzoni A.; Gray J.; Crisanti A. *Methods Mol. Biol.*, 278:271–284, 2004.

[92] Schweitzer B.; Roberts S.; Grimwade B.; Shao W.; Wang M.; Fu Q.; Shu Q.; Laroche I.; Zhou Z.; Tchernev V. T.; Christiansen J.; Velleca M.; Kingsmore S. F. *Nat. Biotechnol.*, 20:359–365, 2002.

[93] Tam S. W.; Wiese R.; Lee S.; Gilmore J.; Kumble K. D. *J. Immunol. Methods*, 261:157–165, 2002.

[94] Emili A. Q.; Cagney G. *Nat. Biotechnol.*, 18:393–397, 2000.

[95] Templin M. F.; Stoll D.; Schwenk J. M.; Potz O.; Kramer S.; Joos T. O. *Proteomics*, 3:2155–2166, 2003.

[96] Sreekumar A.; Nyati M. K.; Varambally S.; Barrette T. R.; Ghosh D.; Lawrence T. S.; Chinnaiyan A. M. *Cancer Res.*, 61:7585–7593, 2001.

[97] Knezevic V.; Leethanakul C.; Bichsel V. E.; Worth J. M.; Prabhu V. V.; Gutkind J. S.; Liotta L. A.; Munson P. J.; Petricoin E. F.; Krizman D. B. *Proteomics*, 1:1271–1278, 2001.

[98] Hanash S. *Nature*, 422:226–232, 2003.

[99] Zhu H.; Bilgin M.; Bangham R.; Hall D.; Casamayor A.; Bertone P.; Lan N. et al. *Science*, 293:2102–2105, 2001.

[100] Wilson D. S. and Nock S. *Curr. Opin. Chem. Biol.*, 6:81–85, 2002.

[101] Zhu H. and Snyder M. *Cur. Opin. Chem. Biol.*, 5:40–45, 2002.

[102] Borrebaeck C. A. K. *Immunol. Today*, 21:379–381, 2000.

[103] Templin M. F.; Stoll D.; Schrenk M.; Traub P. C.; Vohringer C. F.; Joos T. O. *Trends. Biotechnol.*, 20:160–166, 2002.

[104] Paweletz C. P.; Charboneau L.; Bichsel V. E.; Simone N. L.; Chen T.; Gillespie J. W.; Emmert-Buck M. R.; Roth M. J.; Petricoin I. E.; Liotta L. A. *Oncogene*, 20:1981–1989, 2001.

[105] Petricoin E. F.; Zoon K. C.; Kohn E. C.; Barrett J. C.; Liotta L. A. *Nat. Rev. Drug Discov.*, 1:683–695, 2002.

[106] Yan F.; Sreekumar A.; Laxman B.; Chinnaiyan A. M.; Lubman D. M.; Barder T. J. *Proteomics*, 3:1228-1235, 2003.

[107] Lizardi P. et al. *Nat. Genet.*, 19:225–233, 1998.

[108] Gembitsky D. S.; Lawlor K.; Jacovina A.; Yaneva M.; Tempst P. *Mol. Cell Proteom.*, 3(11):1102–18, 2004.

[109] Kodadek T. *Chem. Biol.*, 8:105–115, 2001.

[110] Martin K.; Steinberg T. H.; Cooley L. A.; Gee K. R.; Beechem J. M.; Patton W. F. *Proteomics*, 3(7):1244–1255, 2003.

[111] Patwa T. H.; Zhao J.; Anderson M. A.; Simeone D. M. and Lubman D. M. *Anal. Chem.*, 78:6411–6421, 2006.

[112] Patwa T. H.; Zhao J.; Anderson M. A.; Simeone D. M. and Lubman D. M.

[113] Gygi S. P.; Rist B.; Gerber S. A.; Turecek F. ; Gelb M. H.; Aebersold R. *Nat. Biotechnol.*, 17:994–999, 1999.

[114] Tomlins S. A.; Mehra M.; Rhodes D. R.; Cao X.; Wang L.; Dhanasekaran S. M.; Kalyana-Sundaram S.; We J. T.; Rubin M. A.; Pienta K. J.; Shah R. B. and Chinnaiyan A. M. *Nat. Genet.*, 39(1):41–51, 2007.

[115] Subramanian, A. et al. *Proc. Natl. Acad. Sci. USA*, 102:1554515550, 2005.

CHAPTER II

Differential Phosphoprotein Mapping in Cancer Cells Using Protein Microarrays Produced from 2-D Liquid Fractionation

2.1 Introduction

Phosphorylation is one of the most common posttranslational modifications found for proteins. Phosphorylation and dephosphorylation of proteins is intimately connected to the signaling pathways in the cell. Initial changes in phosphorylation of a receptor usually result in large numbers of changes in protein signaling pathways downstream typically associated with major changes in cellular function [1–5]. As such, alterations in phosphorylation are highly correlated to new pathways that lead to oncogenesis [6]. It becomes essential then to be able to monitor changes in phosphorylation patterns on a global scale in order to identify the critical proteins involved in cell-cycle regulation related to cancer onset and progression.

A number of techniques have been used to detect phosphoprotein expression in cells on a global scale [7–9]. In one approach, cells were incubated with radioactive ^{32}P and then detected following 2-D gel electrophoresis [9]. This method however, requires the handling of radiolabels and the identification of phosphoproteins with slow turnover rates, which only incorporate small amounts of radioactive phosphate lead-

ing to poor detection. Monoclonal and polyclonal antibodies have also been used to detect phosphorylated proteins blotted onto membranes. In particular, changes in signal transduction pathways stimulated using platelet-derived growth factor were studied using anti-phosphotyrosine antibodies [10–12]. Changes in tyrosine phosphorylation could be monitored as a function of time, and large numbers of proteins involved in different signaling processes were observed. This method has been proved to be very sensitive with only a few femtomoles of the target required for detection. However, antibodies for detection of phosphorylated threonine and serine are still unreliable, and phosphorylated antibodies may not detect certain phosphorylated proteins due to steric hindrance [13]. Analytical mass spectrometry based methods, more specifically shotgun proteomics [14, 15], have been developed for monitoring phosphorylation as well. Protein digestion followed by MS/MS analysis of the resulting peptides can identify proteins in complex mixtures after comprehensive database searching [16–20].

In more recent work, ultrasensitive detection of small amounts of phosphorylated proteins has been achieved using a small molecule phosphosensor dye technology [21,22]. This phospho-sensitive dye was capable of quantitatively detecting phosphotyrosine, phosphoserine and phosphothreonine on a global scale. It has been used directly on 2-D gels and also in a microarray format on a variety of surfaces for monitoring substrates of kinase reactions. This has been shown to be a universal method for detection of phosphorylation, which could further discriminate against thio-phosphorylation and sulfation.

It is clear, however, that any global screening of cellular protein expression must em-

ploy methods that can readily separate large numbers of proteins and be amenable to the various techniques possible for phosphoprotein detection. 2-D gel electrophoresis has generally been the technique of choice, but the disadvantages of 2D gel technology are well known [23]. New methodologies for comprehensive protein expression will need to be explored. More recently, we have evaluated microarray formats as a high throughput screening method for studying global protein expression [24, 25]. This format could potentially provide a convenient platform for monitoring not only changes in protein expression but also the effects on protein modifications as a function of time and specific kinase activity.

In the present work, an all-liquid 2-D separation method has been explored to map the protein expression of a cell lysate for differential protein expression to study changes in phosphorylation patterns. This method uses chromatofocusing to fractionate proteins in a first dimension based on their pI, followed by nonporous silica RP-HPLC separation of the pI fractions to further fractionate proteins based on their hydrophobicity [26]. The method provides a means of separating large numbers of proteins in the liquid phase, as expressed in the cells, for deposition on a microarray surface [27]. The resulting protein microarray could be used to study global protein expression using fluorescent phospho-sensor dyes or phospho-specific antibodies. Specifically, the method has been used for differential protein expression to study changes in phosphorylation patterns in the human breast cancer cell lines SUM-52PE before and after inhibition of the fibroblast growth factor receptor 2 (FGFR2) protein [28]. The method provides a new and convenient means for protein identification and phosphorylation site searching by mass spectrometry where each microarray spot can be matched to the original vial (fraction) containing the purified

protein in the liquid phase.

2.2 Experimental

2.2.1 Chemicals

Methanol, ACN, urea, thiourea, iminodiacetic acid, DTT, OG, glycerol, bis-tris, TFA, PMSF, and β-mercaptoethanol were obtained from Sigma (St. Louis, MO). Water was purified using a Milli-Q water filtration system (Millipore, Inc., Bedford, MA) and all solvents used were HPLC grade unless otherwise specified. The reagents used were in the most pure form commercially available. Polybuffer 74 and Polybuffer 96 were purchased from Amersham Pharmacia Biotech (Piscataway, NJ). Pro-Q Diamond phosphoprotein gel stain and Pro-Q Diamond phosphoprotein gel destaining solution were obtained from Molecular Probes (Eugene, OR). BlockIt 1X blocking buffer and ArrayIt 2X printing buffer were obtained from Telechem International, Inc. (Sunnyvale, CA). 1X PBS and ultrapure DNase/RNase free distilled water were obtained from Invitrogen (Carlsbad, CA). Anti-phosphotyrosine antibody 4G10 clone was obtained from Upstate (Charlottesville, VA), Cy5-conjugated secondary antibodies were obtained from Jackson ImmunoResearch Lab (West Grove, PA).

2.2.2 Sample Preparation

Cell Culture

SUM-52PE is a human breast cancer cell line isolated from a patient's pleural effusion and developed in the Ethier laboratory [28]. The SUM-52 cells were cultured in Ham's F12 medium under serum-free conditions. The medium was supplemented

with 0.1% BSA, 0.5 μg/mL fungizone, 5 μg/mL gentamicin, 5 mM ethanolamine, 10 mM HEPES, 5 μg/mL transferrin, 10 μM T3, 50 μM selenium, 1 μg/mL hydrocortisone and 5 μg/mL insulin. All cell culture reagents were obtained from Sigma Chemical Co. The SUM-52PE cells were exposed to 1 μM PD173074 for 24 hr, and untreated cells received DMSO as a vehicle control.

SUM-52PE Cell Lysis

Cell pellets were reconstituted in lysis buffer consisting of 7 M urea, 2 M thiourea, 100 mM DTT, 0.5% biolyte ampholyte 3-10, 2 % OG, and 1 mM PMSF. The cell pellets were lysed at room temperature for 0.5 hr, followed by centrifugation at 35,000 rpm at 4°C for 1 hr. The supernatant was stored at -80°C for future use.

Sample Preparation for Chromatofocusing

A PD10 column (Amersham Biosciences) was equilibrated with a buffer solution containing 25 mM bis-tris in 6 M urea and 0.2% OG and then used to exchange the cell lysate from the lysis buffer to the above buffer according to the manufacturer's protocol.

Chromatofocusing of the SUM-52PE Cell Lysate

A schematic of the experimental apparatus used in this work is shown in Figure 2.1. The liquid separations were performed on the ProteomeLab PF-2D liquid fractionation system (Beckman-Coulter). A 5 mg sample of cell lysate was loaded on to the first-dimension column. The start buffer consisted of 6 M urea, 0.2% OG and 25 mM bis-tris, adjusted to pH 8.5 using IDA. The elution buffer consisted of 6 M urea,

0.2% OG and a 10-fold dilution of Polybuffer 96 and Polybuffer 74 in a ratio of 3:7, the pH adjusted to 4.0 using IDA. A PS-HPCF 1D column (Beckman-Coulter) was equilibrated with the start buffer until the pH of the effluent was the same as that of the start buffer. Sample was applied to the column with multiple injections. Once a stable baseline was achieved, the elution buffer was switched on to elute the proteins on the column in an isocratic mode. UV detection was performed at 280 nm, and the pH of the effluent was monitored using a flow-through on-line pH probe. The pH fractions were collected in 0.3 pH intervals and 15 fractions in all were taken over the range of pH 8.5-4.0. The CF separation was completed when the pH of the effluent reached 4.0. A 1 M NaCl solution followed by 100% IPA were then used to elute the strongly binding proteins as salt wash and IPA wash fractions respectively.

2.2.3 Reversed-Phase HPLC on pI Fractions

RP-HPLC separation was performed using PS-HPRP 2D (4.6×33 mm) columns (Beckman-Coulter). Solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in acetonitrile. The gradient was run from 5 to 15% B in 1 min, 15% to 25% B in 2 min, 25% to 31% B in 2 min, 31 to 41% B in 10 min, 41 to 47% B in 6 min, 47 to 67% B in 4 min, finally up to 100% B in 3 min, held for another 1 min, and then back to 5% B in 1 min at a flow rate of 1 mL/min. The column temperature was 40°C higher than the ambient temperature. The UV absorption profile was monitored at 214 nm. RP fractions were taken using a FC204 fraction collector in 96-well plates. Using a SpeedVac at 75°C, the fractionated proteins were dried down to 20 μL volume and transferred to a 384-well plate after which they were dried down completely. More than 2000 fractions were obtained after the 2-D separation and around half of these fractions from each cell line were used for spotting on the array. The dried protein

fractions (plates) were stored at -80°C until further use.

2.2.4 Protein Microarrays

A 3 μL sample of a 1:1 mixture of PBS and printing buffer was added to each well using a multipipet. Printing was done on super-epoxy slides (TeleChem International) using a Magna Spotter microarray printer (Bioautomation) and SMP4 microarray spotting pins (TeleChem Int). Using these pins, the uptake volume was 0.25 μL and the delivery volume was 1.1 nL, resulting in spot diameters of 135 μm. A minimum spot spacing of 160 μm can be achieved and 2300 spots per 1×3 in. slide can be printed. After spotting, the slides were stained with Pro-Q Diamond phosphoprotein gel dye (Molecular Probes) for 45 min. Destaining was performed three times for 10 min each using destaining solution from Molecular Probes. After destaining, the slides were washed with DNase- and RNase-free water for 10 min and then left to dry. For the antibody detection, the slides were washed 5 times for 5 min each in 1X PBS and incubated with 1:750 anti-phosphotyrosine antibody for 3 hr. The slides were then washed three times with 1X PBS-T and twice with 1X PBS for 5 min each. After washing, the slides were incubated with 1:1000 Cy5-conjugated secondary antibody for 1 hr following which they were washed three times in 1X PBS-T and 1X PBS for 5 min each. The slides were then rinsed with 1X PBS and dried by centrifuging for 1 min on a microarray high-speed centrifuge (Telechem Int.). All steps following the staining with Pro-Q Diamond dye were performed in the dark under aluminum foil wraps. Both antibody solutions above were prepared in 1:1 BlockIt buffer and 1XPBS. Hybridization chambers were used for antibody incubation, and a minirotator (Geneq Inc., Montreal, Canada) was used for all the washing and incubation steps. Scanning was done using an Axon 4000A scanner, and GenePix Pro

3.0 software was used for data acquisition and analysis.

2.2.5 Protein Digestion

The targeted UV peak in the second dimensional RP-HPLC chromatogram, which showed a positive response to the phospho dye, was collected and dried down to eliminate ACN and TFA. 1 M NH_4HCO_3 and 10 mM DTT were then added to a final concentration of 100 mM and 1 mM respectively, and incubated at 60°C for 15 min. Trypsin was then mixed with the denatured proteins at the ratio of 1:50. The mixture was incubated at 37°C for 24 hr.

2.2.6 Enzymatic Dephosphorylation

After completion of the proteolytic cleavage, the samples were divided into two equal parts. The enzymatic dephosphorylation step [29] was performed by treating one part with 5 units calf alkaline phosphatase reconstituted in 25 mM NH_4HCO_3 buffer (pH 8.0). The mixture was incubated at 37°C for 2 hr, and 2.5% TFA was added to stop the enzymatic reaction. The other part was treated as a control.

2.2.7 Matrix Preparation and Spotting

In preparation for MALDI-MS, the samples were first aspirated using Zip Tips, and then 1 μL of the eluent was mixed with an equal volume of CHCA matrix solution prepared in 60% ACN/0.1% TFA and spotted on a MALDI plate. Once the spot dried, 1 μL of 9:1 THAP/DAC matrix solution [30] prepared in 60% ACN/0.1% TFA was applied on top. The spot was allowed to dry slowly afterward.

2.2.8 Protein Identification by MALDI-MS

MALDI-TOF MS (Micromass Inc. TOFSpec2E) was used to generate peptide mass fingerprints and then searched for registered peptide masses of proteins in the existing SwissProt database. The peptide map database search was also used to initially confirm the possible presence of a phosphorylation site. To verify and locate the phosphorylation sites on the proteins, MALDI-MS spectra of the phosphorylated (control) and the dephosphorylated samples were compared.

2.2.9 MW Determination by ESI TOF-MS

An ESI TOF-MS (Micromass Inc. LCT) was used for determination of intact protein molecular weights. The intact molecular weights from the LCT and the PMF obtained from the MALDI-TOF MS analysis provided the complete identification of the proteins in the fractions of interest. Fractions from the second-dimension RP-HPLC analysis for LCT were first dried down using a SpeedVap and then reconstituted in 60% ACN with 2% FA. The samples were directly infused at 10 μL/min using a syringe pump. A desolvation temperature of 150°C and source temperature of 100°C was used. Nitrogen gas flow was maintained at 400 L/hr. The capillary voltage was set at 3200 V, the sample cone voltage at 35 V, the extraction cone voltage at 3 V, and the reflection lens voltage at 750 V. One mass spectrum was acquired every second. The intact molecular weight was obtained by deconvolution of the spectra using the MaxEnt1 software (Micromass Inc.).

2.2.10 LC-MS/MS

The tryspin-digested samples were analyzed by reversed-phase chromatography using a 0.075×150 mm C18 column attached to a Paradigm HPLC pump (Michrome Bio Resources). Peptides were eluted using a 23 min gradient from 5 to 95% B (0.1% FA/95% ACN), where solvent A was 0.1% FA/2% ACN. A Finnigan LTQ mass spectrometer (Thermo Electron Corp.) was used to acquire the spectra, the instrument operating in data-dependent mode with dynamic exclusion enabled. The MS/MS spectra on three most abundant peptide ions in full MS scan were obtained. All MS/MS spectra were analyzed using the MASCOT search tool against the composite, nonidentical protein sequence database MSDB.

2.3 Results and Discussion

SUM-52PE cells highly overexpress FGFR2 at both the message and protein levels. There are nine alternatively spliced isoforms of FGFR2 expressed by the SUM-52PE cells [28]. The isoforms differ in the number of immunoglobulin-like domains, the presence or absence of the acid box, and the carboxyl terminal region. The SUM-52PE cells display the transformed phenotypes of growth factor-independent growth and the ability to grow under anchorage independent conditions and invasion. PD173074 is a small molecule kinase inhibitor against the FGFR family [31]. PD173074 blocks the phosphorylation of FGFR2, as well as the downstream signaling components of the MAP kinase and PI3 kinase pathways [32]. The PD compound also inhibits SUM-52PE cell growth in monolayer and in soft agar [28].

The cell lysates of SUM-52PE and SUM-52PE inhibited by PD173074 were separated

using the 2-D liquid separation method, and protein maps were obtained using the Beckman Coulter ProteoVue Software suite for each of the cell lines. A comparison of the two cell lines before and after inhibition is shown in Figure 2.2 in differential display format (using DeltaVue from Beckman Coulter) for two different pH regions. In Figure 2.2, the protein profile for SUM-52PE is displayed in green and that for the inhibited cell line is shown in red. The differential display in the center lane shows that there are proteins that are up- and down-regulated following inhibition of the SUM-52PE cell line. This might be expected since inhibition of the FGFR2 growth factor results in changes in protein pathways that would change the protein expression in the cell. Nevertheless, most of the proteins observed are similar in the SUM-52PE before and after inhibition.

Each of the protein bands were collected in the liquid phase following 2-D liquid fractionation and spotted on the protein microarray as described above. Each array spots can be associated with a protein band collected during the 2-D liquid separation. The array is then stained with the Pro-Q Diamond phospho-dye to screen for the presence of phosphorylation on the different protein spots. The result is shown in Figure 2.3 for an array cluster with three pH fractions where several spots are clearly lit up by the dye when scanned by the 532 nm excitation source indicating the presence of phosphorylation. The microarray image of Figure 2.3 reveals the limited quality of the spot printing using the contact printer on glass slides. The method, however, does allow one to identify phosphorylated proteins on a global scale using only a limited amount of material.

The result of a differential phosphoprotein array for proteins printed from a single

pH range is shown in Figure 2.4. Each pair of arrays compares the SUM-52PE cell line before and after inhibition of FGFR2. The arrows that point to pairs of protein spots clearly identify proteins that are phosphorylated in the SUM-52PE cells under control conditions but not so following treatment with the inhibitor in response to changes in phosphorylation pathways due to inhibition of the FGFR2. Of the nearly 1000 protein bands printed on the array for each cell line, there were at least 50 proteins showing changes in the state of phosphorylation due to inhibition. In many cases, the protein spot is lit up on one array but not the other, indicating that the protein is phosphorylated in one cell line but not the inhibited counterpart. In other cases, the protein spot is lit up, but the quantitative degree of excitation may change, indicating a different degree of phosphorylation between the two cell lines. There are also some spots that remain unchanged, indicating that these proteins are not involved in the FGFR2 signaling pathway.

It should be noted in Figure 2.4 that the corresponding spots in the arrays for the two cell lines may not contain the same protein. The spotting process is performed according to the 2-D liquid fractionation of protein bands. There are several bands that appear in one cell line but not the other so that the number of protein peaks in each pH fraction is different. The corresponding peaks in the arrays for the two cell lines can be matched using the %B on the chromatographic gradient and then by using MALDI-TOF MS of the protein digests to verify that they are the same proteins. The use of MALDI-TOF MS for definitively matching the protein spots is essential since phosphorylated proteins often show pH shifts, which can significantly shift the position of the spot on the array. These shifts would likewise be observed on 2-D gels.

The use of protein arrays with the Pro-Q dye, which is selective toward phosphorylation, allows one to rapidly detect the presence of phosphorylation in specific proteins. This eliminates the need to perform detailed analysis on a large number of proteins, thus simplifying the problem of studying differential phosphorylation in biological systems. It is essential, though, to perform detailed mass spectrometric analysis of the proteins selected as candidates to establish the identity of the protein and to confirm changes in phosphorylation as outlined in Figure 2.5-2.8. An important aspect of this work is that each spot on the array can be correlated to the original protein well from which it was spotted for further analysis. MALDI-TOF MS on the tryptic digest of proteins was initially performed for identification and confirmed by LC-MS/MS.

When using the matrix CHCA, the negative charge on the phospho groups make the phosphopeptides difficult to detect in positive ion mode. THAP, a less acidic matrix, has been demonstrated to enhance the ionization of phosphopeptides by 10-fold [30]. We experimented with a technique described above using both THAP and CHCA where improved sensitivity for phosphopeptides in the positive ion mode was achieved without affecting the ionization of the non-phosphorylated peptides. This matrix mixture though, required a higher laser power than when using only CHCA to give comparable signal intensities for non-phosphorylated peptides.

A key issue in this work involves using mass spectrometric methods to confirm the presence of phosphorylations in the array spots that light up when stained with Pro-Q Diamond dye. This was performed using CAP to dephosphorylate the proteins that were identified as being phosphorylated on the arrays and performing MALDI-TOF

MS on the digests before and after dephosphorylation. The mass spectra of the peptides should show an 80 Da shift to lower mass after dephosphorylation if they were originally phosphorylated. The MALDI-TOF MS spectra for several phosphorylated proteins and their dephosphorylated counterpart are shown in Figure 2.5-2.8. Figure 2.5 shows the phosphorylated and dephosphorylated counterpart of a peptide from Zinc Finger Protein 492, clearly indicating a shift of the peak at m/z 2333 corresponding to (K)LYKPESCNNACDNIAKISK(Y) to m/z 2253 following dephosphorylation by CAP. Rab13 (Figure 2.6) shows a shift from m/z 1934.96 to 1855, which corresponds to the peptide (K)-GSKPVRPPAPGHGFPLIK(R). Figure 2.7 shows the peptide (-)-MMLGTEGGEGFVVK(V) at m/z 1534.67 from heterogeneous nuclear ribonucleoprotein H shifted to m/z 1454, and Figure 2.8 shows the peptide at m/z 2342.04 with sequence (R)FHTGKTSFACTECGKFSLR(K) from zinc finger protein 615 shifted to m/z 2262.24 following dephosphorylation. In all these cases, the peaks corresponding to the phosphorylated peptide are absent from the dephosphorylated sample spectra, indicating that the enzymatic reaction is complete. This method clearly shows that these proteins which were illuminated by the Pro-Q dye on the microarray are indeed phosphorylated, although the position and type of phosphorylation need to be confirmed by further experimentation.

The MW of the intact protein was also obtained using ESI-TOF MS when there was a sufficient amount of protein available in order to constrain the peptide map and LC-MS/MS search and unambiguously identify the protein. A partial list of differentially expressed proteins that lit up on the array is shown in Table 2.1 as identified by MALDI-TOF MS and LC-MS/MS. The last column in Table 2.1 indicates whether the phosphoprotein is upregulated or downregulated in the SUM-52PE

cell line compared to the inhibited sample. In each case in Table 2.1 performed by LC-MS/MS, the initial database search showed the probable presence of one or more phosphorylation sites, although the specific phosphorylation generally could not be identified. In addition, the experimental MW often did not precisely match the database value, indicating the presence of a modified protein. Though there may be several different modifications on any protein in addition to phosphorylation, there are significant shifts in the measured pI toward lower pH compared to the database values in all the proteins in Table 2.1, which is often indicative of the presence of phosphorylations [33]. Zinc finger protein 492 was isolated in the salt wash fraction, pH <4.0, although the theoretical pI of the unphosphorylated form is 9.3. This shift in pI may be due to the presence of up to 14 phosphorylations based on the MS data.

To further confirm the type of phosphorylation site modified, anti-phosphotyrosine antibodies were used. Figure 2.9 shows two arrays that had been processed with Pro-Q Diamond Dye. In Figure 2.9a, the green spots obtained in this process display all the proteins that have phosphorylated Ser, Thr, or Tyr residues as detected by the Pro-Q dye. In Figure 2.9b, the array was also processed with 4G10 anti-phosphotyrosine antibody after the Pro-Q analysis, and the red spots correspond to Tyr phosphorylations detected by the antibody. The green spots in this image identify proteins that are not phosphorylated at Tyr, and the yellow spots identify those that have only a small number of phosphorylated Tyr. The data clearly show that the spots corresponding to Eps15 and SHPS-1 are phosphorylated on Tyr, although there may also be a small number of Thr or Ser phosphorylations present. In principle, an anti-phosphoserine or anti-phosphothreonine antibody may also be used after the Pro-Q analysis.

It should be noted that in most cases shown in Figure 2.9 the spots lit up by the Pro-Q dye are in concordance with those lit up by the Cy-5-labeled anti-phosphotyrosine antibody. However, the spot marked by 'X' on the array is not detected by the Pro-Q dye but is detected by the anti-phosphotyrosine antibody as shown by the bright red color. The response to the antibody could be due to a possible nonspecific binding of the antibody. Alternatively, the lack of response to the Pro-Q dye may be due to the protein concentration in this spot, which is too low for detection by the dye. This spot has presently not been positively identified by MALDI-MS, and evidence of a phosphorylation site by mass spectrometric analysis has not yet been found.

A quantitative analysis was also conducted with the Pro-Q Diamond dye. The sensitivity of the dye for epoxy-coated surfaces turned out to be $\sim$100 pg of total protein/well, which is equivalent to $\sim$100 fg of protein/spot where β-casein was used as a quantitation standard (Figure 2.10). β-Casein was dephosphorylated and used as a quantitation control. The dye appears to have an improved sensitivity when used with hydrogel slides as shown previously by Patton et al. [21], where they obtained a sensitivity of $\sim$50 fg using the same standard. Epoxy-coated slides were selected for analysis since they resulted in lower background absorbance compared to the amine substrate when working with the Pro-Q dye. This quantitative data should not be treated as absolute since contact printing with its inherent drawbacks is less reliable for quantitative analyses than noncontact printers. Nevertheless, the method can be used as a reference for comparisons within the slide. Figure 2.11 shows the linear dynamic range of the Pro-Q dye, which turned out to agree quite well with the values obtained by Patton et al [21]. Since with contact printing the

spot size is not consistent, a more convenient measure of 'total protein per well' was used to obtain the quantitative data representing a more practical approach.

2.4 Conclusion

The use of 2-D liquid separations can generate protein microarrays that reflect the natural posttranslational modifications as produced in cells. Of critical importance is the detection of changes in phosphorylations, since these PTMs are often responsible for signaling pathways related to essential processes in cells related to cancer. In this work, we have shown that these microarrays can be used to detect changes in phosphorylation in a malignant breast cancer cell line due to inhibition of the FGFR2 receptor. Pro-Q Diamond dye was used as a global means to detect phosphorylations while an anti-phosphotyrosine antibody was used to detect proteins with tyrosine phosphorylations. These arrays can be clearly used to detect the presence of phosphorylated proteins, although the specific phosphorylation sites require further work using LC-MS/MS. When using CAP - which was applied to proteins detected as phosphorylated on the arrays - a shift of -80 Da detected in the mass spectrum resulted in easier identification of proteins by MALDI-TOF MS. Although changes in phosphorylation patterns could be detected due to inhibition of the FGFR2 receptor by a small-molecule inhibitor, this detection was only performed 24 hr after initial stimulation. To obtain meaningful biological data on this system, future work will require a time course study to monitor changes in phosphorylation at various times immediately after inhibition.

Table 2.1: List of differentially phosphorylated proteins identified by MALDI-MS and LC-MS/MS

Protein Name	**Acc No.**	**Protein MW/pI**	**Obs MW**	**Obs pH Range**	**Exprn Level**
Zinc finger protein 492	Q9P255	65952/9.3	66010	salt wash	-
RalBP1-interacting protein 1	Q96D71	80770/5.6	80758	4.3-4.6	-
Eps15	P42566	98675/4.4		4.3-4.6	-
MAPK interacting protein	Q15750	54645/5.3	54436	4.3-4.6	+
Signal transducer and activator of transcription 3 (STAT3)	P40763	88069/5.9		4.3-4.6	+
Heterogeneous nuclear ribonucleoprotein H3	P31942	36927/6.4	36915	6.1-6.4	-
SHPS-1	P78324	54813/6.3	55163	6.1-6.4	-
Zinc finger protein 324 (zinc finger protein ZF5128)	O75467	61104/9.7	61074	6.1-6.4	-
60S ribosomal protein L13	P26373	24116/11.65	24285	6.1-6.4	-
Cofilin-1	P23528	18371/8.26	18425	6.1-6.4	-
Lamin A/C (70 kDa lamin)	P02545	74139/6.57	74304	5.2-4.9	+
Protein kinase C binding protein 1	Q9ULU4	131692/6.83		5.2-4.9	-
Peroxisome proliferator-activated receptor binding protein (PBP)	Q15648	168334/8.88		5.2-4.9	-
Splicing factor 1	Q15637	68287/9.07	68415	5.2-4.9	+
Proto oncogene C-crk	P46108	33850/5.49	33715	5.2-4.9	-
Octamer-binding transcription factor 1	P14859	76426/6.34	76785	5.2-4.9	-
Cytoplasmic protein NCK1	P16333	42838/6.06		5.2-4.9	-
Histamine H1 receptor	P35367	55748/9.33	55637	5.2-4.9	+
Neurofilament triplet M protein	P07197	102256/4.9		salt wash	-
Antigen KI-67	P46013	358526/9.46		salt wash	-
Neuroblast differentiation associated protein AHNAK	Q09666	312295/6.29		salt wash	+
Ribosomal protein S6 kinase alpha 5	O75582	89810/6.63		salt wash	+
Signal transduction protein CBL-C	Q9ULV8	52450/7.83	52522	salt wash	-

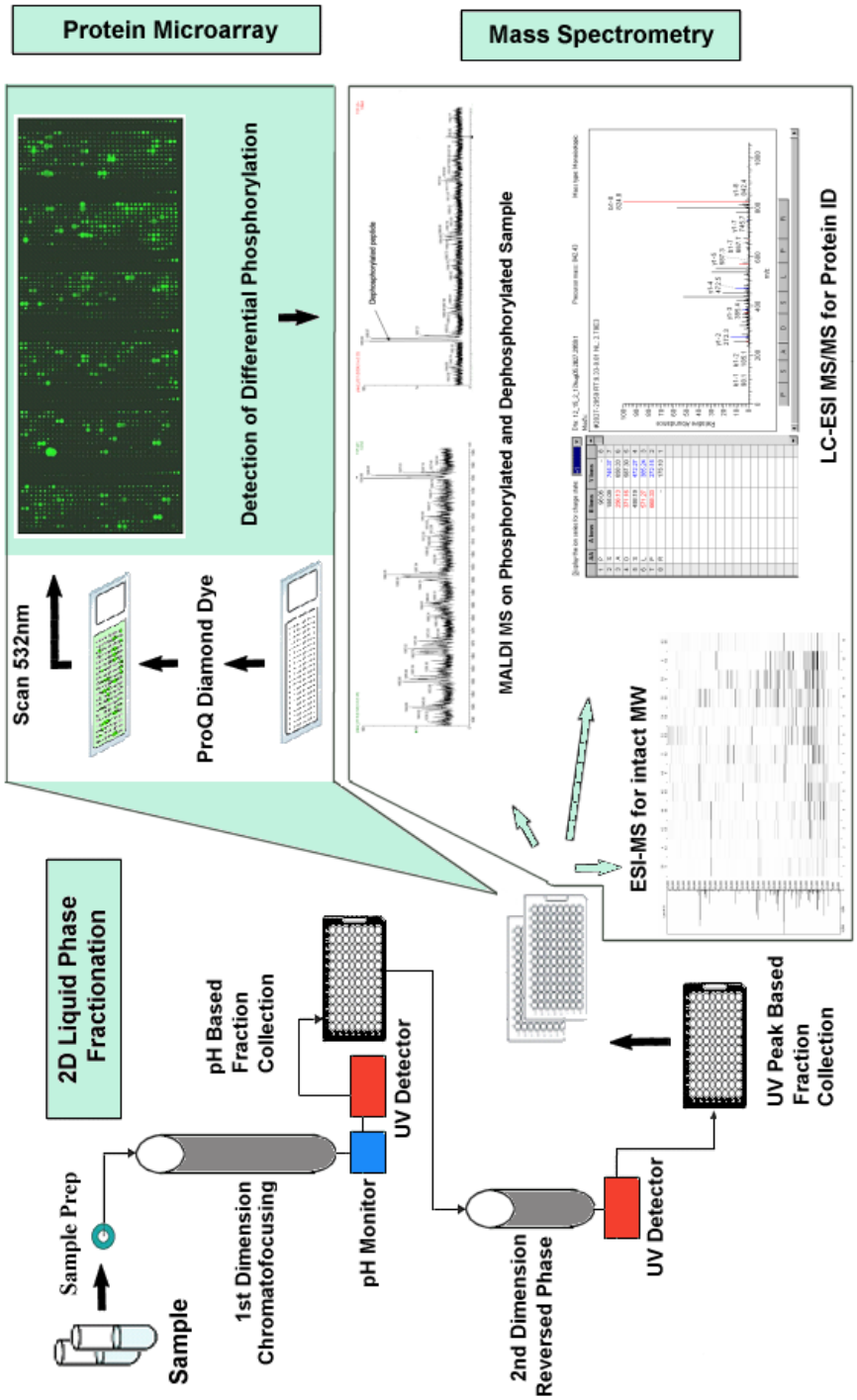

Figure 2.1: Overview of the approach used in this experiment.

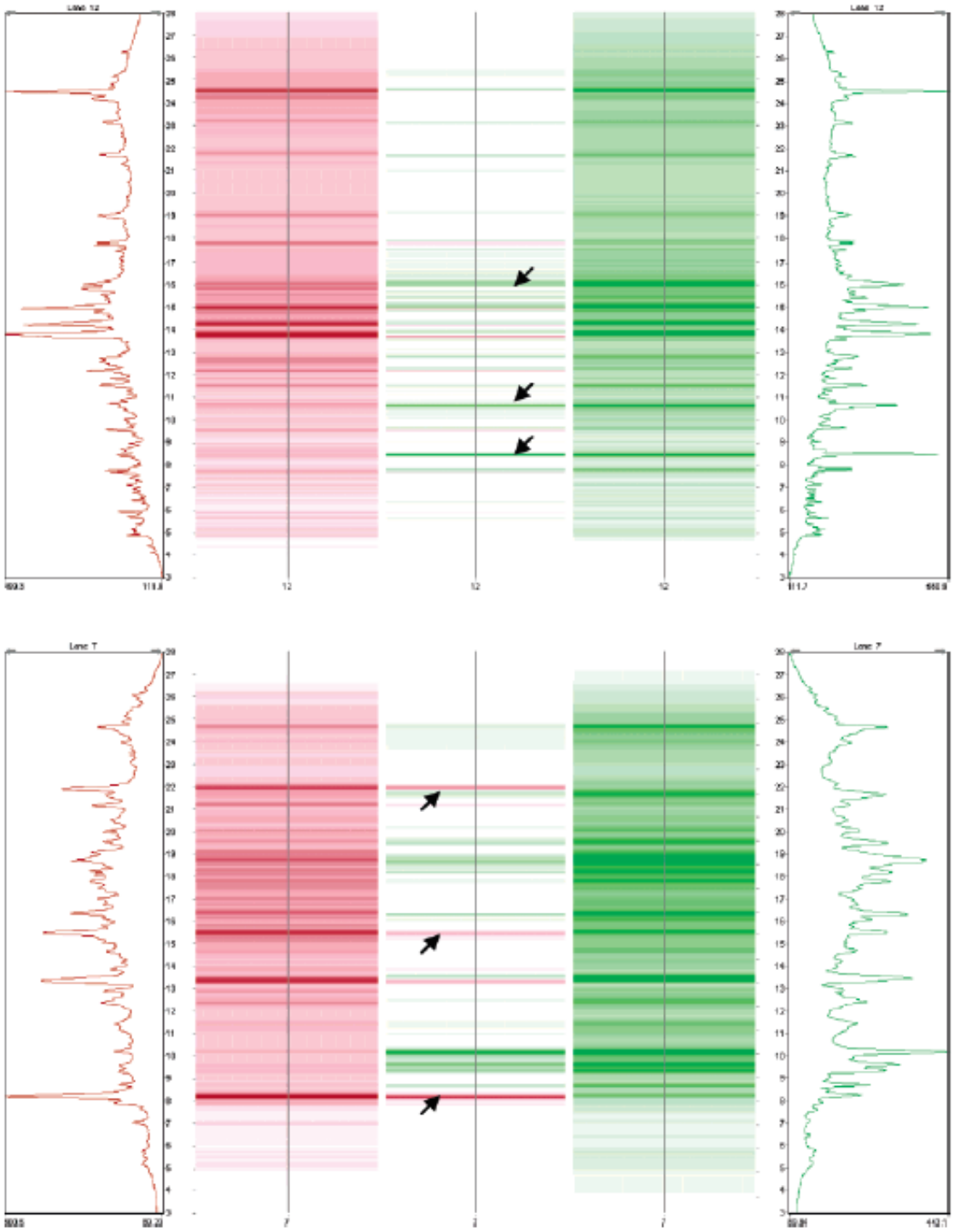

Figure 2.2: 2D-UV difference maps of FGFR2-inhibited SUM-52PE (left) and normal SUM-52PE (right) cell line for two different pH fractions: lane 12 - pH 7.6-7.9 (top); and lane 7 - pH 6.1-6.4 (bottom)

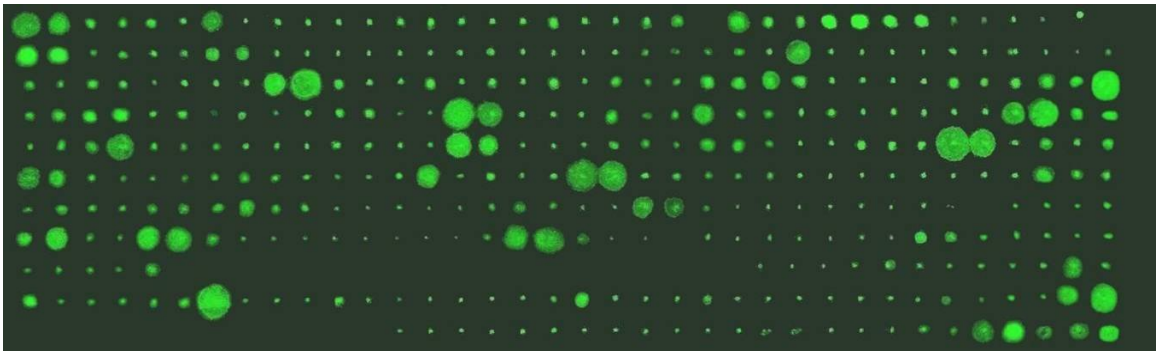

Figure 2.3: Microarray image showing fractions with pH 5.2-4.3 for SUM-52PE where phosphorylation is detected using Pro-Q Diamond dye

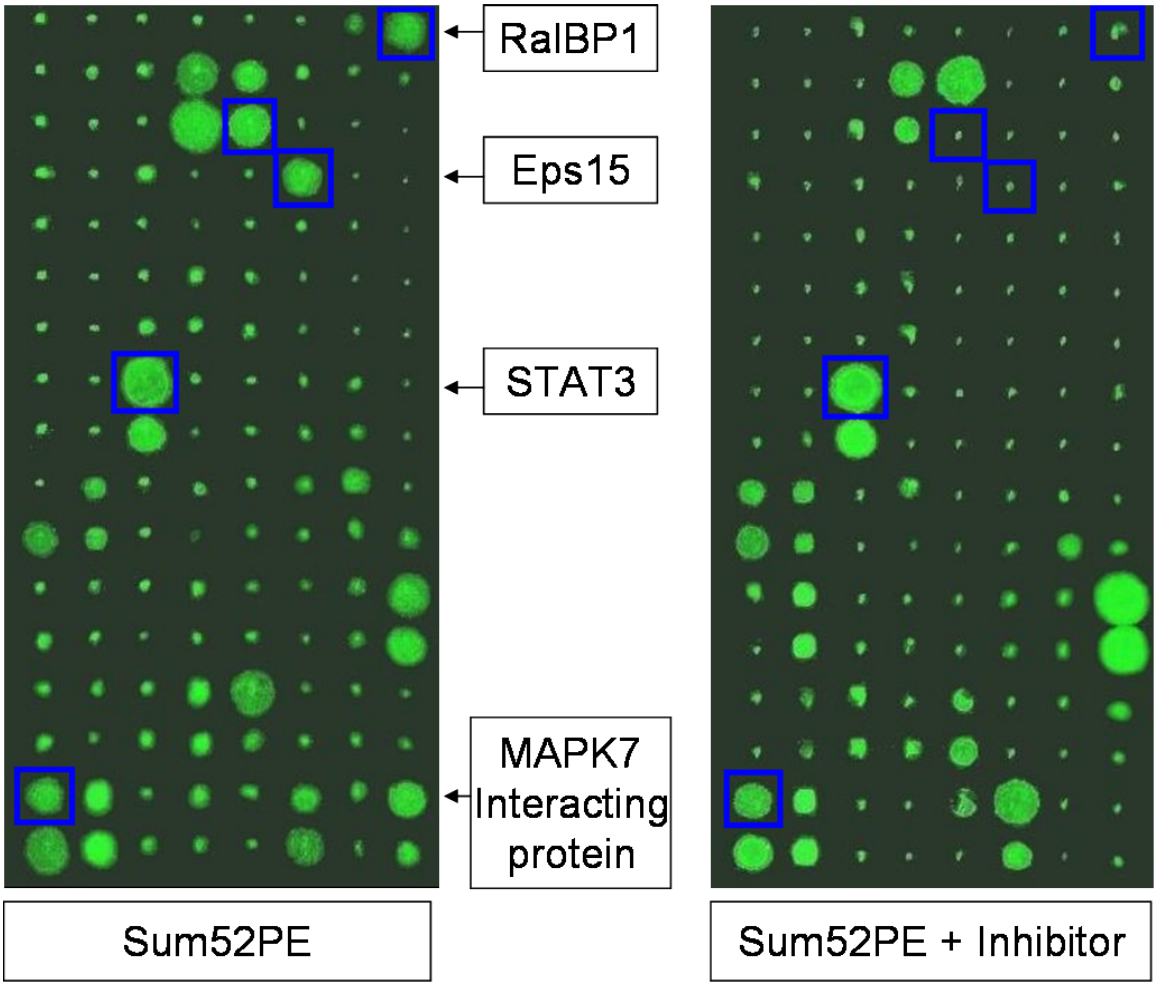

Figure 2.4: Microarray image showing pH fraction 4.6-4.3 for SUM-52PE before (left) and after (right) stimulation by PD173074

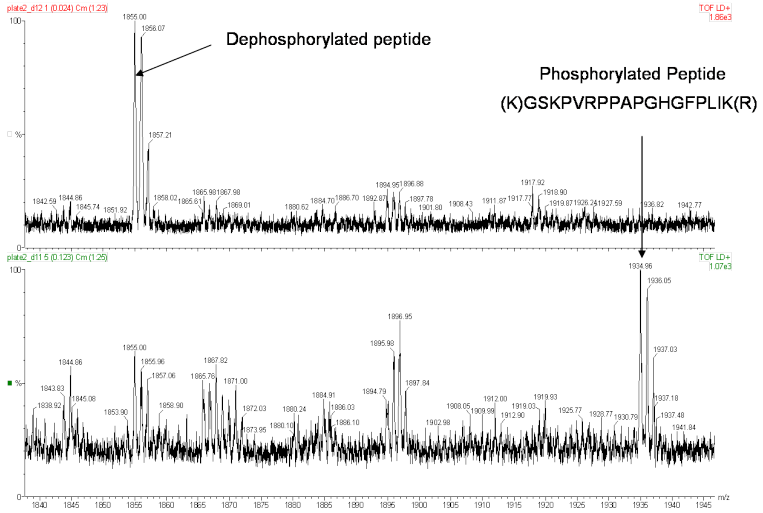

Figure 2.5: MALDI spectrum of zinc finger protein 492 obtained before (bottom) and after (top) dephosphorylation by CAP

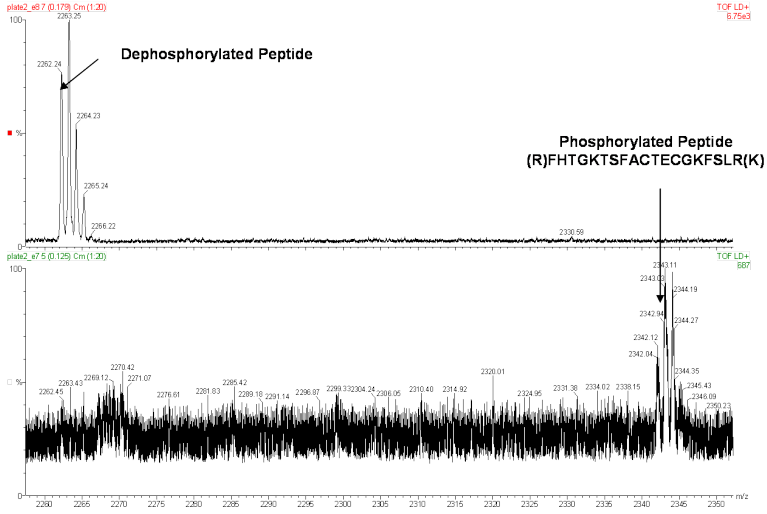

Figure 2.6: MALDI spectrum of Rab13 interacting protein (MIRab13) (MICAL-like protein 1) obtained before (bottom) and after (top) dephosphorylation by CAP

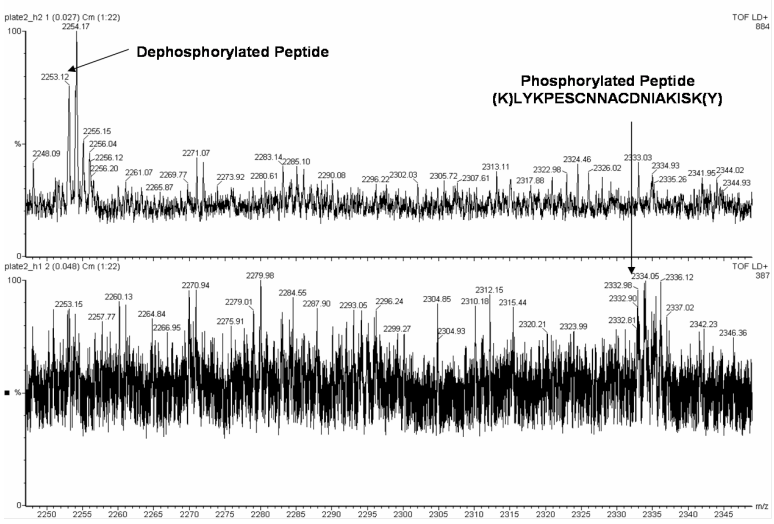

Figure 2.7: MALDI spectrum of heterogeneous nuclear ribonucleoprotein H (hnRNP H) obtained before and after dephosphorylation

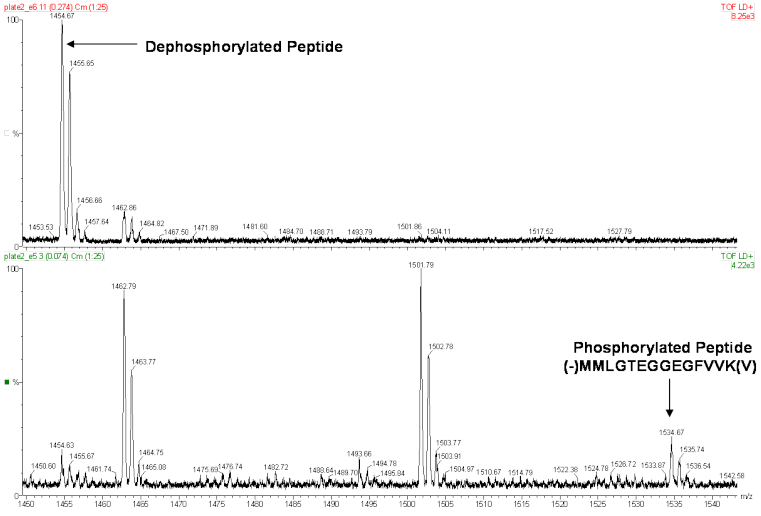

Figure 2.8: MALDI spectrum of zinc finger protein 615 obtained before and after dephosphorylation

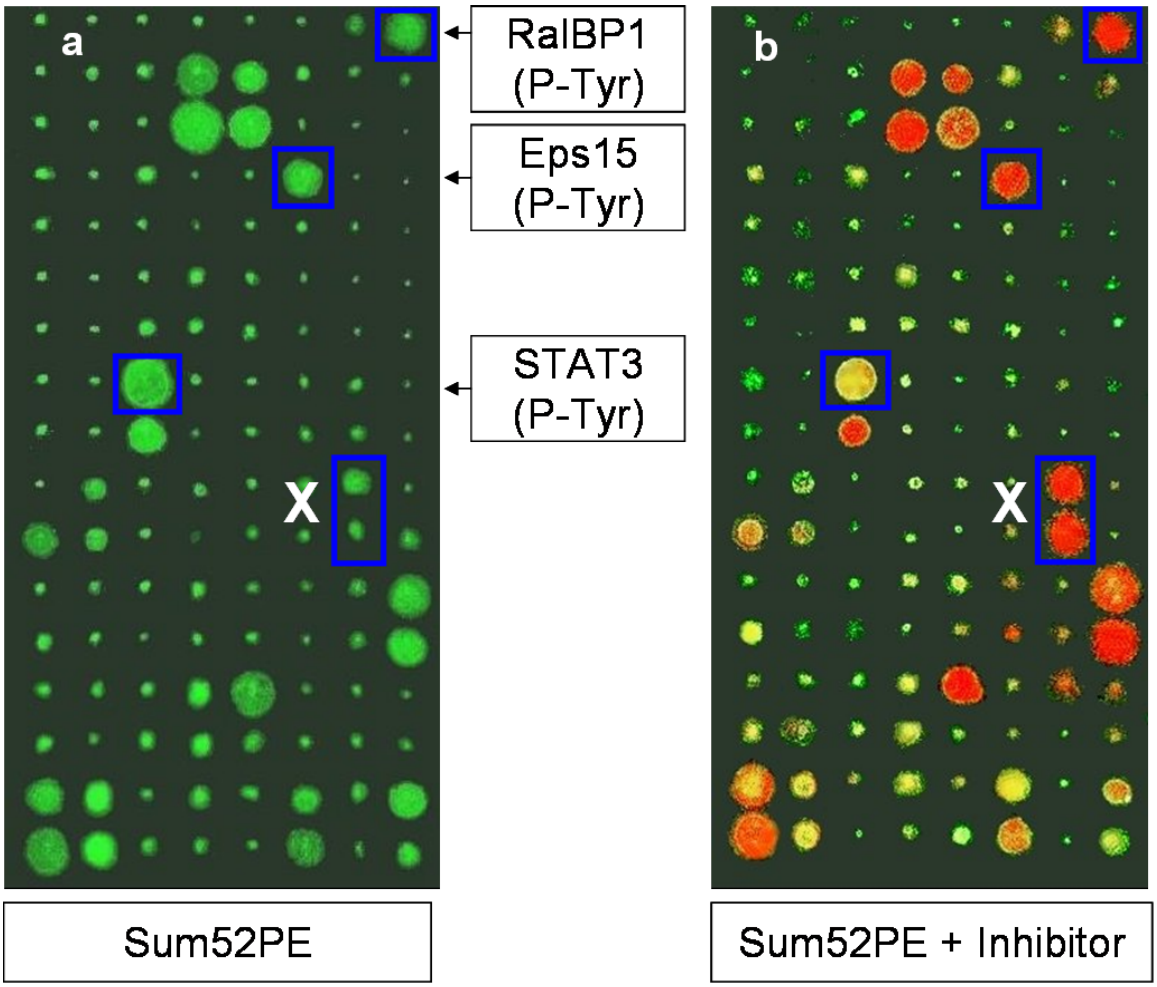

Figure 2.9: Slide image for pH fraction 4.6-4.3 processed with antiphosphotyrosine antibody (b) after having been visualized with Pro-Q Diamond dye (a). Eps15 and RalBP1 show Tyr phosphorylation and STAT3 shows phosphorylation on amino acids other than tyrosine. The spot marked X displays the case where the antibody binds to the protein either through nonspecific interactions or the concentration of the protein in that spot is below the sensitivity limit of Pro-Q Diamond dye. Spots reacting positively to the Pro-Q dye are shown in green while those that bind with the anti-phosphotyrosine antibody are shown in red.

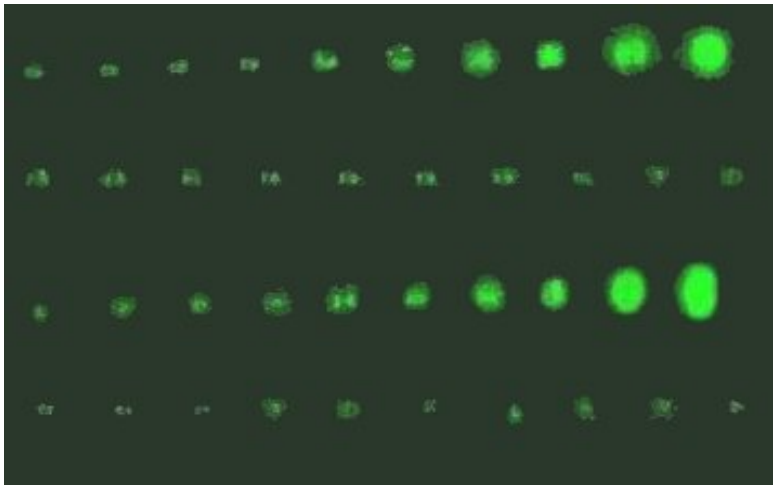

Figure 2.10: Detection sensitivity of β-casein visualized using Pro-Q Diamond dye. Figure shows 10 and 100-fold serial dilutions of β-casein from 100 μg to 100 fg per well spotted on superepoxy slides; The first and the third rows from top show the phosphorylated form and the other two rows show controls (dephosphorylated β-casein treated with calf alkaline phosphatase). A sensitivity limit of $\sim$100 pg of total protein/well was obtained.

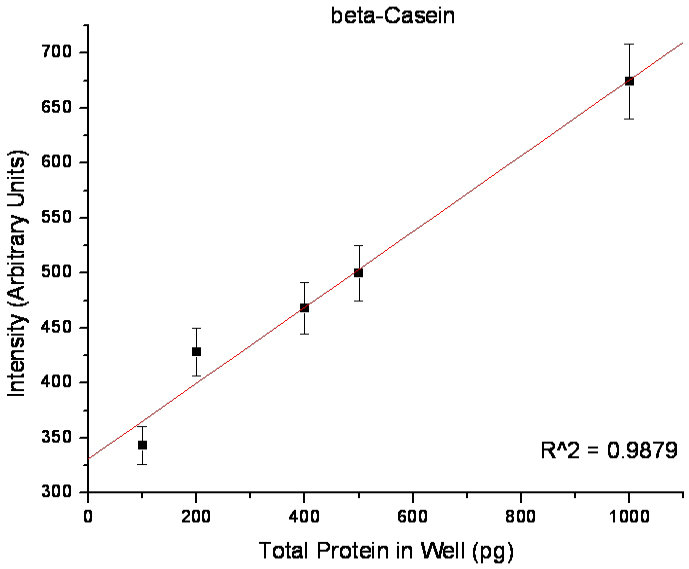

Figure 2.11: Linear dynamic range for β-casein visualized using Pro-Q Diamond phosphoprotein dye. A dynamic range of $\sim$100 was obtained when epoxy-coated glass slides were used for analysis.

References

[1] Hunter T. *Philos. Trans.: Biol. Sci.*, 353(1368):583–605, 1998.

[2] Cohen P. *Eur. J. Biochem.*, 268(19):5001–5010, 2001.

[3] Cohen P. *Nature*, 296:613–617, 1982.

[4] Robinson D. R.; Wu Y. M.; Lin S. F. *Oncogene*, 19:5548–5557, 2000.

[5] Cantin G. T.; Yates J. R. *J. Chromatogr. A*, 1053:7–14, 2004.

[6] Gschwind A.; Fischer O. M.; Ullrich A. *Nat. Rev.*, 4:361–370, 2004.

[7] Sickman A.; Meyer H. E. *Proteomics*, 1:200–206, 2001.

[8] Zeller M.; Konig S. *Anal. Bioanal. Chem.*, 378:898–909, 2004.

[9] Boyle W. J.; van der Greer P.; Hunter T. *Methods Enzymol.*, 201:110–149, 1991.

[10] Soskic V.; Gorlach M.; Poznanovic S.; Boehmer F. D.; Godovac-Zimmermann J. *Biochemistry*, 38:1757–1764, 1999.

[11] Hinsby A. M.; Olsen J. V.; Bennett K. L.; Mann M. *Mol. Cell. Proteomics*, 2:29–36, 2003.

[12] Kaufmann H.; Bailey J. E.; Fussenegger M. *Proteomics*, 1:194–199, 2001.

[13] Gronborg M.; Kristiansen T. Z.; Stensballet A.; Andersen J. S.; Ohara O.; Mann M.; Jensen O. N.; Pandey A. *Mol. Cell. Proteomics*, 1(7):517–527, 2002.

[14] Eng J. K.; McCormack A. L.; Yates J. R. J. *Am. Soc. Mass Spectrom.*, 5(11):976–989, 1994.

[15] Washburn M. P.; Wolters D.; Yates J. R. *Nat. Biotechnol.*, 19(3):242–247, 2001.

[16] Shu H.; Chen S.; Bi Q.; Mumby M.; Brekken D. L. *Mol. Cell. Proteomics*, 3(3):279–286, 2004.

[17] Beausoleil S. A.; Jedrychowski M.; Schwartz D.; Elias J. E.; Villen J.; Li J.; Cohn M. A.; Cantley L. C.; Gygi S. P. *Proc. Natl. Acad. Sci. U.S.A.*, 101(33):12130–12135, 2004.

[18] Zhou H.; Watts J.; Aebersold R. *Nat. Biotechnol.*, 19:375–378, 2001.

[19] Ficarro S. B.; McCleland M. L.; Stukenberg P. T.; Burke D. J.; Ross M. M.; Shabanowitz J.; Hunt D. F.; White F. M. *Nat. Biotechnol.*, 20:301–305, 2002.

[20] Salomon A. R.; Ficarro S. B.; Brill L. M.; Brinker A.; Phung Q. T.; Ericson C.; Sauer K.; Brock A.; Horn D. M.; Schultz P. G.; Peters E. C. *Proc. Natl. Acad. Sci. U.S.A.*, 100(2):443–448, 2003.

[21] Martin K.; Steinberg T. H.; Cooley L. A.; Gee K. R.; Beechem J. M.; Patton W. F. *Proteomics*, 3:1244–1255, 2003.

[22] Steinberg T. H.; Agnew B. J.; Gee K. R.; Leung W. Y.; Goodman T.; Schulenberg B.; Hendrickson J.; Beechem J. M.; Haugland R. P.; Patton W. F. *Proteomics*, 3:1128–1144, 2003.

[23] Nilsson C. L.; Davidsson P. *Mass Spectrom. Rev.*, 19:390–397, 2000.

[24] Zhu H.; Snyder M. *Curr. Opin. Chem. Biol.*, 7:55–63, 2003.

[25] Cutler P. *Proteomics*, 3:3–18, 2003.

[26] Yan F.; Subramanian B.; Nakeff A.; Barder T. J.; Parus S. J.; Lubman D. M. *Anal. Chem.*, 75(10):2299–2308, 2003.

[27] Yan F.; Sreekumar A.; Laxman B.; Chinnaiyan A.; Lubman D. M. *Proteomics*, 3:1210–1217, 2003.

[28] Tannheimer S. L.; Rehemtulla A.; Ethier S. P. *Breast Cancer Res.*, 2(4):311–320, 2000.

[29] Liao P.; Leykam J.; Andrews P. C.; Gage D. A.; Allison J. *Anal. Biochem.*, 219:9–20, 1994.

[30] Yang X.; Wu H.; Kobayashi T.; Solaro R. J.; van Breemen R. B. *Anal. Chem.*, 76:1532–1536, 2004.

[31] Mohammadi M.; Froum S.; Hamby J. M.; Schroeder M. C.; Panek R. L.; Lu G. H.; Eliseenkova A. V.; Green D.; Schlessinger J.; Hubbard S. R. *EMBO J.*, 17(20):5896–5904, 1998.

[32] Moffa A. B.; Tannheimer S. L.; Ethier S. P. *Mol. Cancer Res.*, 2(11):643–652, 2004.

[33] Zhu K.; Zhao J.; Lubman D. M.; Miller F. R.; Barder T. J. *Anal. Chem.*, 77:2745–2755, 2005.

CHAPTER III

Natural Protein Microarrays using Liquid Phase Fractionation of Panc-1 cell-lines for the study of Humoral Response in Pancreatic Cancer

3.1 Introduction

Major advances in cancer control will be greatly aided by early detection so as to diagnose and treat cancer in its pre-invasive stage prior to metastasis. Unfortunately, for Pancreatic ductal adenocarcinoma (PDAC), the fourth leading cause of cancer-related death in the United States [1], effective early detection and screening are currently not available and tumors are typically diagnosed at a late stage, frequently after metastasis. PDAC is generally considered to be incurable by available treatment modalities, with a 5-year survival rate of less than 4 percent. Existing biomarkers for PDAC are inadequate [2]. CA19-9 has been tested for its utility as an early detection marker in pancreatic cancer [2–5]. However, the sensitivity and specificity of this biomarker are not high, and serum levels are significantly increased in inflammatory diseases of the pancreas and biliary tract. Therefore, CA19-9 is not useful for early diagnosis, mass screening or for distinguishing between PDAC and chronic pancreatitis. Thus, there is a great need for new biomarkers for PDAC. In the absence of good biomarkers, 80% to 90% of PDAC cases are diagnosed too late in the disease process for surgical resection to be an effective option. Among the 10%

to 20% of PDAC cases where surgical resection is an option, most patients ultimately die of recurrent or metastatic disease [6].

The development and progression of PDAC is generally believed to be caused by the activation of oncogenes, inactivation of tumor suppressor genes and the dysregulation of cellular signal transduction pathways, e.g. EGFR, Akt and NFκB [7]. The analysis of gene mutations, growth factors and their receptors and other downstream signaling proteins may have utility in the early detection of PDAC. For instance, activating point mutations in codon 12 of the K-ras gene are present in the large majority of PDAC [8–10]. Mutations in p53 have also been observed in approximately 50% of PDAC tumors [11].

Mutations in a few key signaling proteins may deregulate the expression of a large number of downstream proteins that interact with each other, possibly through post-translational modifications such as phosphorylation and glycosylation. While both N-linked glycosylation and sialylation regulate receptor expression and signaling by modifying ligand binding sites [12], protein phosphorylation acts as a molecular switch to activate or deactivate diverse cellular signaling networks. Importantly, identification of deregulated proteins has utility in elucidation of the protein signaling networks. Such identifications may be facilitated by protein microarray technologies for proteome-wide screening.

Protein microarrays have utility as a high-throughput screening method for whole-cell lysates, fractionated proteomes, tissues, and antigen-antibody reactions [13–21]. Such microarrays, arrayed with naturally produced proteins that have been sepa-

rated using multi-dimensional liquid-based separation of a proteome, followed by the arraying of all proteins found in the individual fractions, can subsequently be probed with a variety of detection agents, including lectins for glycoprotein detection [22, 23]. In the present work, we have utilized protein microarrays to analyze post-translational modifications as potential epitopes that elicit a humoral response in PDAC. We have used 2-D HPLC based fractionation of Panc-1 derived proteins, followed by non-contact piezoelectric spotting for generation of protein microarrays. These microarrays were utilized for evaluation of the humoral response in patients with PDAC, and for assessing post-translational modifications on Panc-1 derived tumor antigens. Statistical analysis of the humoral response data facilitated a quantitative estimate of the humoral response against specific tumor antigens between the diagnosis groups. Specific tyrosine phosphorylation patterns were elucidated for each protein fraction using anti-phosphotyrosine antibodies. In addition, analysis of the glycoproteins involved in generation of a humoral response was obtained using five different biotinylated lectins to analyze the glycan structure of glycoproteins.

3.2 Experimental

3.2.1 Chemicals

Methanol, acetonitrile, urea, thiourea, iminodiacetic acid, DTT, OG, glycerol, bis-tris, TFA, and PMSF were obtained from Sigma (St. Louis, MO). Water was purified using a Milli-Q water filtration system (Millipore Inc., Bedford, MA) and all solvents were HPLC grade unless otherwise specified. Reagents used were in the purest form commercially available. Polybuffer 74 and polybuffer 96 were purchased from Amersham Pharmacia Biotech (Piscataway, NJ). Pro-Q Diamond phosphoprotein gel stain and Pro-Q Diamond phosphoprotein gel destaining solution were obtained from

Molecular Probes (Eugene, OR). 1X PBS and ultra-pure DNase/RNase free distilled water were obtained from Invitrogen (Carlsbad, CA). Mouse anti-phosphotyrosine antibody, clone 4G10 was obtained from Upstate (Charlottesville, VA), Mouse anti-phosphotyrosine antibody, clone PY20, was obtained from Perkin-Elmer and the Cy5-conjugated secondary antibodies were obtained from Jackson ImmunoResearch Lab Inc. (West Grove, PA.). Five biotinylated lectins (Aleuria aurentia (AAL), Maackia amurensis lectin II (MAL), peanut agglutinin (PNA), Sambucus nigra bark lectin (SNA) and Concanavalin A (ConA)) were all purchased from Vector Laboratories (Burlingame, CA). The streptavidin-AlexaFluor555 conjugate was obtained from Invitrogen (Carlsbad, CA.).

3.2.2 Serum Samples

Serum was obtained at the time of diagnosis following informed consent using IRB-approved guidelines. Sera were obtained from 15 patients with a confirmed diagnosis of pancreatic adenocarcinoma in the Multidisciplinary Pancreatic Tumor Clinic at the University of Michigan Hospital. These sera were randomly selected from a clinic population that sees, on average, at the time of initial diagnosis, 15% of pancreatic adenocarcinoma patients presenting with early stage (i.e., stage 1/2) disease and 85% presenting with advanced stage (i.e., stage 3/4). Inclusion criteria for the study included patients with a confirmed diagnosis of pancreatic cancer, the ability to provide written, informed consent, and the ability to provide 40 mL of blood. Exclusion criteria included inability to provide informed consent, patient's actively undergoing chemotherapy or radiation therapy for pancreatic cancer, and patients with other malignancies diagnosed or treated within the last 5 years. Sera were also obtained from 8 patients with chronic pancreatitis who were seen in the Gastroenterology

Clinic at University of Michigan Medical Center, and from 15 control healthy individuals collected at University of Michigan under the auspices of the Early Detection Research Network (EDRN). The mean age of the tumor group was 65.4 years (range 54-74 years) and from the chronic pancreatitis group was 54 years (range 45-65). The sera from the normal subject group was age and sex-matched to the tumor group. All of the chronic pancreatitis sera were collected in an elective setting in the clinic in the absence of an acute flare. All sera were processed using identical procedures. The samples were permitted to sit at room temperature for a minimum of 30 minutes (and a maximum of 60 minutes) to allow the clot to form in the red top tubes, and then centrifuged at 1,300×g at 4°C for 20 minutes. The serum was then removed, transferred to a polypropylene, capped tube in 1 mL aliquots, and frozen. The frozen samples were stored at -70°C until assayed. All serum samples were labeled with a unique identifier to protect the confidentiality of the patient. The handling of all serum samples was similar in that none of the samples were thawed more than twice before analysis.

3.2.3 Sample Preparation

Cell Culture

The Panc-1 PDAC cell line was cultured in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum, 100 units/mL penicillin and 100 units/mL streptomycin (Invitrogen, Carlsbad, CA). Upon reaching 80% confluence, the cells were washed twice in 10 mL 1XPBS containing 4 mM Na_3VO_4, 10 mM NaF and one half of a protease inhibitor cocktail tablet. The sample was then solubilized in 300 μL lysis buffer consisting of 7 M urea, 2 M thiourea, 100 mM DTT, 0.5% biolyte ampholyte 3-10, 2% OG, 4 mM Na_3VO_4, 10 mM NaF and 1 mM PMSF at room

temp for 30 min, followed by centrifugation at 35,000 rpm at 4°C for 1 hr. The supernatant was stored at -80°C until use.

Sample Preparation for Chromatofocussing (CF)

A PD10 column (Amersham Biosciences) was equilibrated with a pH 7.9 buffer solution containing 25 mM bis-tris in 6 M urea and 0.2% OG and then used to exchange the cell lysate from the lysis buffer to the CF buffer according to the manufacturer's protocol.

3.2.4 Chromatofocusing of Panc-1 Cell Lysate

The liquid separations were performed on the Gold Model 128 HPLC Pump along with a Model 166 UV Detector (Beckman-Coulter, Inc.). An AX300 4.6×250 mm (Eprogen Inc.) weak anion exchanger column was utilized for the first dimension separations. The start buffer consisted of 6 M urea, 0.2% OG, 25 mM bis-tris (pH adjusted to 7.9 using saturated IDA). The elution buffer consisted of 6 M urea, 0.2% OG and a 10-fold dilution of polybuffer 96 and polybuffer 74 in a ratio of 3:7 (the pH was adjusted to 4.0 using saturated IDA). The chromatofocusing column was pre-equilibrated with the start buffer until the pH of the effluent was the same as that of the start buffer. 13 mg of the cell lysate was applied to the chromatofocusing column with multiple injections. Once a stable baseline was achieved, the elution buffer was switched on to elute the proteins on the column in an isocratic mode. UV detection was performed at 280 nm and the pH of the effluent was monitored using the PF2D's flow-through on-line pH probe. The pH fractions were collected in 0.3 pH intervals and 15 fractions in total were collected over the range of pH 7.9-4.0. The

CF separation was completed when the pH of the effluent reached 4.0. The column was then washed with a 1 M NaCl solution followed by 100% IPA to elute out the strongly binding proteins as salt-wash and isopropanol-wash fractions respectively.

3.2.5 Reverse Phase HPLC Separations

RP-HPLC was performed using an ODS-I (8×33 mm) column (Eprogen Inc.). Solvent A was 0.1% TFA in water and Solvent B 0.1% TFA in acetonitrile. The gradient was run from 5% to 15% in 1 min, 15% B to 25% in 2 min, 25% to 31% in 2 min, 31% to 41% in 10 min, 41% to 47% in 6 min, 47% to 67% in 4 min, then up to 100% B in 3 min where it was held for 1 min, and then reduced to 5% in 1 min at a flow rate of 1 mL/min. The column temperature was 40°C higher than the ambient temperature. The UV absorption profile was monitored at 214 nm. The RP fractions were collected automatically in 96 well plates, then lyophilized to dryness using vacuum centrifugation at 75°C, and stored at -80°C until use.

3.2.6 Protein Microarrays

Printing

The fractionated Panc-1 proteins (1152 fractions) were resuspended in 60% ACN with 0.1% TFA and transferred into 96-well microtiter plates. The samples were then lyophilized to dryness, and then reconstituted in 15 μL printing buffer comprising of 125 mM Tris-HCL (pH 6.8), 1% w/v SDS, 5 w/v DTT, 1% glycerol and protease inhibitors in 1X PBS. All of the fractions were printed on nitrocellulose (Whatman) and/or SuperEpoxy (Telechem International) slides using a GeSiM Nanoplotter2 non-contact piezoelectric printer. Each spot measured approximately 300 μm in di-

ameter with a spot spacing of 600 μm. The slides were dried overnight at room temp and were either used immediately or stored in a desiccated chamber at -20°C.

Hybridization of Slides for Humoral Response Analysis

Nitrocellulose slides spotted with Panc-1 protein fractions were blocked in 1X PBS containing 1% BSA (Sigma, St Louis) and 0.1% Tween-20 overnight at 4°C. The slides were then incubated individually with serum from pancreatic adenocarcinoma patients, from normal subjects or from patients with chronic pancreatitis. The sera used were diluted to 1:400 in probe buffer (1X PBS, pH 7.4 containing 1% BSA, 5 mM $MgCl_2$, 0.5 mM DTT, 0.05% Triton X-100 and 5% glycerol) and hybridized to the microarray slides. The slides were incubated for 2 hr at 4°C, washed 5 times with probe buffer for 5 min each, and then incubated with Alexa Fluor 647-conjugated anti-human IgG (1:2000, Invitrogen, Carlsbad, CA) for 1 hr at 4°C. The microarrays were then washed 5 times for 5 min each with probe buffer, dried and scanned on an Axon 4000A scanner (Axon Instruments Inc., Foster City, CA). GenePix Pro 6.0 software (Molecular Devices, Sunnyvale, CA) was used for data acquisition and preliminary data filtering.

Hybridization with Lectins to Delineate the Glycan Structure of Panc-1 Glycoproteins

The microarrays were blocked in 1% BSA in 1X PBS-T (with 0.1% Tween-20) overnight at 4°C. The slides were then incubated with biotinylated lectins diluted in 1X PBS-T. The lectins used were biotin conjugated Aleuria aurentia (AAL), Maackia amurensis lectin II (MAL), peanut agglutinin (PNA), Sambucus nigra bark lectin (SNA) and Concanavalin A (ConA). The working concentration of all lectins used

was 5 μg/mL except for SNA, which was used at 10 μg/mL. After primary incubation, all slides were washed 5 times with 1X PBS-T for 5 min each. Secondary hybridization was performed with a streptavidin-Alexa Fluor 555-conjugate (Invitrogen, Carlsbad, CA) in a working concentration of 1μg/mL in 1X PBS-T containing 0.5% BSA. After secondary hybridization, the microarrays were washed 5 times for 5 min each in 1X PBS-T and then completely dried by centrifugation. The dried microarray slides were subsequently scanned on an Axon 4000A scanner. GenePix Pro 6.0 software was used for data acquisition and preliminary data filtering.

Hybridization to Delineate Panc-1 Phosphoproteins

The microarray slides were blocked overnight in 1% BSA in 1X PBS-T The microarrays were then incubated in goat monoclonal anti-phosphotyrosine (Upstate, Charlottesville, VA) antibody diluted to 2 μg/mL in probe buffer After primary hybridization, the slides were washed 5 times for 5 min each in probe buffer. Secondary hybridization was performed for 1 hr using Cy3 conjugated anti-mouse antibody at a concentration of 1 μg/mL in probe buffer. The microarray slides were washed 5 times for 5 min each in probe buffer, dried by centrifugation and scanned on an Axon 4000A scanner. GenePix Pro 6.0 software was used for data acquisition and preliminary data filtering.

Microarray Data Acquisition and Filtering

All processed slides were scanned using an Axon 4000A scanner. GenePix Pro 6.0 software was used for data acquisition and preliminary data filtering. Single channel intensity values for the green channel were obtained for each microarray processed

with lectins and the intensity values for the red channel were obtained for each microarray processed with human sera. Initial spot analysis was performed with GenePix software, where all irregularly formed spots manually flagged. The background corrected median intensities were calculated. Spots that had negative intensity for greater than 50% of the serum samples were removed. Each array was then centered on its median intensity value and scaled by the median absolute deviation (MAD). Quantile matching was used to standardize the data across arrays.

3.2.7 Statistical Analysis of Humoral Response Data

A supervised analysis was conducted with the humoral response array data (all 1152 fractions) to select the protein fractions that were predictive of cancer. Array data, normalized as described above, was applied to a test statistic-based feature selection procedure. Protein signatures of varying length (10 to 100 proteins) were used to build a Support Vector Machine (SVM) prediction model. The SVM over multiple kernel test permutations was embedded in a finite grid-search of paired values of exponentially growing sequences of cost (C) and gamma (γ). Leave-one-out cross-validation (LOOCV) was used to evaluate the performance of the signatures. A signature that yielded the best accuracy was selected. Principal component analysis (PCA) was also used to verify the accuracy of the differentiation among the diagnostic groups. All statistical analyses were performed using R 2.3.0 and SPSS.

3.2.8 Protein Digestion

The UV fractions collected from the 2^{nd} dimension RP-HPLC chromatogram that showed a positive response to PTM detection experiments or that demonstrated pos-

itive reactivity from the humoral response experiment were collected and dried down to 20 μL in order to eliminate ACN and TFA. 1 M NH_4HCO_3 and 10 mM DTT were then added to a final concentration of 100 mM and 1 mM, respectively and incubated at 60°C for 15 min. Trypsin was then mixed with the denatured proteins at the ratio of 1:50. The mixture was incubated at 37°C for 24 hr.

3.2.9 Peptide Sequencing by LC-MS/MS

Trypsin digested samples were separated by a capillary reversed phase chromatography column (MagicAQ C18, 0.1×150 mm) (Michrom Biosciences, Auburn, CA) on a Paradigm MG4 micro-pump (Michrom Biosciences) with a flow rate of 300 nL/min. Peptides were eluted using a 45 min gradient which was started at 3% B, increased to 35% B in 25 min, 60% B in 15 min, 90% in B min, maintained at 90% B for 1 min and finally changed back to 3% B in 3 min. Both solvents A (water) and B (acetonitrile) contained 0.1% formic acid and 0.05% HFBA. The resolved peptides were analyzed on an LTQ mass spectrometer (Thermo, San Jose, CA) with a nano-ESI platform (Michrom Biosciences) operating in data dependent mode with dynamic exclusion enabled. The capillary temperature was set at 200°C, the spray voltage was 2.4 kV, and the capillary voltage was 20 V. The normalized collision energy was set at 35% for MS/MS. The MS/MS spectra for the top three peaks from full MS scan were obtained. The spectra were searched using SEQUEST algorithm against the non-redundant Swiss-Prot human protein database. One missed cleavage was allowed during the database search. Protein identification was considered positive for a peptide with Xcorr of greater than or equal to 3.0 for +3, 2.5 for +2, and 1.8 for +1 charged ions. Peptide identification accuracy for protein identification was further increased using PeptideProphet (maintained by Trans Proteomic

Pipeline Project). PeptideProphet [24] validates peptide assignments from MS/MS spectra which in turn is used to validate protein identities using ProteinProphet [25].

3.3 Results and Discussion

Clinical detection of early stage pancreatic cancer has been hampered, in part due to the lack of suitable biomarkers. In this study, we used the Panc-1 human pancreatic ductal adenocarcinoma cell line to analyze the humoral response in pancreatic cancer, as Panc-1 cell lines have been shown to maintain some of the differentiated characteristics of normal mammalian pancreatic ductal epithelial cells [26]. The analytical work flow is depicted diagrammatically in Figure 3.1. The Panc-1 cell line was lysed as described in Methods. 13 mg solubilized protein was subjected to 2-D HPLC in order to resolve sufficient protein for the microarray analysis. The first dimension separation, between pH 7.9-4.0, was achieved using a weak anion exchange column. A 4.6×250 mm column was used to resolve sufficient protein for all the experiments. Each of the collected pI fractions from the first dimension separation were then resolved in the second dimension using nonporous silica reverse phase HPLC on an 8×33 mm NPS C18 column. 1152 protein fractions were obtained following RP-HPLC. All fractions were subsequently printed on nitrocellulose-coated glass slides as described in Methods.

For the humoral response experiment, 38 serum samples were hybridized individually to the protein microarrays. The serum included 15 sera from pancreatic cancer patients, 15 from normal subjects and 8 serum samples from patients with chronic pancreatitis. Following hybridization of all the slides with patient serum, in parallel

so as to mitigate any day-to-day variation, the obtained microarray data was used for rigorous statistical analysis.

Statistical Analysis for Protein Selection: To eliminate the humoral response signature of all protein fractions that are related to pancreatic inflammation, but which are not cancer-specific, protein fractions which distinguished chronic pancreatitis sera from normal sera were first identified. SVM algorithm was used on the data from both chronic pancreatitis and normal samples, and LOOCV was used to train and test possible protein signatures. A signature comprising 68 protein fractions had the highest accuracy at 69.6%. 72 proteins were found to be in common among the 23 protein signatures generated by LOOCV. After exclusion of these 72 proteins from the data set (group A; Figure 3.4), a comparison between the control and cancer serum samples yielded a 28 protein fraction signature with 60% accuracy, with 33 common protein fractions among the signatures (group B, Figure 3.4). Further, a broader comparison between normal and cancer samples conducted without excluding the protein fractions in the chronic pancreatitis signature yielded a 23 protein signature with 60% accuracy, comprising 23 common protein fractions (group C, Figure 3.4). The heat-map of protein fractions obtained from the above analysis is shown in Figure 3.2. To obtain an overview of how well the hybridization intensities of the protein fractions obtained in this experiment would differentiate the three diagnostic groups, principal component analysis (PCA) was performed. Figure 3.3 shows the first two components from PCA. Although exclusive separation was not observed, very different clustering patterns were seen between the three diagnostic groups.

Humoral Response: Following the low accuracy of the results from the statistical analysis, a complementary filtering method was used to increase the reliability of the biomarker selections in which manual analysis of the fluorescence intensities indicating differential humoral response was conducted. The differences in overall response to the three diagnostic serum groups for some spots was found to be very small. Because of the likelihood of experimental variations arising from sample handling, such small differences were ignored even when the centroid of signal intensity for the differential response were different. Thereafter, the data was reprocessed to retain those spots that showed higher response for cancer sera for at least 30% of the samples compared to both normal and pancreatitis. In this respect, it is important to note that spots showing higher humoral response in normal compared to cancer were also registered in small numbers and may indicate a loss in humoral response. It is argued that around 30-40% of cancer patients will respond to any given cancer antigen. This arises from the fact that serum samples obtained from different patients are biologically unique as each person in the study has an unique genetic makeup. Various environmental factors like food habits also critically affect immune response. This feature relating to the diversity of response to cancer antigens is difficult to process using statistical algorithms. As such, the use of complementary manual inspection of the data proved to be useful. The candidates short-listed after filtering of data using this method were then plotted as scatter-plots as shown in Figure 3.7. Using the above technique the discriminating nature of 16 fractions were verified and 39 more protein fractions were selected. These 39 fractions showed differential response but were not detected by the statistical analysis. All the short-listed protein fractions were then digested and MS/MS analysis was conducted for their identification as described in Section 3.2.9. The use of complementary techniques provided a larger set

of proteins that showed differential response and reduced the chance of false positives.

Table 3.1 shows the proteins that had been identified in the first set (16 fractions) showing differential response between the three diagnostic groups (A-C; Figure 3.4), among which two proteins had previously been associated with pancreatic cancer and four of the others were observed to be up-regulated in breast cancer. The proteins identified in the second group (39 fractions) are shown in Table 3.3 and Table 3.2 shows the peptides identified for each protein. 37 proteins were identified in this group including previously identified pancreatic cancer biomarkers like Cu-Zn superoxide dismutase. Some overlap in proteins identified between the two sets was observed where often neighboring fractions were selected as part of the two groups. Figure 3.4 depicts the humoral response pattern against all the tested serum samples. The top 10 proteins in the figure are from group B which lists the proteins that are differentially responsive to cancer and normal sera (with proteins reactive to chronic pancreatitis sera eliminated from the analysis). PDZ and LIM domain 1 protein, Histidine Triad Nucleotide and RAD50 Homolog isoform 1 protein showed a differential humoral response between chronic pancreatitis and normal sera. PDZ domain proteins are common protein interaction modules that play key roles in cellular signaling [27]. The majority of PDZ-containing proteins are associated with the plasma membrane [28], where they take part in signaling, mediating the adhesive properties of particular cells, ion transport and formation of tight junctions. Over-expression of PDZ domain-containing proteins in chronic pancreatitis, which change the nature of interaction in the plasma membrane or epithelial region [29], may be a differentiating feature between normal pancreata and chronic pancreatitis. Histidine Triad Nucleotide protein (spectra shown in Figure 3.5a), also known as PKC-interacting

protein (PKCI), has been identified in MFC-7 human breast carcinoma cells and it probably plays the role of a tumor suppressor [30] protein, though its role in the development of chronic pancreatitis is unknown. RAD50 Homolog forms a complex with MRE11 and NSB1 and subsequently binds to DNA and plays important role in DNA double-strand break repair. It exhibits decreased humoral response in PDAC and chronic pancreatitis serum, as compared to normal sera. Mutations in RAD50 have been observed in breast cancer [31]. Though a phosphorylated peptide was not detected by MS/MS analysis, the microarray data indicates possible tyrosine phosphorylation, as evident by reactivity to the PY20 anti-phosphotyrosine antibody. 10 proteins were identified from among the group B proteins, including Vimentin and α-Enolase (spectra shown in Figure 3.5b). Both of these proteins had previously been observed to be up-regulated in pancreatic cancer tissue when compared with normal (and chronic pancreatitis) tissues using 2-D gel electrophoresis and mass spectrometry [32]. The cytoskeleton-associated protein Vimentin has been found to play an important role in the (TGF-β)-induced cell migration and invasion [33]. In a recent study, a single isoform of vimentin has been shown to elicit a humoral response in pancreatic cancer, as compared to both chronic pancreatitis and normal sera [34]. Interestingly, in the present study vimentin was shown to elicit a humoral response in both PDAC and chronic pancreatitis sera. Similar reactivity was also seen with α-Enolase, which was previously shown to have utility in differentiating PDAC tissue from normal pancreata.

Thrombospondin-2 and elongation factor alpha-1 (eEF1A) were both observed to have greater reactivity against PDAC sera as compared to control. The microarray image in Figure 3.6 indicates the spot associated with Thrombospondin-2 that shows

higher binding to auto-antibodies in PDAC serum. Figure 3.7 shows the spot fluorescence intensity pattern of Thrombospondin-2 and eEF1A in which greater reactivity is observed for PDAC sera when compared to the others. eEF1A is responsible for the enzymatic delivery of aminoacyl tRNAs to the ribosome and is expressed in other tissues besides pancreas. eEF1A has been identified as an auto-antigen in 66% of patients with Felty's syndrome [35]. Amplification and over-expression of Elongin C was detected in both the breast cancer cell line SKBR-3 and in prostate cancer by cDNA microarrays [36]. In our study, however, the immunoreactivity of Elongin C between the diagnostic groups was not significantly different. Among the proteins in group C, Heat shock cognate (71 kDa) protein, a chaperone protein that is involved in the cellular transport, was observed to exhibit lower immunoreactivity with PDAC sera as compared to normal, while the regulatory protein hnRNPA2/B1 showed greater immunoreactivity with cancer sera.

Analysis of Post-Translational Modifications: The nature and extent of post-translational modifications in critical cell-signaling proteins depends, to some extent, upon the progression of cancer. For example, changes in phosphorylation patterns have been observed in breast cancer progression in response to therapeutic drugs. Our work utilizes a novel method in which the modification pattern of the proteins on the microarray can be utilized to obtain information on the deregulated pathway, and to aid in protein identification using mass spectrometry. The type of modification detected through a blind database search of the MS/MS data can be verified with data from the microarray experiments, which provides practical approach for identification of modifications in those situations where very high sequence coverage is difficult to obtain. To this end, we utilized protein microarrays that were printed

at the same time as those utilized in the humoral response experiments above. Mouse Anti-phosphotyrosine antibodies were used for the detection of protein phosphorylation and five biotinylated lectins were used to map the glycosylation pattern of the Panc-1 proteome. Figure 3.8 demonstrates the glycosylation and phosphorylation patterns in the Panc-1 proteome along with humoral response. It is interesting to note that after the spot intensities from the microarray data were normalized for UV peak intensities, and the spots having intensities below the cut-off value dictated by negative standards spotted on the array were eliminated, almost all the proteins showing intense humoral response (top portion of the Figure 3.8) were observed not to be glycosylated, or only marginally so, with very low spot intensities. With the exception of a few proteins that showed sialylation (as dictated by their response to SNA lectin), glycosylations were largely detected in proteins from the injection peak fraction comprising of proteins having a pI greater than 7.9 (lower half of Figure 3.8). Since glycosylation adds increasing amounts of negatively charged sugars and eliminates a positive charge on asparagine, the modified protein will have a lower pI and bulky glycan side chains can make the proteins difficult to elute from the first dimension CF column.

As such, the glycosylation pattern on the microarray may be incomplete and hence, the data presented here only demonstrates a methodology that can be applied for effectively increasing the amount of data that can be obtained from a microarray experiment. Though an analysis of the flow-through during loading of both the chromatofocusing column and the reverse phase column was not conducted, there is a possibility that some of the heavily glycosylated proteins were lost due to poor chromatographic separation. Figure 3.9 demonstrates the above method where ni-

trocellulose slides printed for the purpose of the humoral response experiment was processed using SNA lectin. The spot labeled 'c' (Thrombospondin-2) is thus probably sialylated. Further MS/MS experiments for the verification of the abovementioned type of glycosylation or for the identification of the glycosylated peptide/s were not pursued and the glycosylation state of the protein was only used as a factor in improving protein identification and selection from the MS/MS database search results. However, since blind database search for phosphorylated peptides was easier to perform, the phosphorylation data from the microarray was used to verify the database search results. Figure 3.10 shows the phosphorylated peptide DMRQpTVAVGVIK from Elongation Factor 1A. Through the blind search of MS/MS data, the phosphorylated peptides TVETRDGQVINEpTpSQHHDDLE and SGAQASSTPLpSPTR were also identified, helping in validating Vimentin and Lamin A/C, respectively, as the correct protein IDs. Such a method which incorporates experimental modification data to verify protein IDs obtained through mass spectrometry is more reliable than mass spectrometry data alone.

3.4 Conclusion

The use of complementary data analysis techniques for biomarker discovery using protein microarrays is essential to reduce the chance of false positives. Because of highly diverse nature of humoral response in general, a clear differentiation is difficult to observe. Moreover, for a complex disease like pancreatic cancer which is characterized by a lack of strong humoral response and difficulty in detection till late stages, the possibility of observing a clear differentiation is very low. In this respect, the use of complementary techniques to analyze the microarray data is essential.

The top-down mass spectrometric method that has been used in this experiment for protein identification is able to more easily identify modified peptides, as compared to a bottom-up method where enrichment is usually necessary. Without peptide enrichment in the positive ion mode, ionization efficiency of the phosphopeptides is suppressed by the presence of other non-phosphorylated peptides to some extent. Thus, the purity of the sample fractionated using multidimensional liquid-based separation methodology allows us to use the modification information from the microarray in eliminating false positives. Unlike bottom-up methods, where the cell lysate proteins are digested and the protein-to-peptide correlation is lost, in this method, all identified peptides can be assigned to a single protein. Owing to the complex and competitive process associated with generation of sample ions in mass spectrometry, peptide coverage is usually low for a moderately high molecular weight protein. In this respect, the availability of information regarding the phosphorylation or glycosylation state of a protein is immensely helpful in correctly identifying the protein. Though humoral response experiments using protein microarrays and mass spectrometry method were previously developed, the use of protein modification on the microarray and its subsequent application for improving the reliability of the mass spectrometry based protein identifications was not previously described. This work undertakes a proof-of-concept study to demonstrate the effectiveness and simplicity of such methodology. The power of this method also lies in its ability to identify a large group of proteins in a single experiment that are co-regulated in their post translational modifications and which also elicit a humoral response. Detection and analysis of such co-regulated proteins will enable delineation of functional pathways that play an important role during cancer initiation and progression.

Table 3.1: List of proteins showing humoral response and identified by nESI-LC-MS/MS. Positive change (+), No change (∼) and Negative Change (−).

Protein Name	**Acc No.**	**Identified Phosphorylated Peptide**	**Theo. Protein MW(Da) /pI**	**Obs pH Range**	**Humoral Response**	
					C/N	P/N
Thrombospondin-2	P35442		129955/4.6	7.6-7.3	+	∼
Eukaryotic Elongation Factor 1 Alpha 1 (eEF1A)	NP001393	DMRQpTVAVG VIK	50161/9.1	6.4-6.1	+	∼
Elongin C	Q15369		12473/4.77	7.3-7.0	∼	∼
Acyl-CoA dehydrogenase	P28330		47669/7.61	7.0-6.7	∼	∼
Transcription elongation factor A protein 1	P23193		33969/8.38	6.4-6.1	-	∼
Lamin A/C	P02545	SGAQASSTPL pSPTR	74139/6.57	5.2-4.9	-	∼
Vimentin	P08670	TVETRDGQVI NEpTpSQHHD DLE	53652/5.06	4.9-4.6	+	+
Rab-17	Q9H0T7		23490/7.84	6.1-5.8	-	∼
Heterogeneous Nuclear Ribonuclear Protein A2/B1	P22626		37429/8.95	7.9-7.6	+	∼
α-Enolase	P06733		47168/7.38	5.8-5.5	+	+
PDZ and LIM Domain 1	NP066272		36071/7.02	7.3-7.0	-	+
Histidine Traid nucleotide binding protein	P49773		13801/6.95	7.3-7.0	-	+
RAD50 Homolog Isoform 1	NP005723		153891/6.89	7.9-7.6	-	-
ATP Synthase	NP001676		12587/9.53	7.6-7.3	∼	∼
Glyceraldehyde 3-Phosphate Dehydrogenase	AAP36549		36166/8.45	7.3-7.0	-	∼
WD Repeat Domain 35	Q9P2L0		133546/6.37	7.0-6.7	∼	∼
Heat Shock Protein 7C	P11142		70897/5.52	5.8-5.5	-	∼

Table 3.2: Database search results for proteins analyzed using ESI LC-MS/MS. Proteins were identified using atleast 3 unique peptides.

Protein Name (Acc#)	MW Theo.	pH Theo.	pH Obs.	%AA Cov.	[M+H]+ Exp.	[M+H]+ Theo.	AA Position	Peptide (Charge)	Xcorr	ΔCn
10 kDa heat shock protein, mitochondrial (P61604)	10794.8	8.91	> 7.9	**66.34**	717.20	717.38	92-98	**R.DGDILGK.Y (1+)**	1.89	0.23
					860.32	860.46	28-35	**K.GGIM'LPEK.S (2+)**	2.65	0.19
					1013.36	1013.56	56-65	**K.GGEIQPVSVK.V (2+)**	3.57	0.22
					1198.52	1198.68	54-65	**K.GKGGEIQPVSVK.V (2+)**	4.13	0.55
					1260.48	1260.66	28-39	**K.GGIM'LPEKSQGK.V (2+)**	3.35	0.25
					1315.52	1315.76	40-53	**K.VLQATVVAVGSGSK.G (2+)**	4.99	0.59
					1475.56	1475.81	66-79	**K.VGDKVLLPEYGGTK.V (2+)**	4.54	0.50
					2048.34	2048.07	20-39	**R.SAAETVTKGGIM'LPEKSQGK.V (3+)**	4.49	0.51
Heterogeneous nuclear ribonucleoprotein A1 (P09651)	38692.1	9.26	> 7.9	**23.18**	1234.48	1234.64	130-139	**K.IEVEIM'TDR.G (2+)**	3.91	0.42
					1299.44	1299.65	3-13	**K.SESPKEPEQLR.K (2+)**	2.83	0.35
					1427.63	1427.75	3-14	**K.SESPKEPEQLRK.L (2+)**	3.01	0.36
					1437.59	1437.75	92-104	**R.EDSQRPGAHLTVK.K (2+)**	3.15	0.14
					1565.64	1565.84	92-105	**R.EDSQRPGAHLTVKK.I (2+)**	2.87	0.27
					1628.60	1628.78	336-351	**R.SSGPYGGGGQYFAKPR.N (2+)**	4.26	0.54
					1694.43	1694.70	352-369	**R.NQGGYGGSSSSSSYGSGR.R (2+)**	5.14	0.65
					1879.82	1879.97	106-121	**K.IFVGGIKEDTEEHHLR.D (3+)**	3.98	0.55
Heterogeneous nuclear ribonucleoproteins A2/B1(P22626)	37407.7	8.97	> 7.9	**16.43**	1013.19	1013.44	204-213	**R.GGNFGFGDSR.G (2+)**	2.98	0.20
					1013.24	1013.44	204-213	**R.GGNFGFGDSR.G (2+)**	2.77	0.18
					1188.39	1188.65	138-147	**K.IDTIEITDR.Q (2+)**	3.83	0.50
					1220.19	1221.55	192.200	**R.QEMQEVQSSR.S (2+)**	2.71	0.25
					1338.52	1338.70	100-112	**R.EESGKPGAHVTVK.K (2+)**	3.37	0.48
					1377.47	1377.63	214-228	**R.GGGGNFGPGPGSNFR.G (2+)**	3.12	0.41
					1377.40	1377.63	214-228	**R.GGGGNFGPGPGSNFR.G (2+)**	3.19	0.39
Profilin-1(P07737)	14914.5	8.47	> 7.9	**19.42**	1234.43	1234.71	105-115	**K.TDKTLVLLM'GK.E (2+)**	3.76	0.52
					1397.44	1379.72	91-104	**K.STGGAPTFNVTVTK.T (2+)**	3.73	0.49
					1952.73	1953.03	89-107	**R.TKSTGGAPTFNVTVTKTDK.T (3+)**	5.07	0.52
Nuclease sensitive element binding protein 1(P67809)	35903.7	9.87	7.9-7.6	**30.56**	1587.43	1587.70	171.185	**K.NEGSESAPEGQAQQR.R (2+)**	5.47	0.51
					1695.67	1695.87	119-137	**K.GAEAANVTGPGGVPVQGSK.Y (2+)**	5.16	0.61
					1897.43	1897.81	305-324	**K.AADPPAENSSAPEAEQGGAE.- (2+)**	3.64	0.47
					2871.75	2871.31	205-231	**R.RPQYSNPPVQGEVMEGADNQGAGEQGR.P (3+)**	6.78	0.62
					3240.26	3239.53	205-234	**R.RPQYSNPPVQGEVM'EGADNQGAGEQGRPVR.Q (4+)**	5.32	0.56
60S ribosomal protein L13 (P26373)	24116.5	11.65	7.6-7.3	**34.76**	847.27	847.44	167-173	**R.VITEEEK.N (2+)**	2.53	0.27
					950.31	950.49	74-81	**R.GFSLEELR.V (2+)**	2.94	0.29
					973.47	973.62	121-128	**R.SKLILFPR.K (2+)**	2.87	0.39
					1063.35	1063.53	135-144	**K.KGDSSAEELK.L (2+)**	3.09	0.31
					1217.49	1217.60	198-208	**R.AKEAAEQDVEK.K (3+)**	4.19	0.32
					1233.27	1232.62	105-115	**K.STESLQANVQR.L (2+)**	3.80	0.35
					1236.48	1236.65	167-176	**R.VITEEEKNFK.A (2+)**	3.24	0.39
					1382.52	1382.78	145-157	**K.LATQLTGPVMPVR.N (2+)**	3.52	0.52
					1398.47	1398.78	145-157	**K.LATQLTGPVM'PVR.N (2+)**	3.88	0.56
					1475.55	1474.76	103-115	**R.NKSTESLQANVQR.L (2+)**	4.45	0.55
					2443.05	2443.29	135-157	**K.KGDSSAEELKLATQLTGPVM'PVR.N (3+)**	5.43	0.48
60S ribosomal protein L27 (P61353)	15657.7	10.56	7.6-7.3	**34.07**	826.42	826.55	9-16	**K.VVLVLAGR.Y (+1)**	1.86	0.34
					933.39	933.56	51-59	**R.KVTAAMGKK.K (2+)**	2.50	0.28
					1049.35	1049.55	84-92	**R.YSVDIPLDK.T (2+)**	2.95	0.41
					1077.48	1077.61	93-101	**K.TVVNKDVFR.D (2+)**	2.55	0.41
					1407.48	1407.68	73-83	**K.VYNYNHLM'PTR.Y (2+)**	3.76	0.48
					1423.48	1423.68	73-83	**K.VYNYNHLM'PTR.Y (2+)**	3.38	0.49
					1590.72	1590.87	84-97	**R.YSVDIPLDKTVVNK.D (2+)**	4.36	0.41
60S ribosomal protein L21 (P46778)	18423	10.49	7.6-7.3	**40.88**	821.44	821.47	129-135	**K.RQPAPPR.E (2+)**	2.68	0.20
					1243.51	1243.66	21-31	**K.HGVVPLATYM'R.I (2+)**	3.05	0.34
					1259.47	1259.66	21-31	**K.HGVVPLATYM'R.I (2+)**	2.98	0.48
					1291.59	1291.76	32-42	**R.IVKKGDIVDIK.G (2+)**	3.13	0.38
					1640.72	1640.91	63-77	**R.VYNVTQHAVGIVVNK.Q (2+)**	5.52	0.54
					1995.98	1996.13	63-80	**R.VYNVTQHAVGIVVNKQVK.G (3+)**	4.20	0.55
					2094.79	2094.01	142-159	**R.TNGKEPELLEPIPYEFM'A.- (2+)**	3.11	0.43
DNA-binding protein A (Cold shock domain protein A) (P16989)	40037	9.77	7.3-7.0	**8.87**	1744.89	1744.89	110-124	**R.NDTKEDVFVHQTAIK.K (3+)**	3.90	0.42
					1796.51	1795.82	134-150	**R.SVGDGETVEFDVVEGEK.G (2+)**	5.11	0.44
					1873.64	1872.98	110-125	**R.NDTKEDVFVHQTAIKK.N (3+)**	4.79	0.43
Calgranulin B (P06702)	13234.5	5.71	7.3-7.0	**44.74**	877.26	877.48	44-50	**K.DLQNFLK.K (1+)**	1.83	0.06
					1005.47	1005.57	43-50	**R.KDLQNFLK.K (2+)**	3.23	0.11
					1455.43	1455.72	26-38	**K.LGHPDTLNQGEFK.E (2+)**	3.44	0.47
					1758.48	1758.82	58-72	**K.VIEHIM'EDLDTNADK.Q (2+)**	5.53	0.56
					1806.63	1806.94	11-25	**R.NIETIINTFHQYSVK.L (2+)**	4.88	0.52
Histone H1.5 (Histone H1a) (P16401)	22436.4	10.91	7.3-7.0	**15.11**	1014.43	1014.63	57-66	**R.NGLSLAALKK.A (2+)**	2.57	0.29
					1093.31	1093.55	67-77	**K.ALAAGGYDVEK.N (2+)**	2.75	0.46
					1212.39	1212.68	37-48	**K.ATGPPVSELITK.A (2+)**	3.34	0.49

					1340.48	1340.78	36-48	R.KATGPPVSELITK.A (3+)	3.53	0.43
Cytochrome c oxidase polypeptide Vb, mitochondrial precursor (P10606)	13688	9.07	7.3-7.0	29.46	775.24	775.44	50-56	R.EIMLAAK.K (1+)	1.93	0.18
					1187.35	1186.65	58-68	K.GLDPYNVLAPK.G (2+)	3.36	0.57
					1312.39	1312.67	75-86	R.EDPNLVPSISNK.R (2+)	2.77	0.51
					1315.43	1314.74	57-68	K.KGLDPYNVLAPK.G (3+)	4.05	0.16
					1314.35	1314.74	57-68	K.KGLDPYNVLAPK.G (2+)	2.66	0.27
					1841.67	1841.94	69-86	K.GASGTREDPNLVPSISNK.R (2+)	4.25	0.42
					1997.91	1998.04	69-87	K.GASGTREDPNLVPSISNKR.I (2+)	3.57	0.30
Phosphoglycerate kinase 1 (P00558)	44456.1	8.30	7.0-6.7	64.42	734.22	734.45	91-96	K.DVLFLK.D (1+)	2.12	0.21
					809.22	809.42	323-329	K.YAEAVTR.A (2+)	2.52	0.52
					867.35	867.48	184-191	K.AGGFLM'KK.E (2+)	2.57	0.21
					929.22	920.48	353-360	K.ALM'DEVVK.A (2+)	2.78	0.35
					927.43	926.60	39-47	R.IKAAVPSIK.F (2+)	2.71	0.18
					937.44	937.51	322-329	K.KYAEAVTR.A (2+)	2.64	0.36
					950.28	949.48	22-29	R.VDFNVPMK.N (2+)	2.78	0.31
					966.43	965.48	22-29	R.VDFNVPM'K.N (2+)	2.84	0.27
					974.28	974.46	123-130	R.FHVEEEGK.G (2+)	2.70	0.39
					986.51	986.60	206-215	R.PFLAILGGAK.V (2+)	3.48	0.40
					1012.39	1012.55	191-198	K.KELNYFAK.A (2+)	3.05	0.21
					1044.40	1044.63	6-14	K.LTLDKLDVK.G (2+)	2.63	0.27
					1097.27	1097.62	406-416	K.VLPGVDALSNL.- (2+)	3.17	0.52
					1101.08	1101.54	30-38	K.NNQITNNQR.I (2+)	3.20	0.15
					1131.39	1131.62	141-150	K.AEPAKIEAFR.A (2+)	2.67	0.43
					1159.40	1159.57	123-132	R.FHVEEEGKGK.D (2+)	3.11	0.44
					1202.35	1201.66	220-229	K.IQLINNMLDK.V (2+)	3.44	0.19
					1217.40	1217.66	220-229	K.IQLINNM'LDK.V (2+)	3.10	0.35
					1247.47	1247.69	75-85	K.YSLEPVAVELK.S (2+)	3.38	0.48
					1318.43	1318.73	279-290	K.ITLPVDFVTADK.F (2+)	3.07	0.48
					1367.31	1367.71	171-183	R.AHSSMVGVNLPQK.A (2+)	4.25	0.53
					1384.51	1383.71	171-183	R.AHSSM'VGVNLPQK.A (2+)	3.93	0.40
					1615.59	1614.89	216-229	K.VADKIQLINNMLDK.V (2+)	4.76	0.35
					1630.64	1630.88	216-229	K.VADKIQLINNM'LDK.V (3+)	3.92	0.31
					1635.47	1634.79	156-170	K.LGDVYVNDAFGTAHR.A (2+)	4.58	0.56
					1740.25	1740.91	388-405	K.VSHVSTGGGASLELLEGK.V (2+)	5.34	0.54
					1743.59	1743.88	230-245	K.VNEMIGGGM'AFTFLK.V (2+)	4.13	0.30
					1743.56	1743.88	230-245	K.VNEM'IGGGMAFTFLK.V (2+)	5.44	0.27
					1759.59	1759.88	230-245	K.VNEM'IGGGM'AFTFLK.V (2+)	5.37	0.05
					1768.77	1769.00	199-215	K.ALESPERPFLAILGGAK.V (2+)	5.07	0.33
					1966.59	1966.94	246-263	K.VLNNMEIGTSLFDEEGAK.I (2+)	5.75	0.56
					1983.59	1982.94	246-263	K.VLNNM'EIGTSLFDEEGAK.I (2+)	5.74	0.59
					2022.68	2023.04	279-296	K.ITLPVDFVTADKFDENAK.T (2+)	5.34	0.57
					2034.81	2035.05	56-74	K.SVVLMSHLGRPDGVPMPDK.Y (3+)	4.65	0.26
					2050.77	2051.04	56-74	K.SVVLMSHLGRPDGVPM'PDK.Y (3+)	3.93	0.32
					2513.97	2514.25	382-405	K.WNTEDKVSHVSTGGGASLELLEGK.V (3+)	5.41	0.60
					3264.41	3263.72	56-85	K.SVVLMSHLGRPDGVPMPDKYSLEPVAVELK.S (4+)	6.04	0.54
					3280.26	3279.71	56-85	K.SVVLM'SHLGRPDGVPMPDKYSLEPVAVELK.S (4+)	5.13	0.37
					3296.26	3295.71	56-85	K.SVVLM'SHLGRPDGVPM'PDKYSLEPVAVELK.S (4+)	4.59	0.47
Cyclin-dependent kinase inhibitor 1B (P46527)	22060.7	6.54	7.0-6.7	28.79	2386.76	2387.11	167-189	R.ANRTEENVSDGSPNAGSVEQTPK.K (3+)	4.93	0.53
					3454.25	3453.73	101-134	K.VPAQESQDVSGSRPAAPLIGAPANSEDTHLVDPK.T (4+)	4.51	0.41
28 kDa heat- and acid-stable phosphoprotein (Q13442)	20618.6	8.84	6.4-6.1	44.75	936.35	936.48	174-181	R.M'QSLSLINK.- (2+)	3.02	0.15
					1076.44	1076.59	173-181	K.RMQSLSLINK.- (2+)	2.82	0.20
					1085.35	1085.58	96-105	K.VTQLDLDGPK.E (2+)	3.07	0.34
					1133.52	1133.62	162-172	R.KAKDDATLSGK.R (2+)	3.28	0.34
					1161.43	1161.62	163-173	K.AKDDATLSGKR.M (2+)	3.29	0.34
					1213.43	1213.68	95-105	K.KVTQLDLDGPK.E (2+)	3.43	0.47
					1425.47	1425.73	76-88	K.GVEGLIDIENPNR.V (2+)	3.92	0.54
					1553.85	1553.83	75-88	R.KGVEGLIDIENPNR.V (2+)	5.21	0.52
					1709.73	1709.93	74-88	K.RKGVEGLIDIENPNR.V (3+)	4.61	0.15
					1850.60	1849.88	16-31	R.QYTSPEEIDAQLQAEK.Q (2+)	4.97	0.59
					1973.55	1972.88	34-52	K.AREEEEQKEGGDGAAGDPK.K (2+)	3.15	0.35
Splicing factor, arginine/serine -rich 9 (Q13242)	25527.4	8.74	6.4-6.1	33.94	1170.35	1170.547	155-163	R.KEDM'EYALR.K (2+)	3.101	0.362
					1009.20	1009.46	64-72	R.DAEDAIYGR.N (2+)	2.99	0.37
					1154.28	1154.55	49-58	R.KEDMEYALR.K (2+)	3.10	0.32
					1142.47	1154.55	145-154	K.DGVGM'VEYLR.K (2+)	2.95	0.38
					1138.27	1138.56	145-154	K.DGVGMVEYLR.K (2+)	3.63	0.42
					1299.95	1299.65	27-36	R.EKDLEDLFYK.Y (2+)	2.79	0.29
					1682.75	1682.912	113-128	R.VLVSGLPPSGSWQDLK.D (2+)	3.352	0.338
					1142.43	1142.648	49-58	R.HGLVPFAFVR.F (2+)	3.24	0.524
Histone H4 (P62805)	11230.3	11.36	6.4-6.1	64.71	714.15	714.3463	96-102	R.TLYGFGG.- (1+)	1.939	0.595
					989.31	989.5784	60-67	K.VFLENVIR.D (2+)	3.055	0.253
					1180.35	1180.621	46-55	R.ISGLIYEETR.G (2+)	3.996	0.482
					1290.47	1290.644	68-78	R.DAVTYTEHAKR.K (2+)	3.085	0.38
					1325.47	1325.754	24-35	R.DNIQGITKPAIR.R (2+)	3.626	0.428
					1326.39	1326.698	80-91	K.TVTAM'DVVYALK.R (2+)	3.395	0.38
					1336.55	1336.723	45-55	K.RISGLIYEETR.G (2+)	2.952	0.426
					1454.49	1454.793	79-91	R.KTVTAM'DVVYALK.R (2+)	4.544	0.534
					1482.67	1482.799	80-92	K.TVTAM'DVVYALKR.Q (2+)	3.322	0.555
					1610.73	1610.894	78-91	K.RKTVTAM'DVVYALK.R (3+)	3.753	0.381

					1610.73	1610.894	79-92	R.KTVTAM'DVVYALKR.Q (2+)	4.323	0.481
					1693.89	1694.008	21-35	K.VLRDNIQGITKPAIR.R (3+)	5.217	0.472
Isocitrate dehydrogenase (O75874)	46630.5	6.53	6.4-6.1	15.7	1087.39	1087.579	141-151	R.ATDFVVPGPGK.V (2+)	3.323	0.503
					1250.35	1250.613	250-260	R.LIDDM'VAQAMK.S (2+)	3.806	0.059
					1250.31	1250.613	250-260	R.LIDDMVAQAM'K.S (2+)	3.729	0.006
					1341.47	1341.676	302-314	K.TVEAEAAHGTVTR.H (2+)	3.772	0.485
					1438.40	1438.707	152-164	K.VEITYTPSDGTQK.V (2+)	3.492	0.459
					1695.60	1695.768	5-20	K.ISGGSVVEMQGDEMTR.I (2+)	5.205	0.583
					1711.47	1711.763	5-20	K.ISGGSVVEM'QGDEMTR.I (2+)	4.619	0.118
					1711.47	1711.763	5-20	K.ISGGSVVEMQGDEM'TR.I (2+)	4.073	0.591
					1839.57	1839.858	4-20	K.KISGGSVVEM'QGDEMTR.I (3+)	4.423	0.16
					1840.53	1839.858	4-20	K.KISGGSVVEMQGDEM'TR.I (3+)	3.714	0.422
26S protease regulatory subunit 8 (P62195)	45598.1	7.11	6.4-6.1	40.15	946.27	946.5209	214-222	R.VSGSELVQK.F (2+)	3.333	0.381
					1051.35	1051.509	394-402	K.DSEKHMSIK.K (2+)	2.943	0.319
					1105.48	1105.673	85-94	K.VLVKVHPEGK.F (2+)	2.855	0.296
					1158.85	1158.652	185-196	K.GVLLYGPPGTGK.T (2+)	3.516	0.465
					1285.55	1284.702	45-55	R.LQAQRNELNAK.V (2+)	3.245	0.315
					1290.43	1290.593	259-271	R.LEGGSGGDSEVQR.T (2+)	3.566	0.521
					1313.15	1313.732	28-38	K.IEELQLIVNDK.S (2+)	4.377	0.445
					1417.40	1416.76	114-125	R.VALRNDSYTLHK.I (3+)	4.748	0.294
					1440.60	1440.804	44-55	R.RLQAQRNELNAK.V (2+)	2.559	0.41
					1500.60	1500.81	346-360	R.KIAELMPGASGAEVK.G (2+)	5.061	0.5
					1516.60	1516.805	346-360	R.KIAELM'PGASGAEVK.G (2+)	4.786	0.528
					1524.47	1524.726	143-156	K.VPDSTYEMIGGLDK.Q (2+)	4.106	0.622
					1540.48	1540.721	143-156	K.VPDSTYEM'IGGLDK.Q (2+)	4.409	0.614
					1603.41	1603.779	376-389	R.VHVTQEDFEMAVAK.V (3+)	4.087	0.455
					1619.37	1619.774	376-389	R.VHVTQEDFEM'AVAK.V (2+)	4.388	0.552
					1759.85	1759.88	375-389	R.RVHVTQEDFEMAVAK.V (3+)	4.854	0.505
					1775.73	1775.875	375-389	R.RVHVTQEDFEM'AVAK.V (3+)	5.097	0.594
					1838.60	1838.921	272-287	R.TM'LELLNQLDGFEATK.N (2+)	5.556	0.625
					1910.72	1909.958	143-159	K.VPDSTYEM'IGGLDKQIK.E (3+)	3.963	0.464
Seryl-tRNA synthetase, mitochondrial precursor (Q9NP81)	58246.9	8.35	6.4-6.1	18.34	816.20	816.4579	111-118	K.AAVTEAVR.A (2+)	2.701	0.408
					1490.51	1490.749	50-62	R.EGYSALPQLDIER.F (2+)	3.347	0.379
					1531.60	1531.797	105-118	R.SLEEEKAAVTEAVR.A (2+)	4.949	0.502
					1712.82	1712.846	119-134	R.ALLANQDSGEVQQDPK.Y (2+)	5.357	0.48
					1873.04	1872.979	188-203	R.VLHM'VGDKPVFSFQPR.G (3+)	4.127	0.504
					1959.56	1959.952	170-187	K.LPNQTHPDVPVGDESQAR.V (2+)	5.128	0.545
					2001.81	2002.087	490-507	R.ITAPTHVPLQYIGPNQPR.K (3+)	5.087	0.427
Heat-shock protein beta-1 (HspB1) (P04792)	22769.5	5.98	6.1-5.8	59.02	917.24	917.4944	115-123	K.DGVVEITGK.H (2+)	2.815	0.458
					941.27	941.5056	189-198	R.AQLGGPEAAK.S (2+)	2.718	0.292
					987.51	987.6104	5-12	R.RVPFSLLR.G (2+)	2.674	0.208
					1075.28	1075.575	80-89	R.QLSSGVSEIR.H (2+)	2.544	0.32
					1146.39	1146.637	113-123	K.TKDGVVEITGK.H (2+)	3.587	0.527
					1155.35	1155.572	80-89	R.QLS#SGVSEIR.H (2+)	2.711	0.191
					1163.31	1163.621	28-37	R.LFDQAFGLPR.L (2+)	4.239	0.458
					1468.43	1468.74	115-127	K.DGVVEITGKHEER.Q (2+)	3.626	0.345
					1643.59	1643.824	189-205	R.AQLGGPEAAKSDETAAK.- (2+)	5.108	0.618
					1655.51	1655.753	124-136	K.HEERQDEHGYISR.C (2+)	4.669	0.484
					1697.73	1697.882	113-127	K.TKDGVVEITGKHEER.Q (3+)	4.065	0.478
					1783.71	1783.923	97-112	R.VSLDVNHFAPDELTVK.T (2+)	4.413	0.46
					1877.68	1878.013	57-75	R.PLPPAAIESPAVAAPAYSR.A (2+)	3.781	0.623
					1905.83	1905.992	172-188	K.LATQSNEITIPVTFESR.A (2+)	4.242	0.601
Myotrophin (V-1 protein) (P58546)	12756.6	5.28	6.1-5.8	44.44	900.23	900.4427	57-65	K.GADINAPDK.H (2+)	2.718	0.262
					1265.31	1265.613	19-29	K.DYVAKGEDVNR.T (2+)	2.831	0.403
					1465.85	1465.718	11-23	K.NGDLDEVKDYVAK.G (2+)	4.481	0.552
					1747.59	1747.85	97-113	K.GPDGLTAFEATDNQAIK.A (2+)	6.016	0.598
					2075.67	2076.061	94-113	K.TVKGPDGLTAFEATDNQAIK.A (3+)	3.904	0.292
					2636.63	2136.021	11-29	K.NGDLDEVKDYVAKGEDVNR.T (3+)	4.967	0.36
					2447.18	2447.242	90-113	K.GADKTVKGPDGLTAFEATDNQAIK.A (3+)	7.11	0.614
Growth factor receptor-bound protein 2 (P62993)	25191.4	5.89	6.1-5.8	27.65	918.30	918.5161	143-149	R.NQQIFLR.D (1+)	1.964	0.112
					928.36	928.5005	79-86	R.HDGAFLIR.E (2+)	2.992	0.475
					1006.32	1006.524	68-76	R.AKAEEMLSK.Q (2+)	2.956	0.255
					1022.36	1022.519	68-76	R.AKAEEM'LSK.Q (2+)	2.763	0.25
					1096.26	1096.516	11-20	K.ATADDELSFK.R (2+)	3.318	0.581
					1452.35	1452.686	87-100	R.ESESAPGDFSLSVK.F (2+)	3.775	0.505
					1507.35	1506.734	125-136	K.FNSLNELVDYHR.S (2+)	2.534	0.109
Superoxide dismutase [Cu-Zn] (P00441)	15795.9	5.7	6.1-5.8	35.29	1225.43	1225.618	80-91	R.HVGDLGNVTADK.D (2+)	3.822	0.576
					1420.43	1420.656	123-136	K.ADDLGKGGNEESTK.T (2+)	3.723	0.446
					1501.51	1501.765	10-23	K.GDGPVQGIINFEQK.E (2+)	3.927	0.465
					2063.67	2063.959	123-143	K.ADDLGKGGNEESTKTGNAGSR.L (3+)	4.706	0.416
					2227.05	2227.121	116-136	R.TLVVHEKADDLGKGGNEESTK.T (3+)	5.645	0.511
Stress-induced-phospho-protein 1 (P31948)	62600.4	6.4	5.8-5.5	22.28	800.35	800.4882	332-338	R.TPDVLKK.C (2+)	2.612	0.207
					854.15	854.4082	447-453	K.AMDVYQK.A (2+)	2.759	0.426
					926.31	926.4219	479-486	R.HDSPEDVK.R (2+)	2.901	0.351
					1004.27	1004.553	79-87	K.AAALEFLNR.F (2+)	2.968	0.445
					1116.43	1116.645	534-543	K.LM'DVGLIAIR.- (2+)	3.831	0.491

					1137.39	1137.59	514-523	K.DPQALSEHLK.H (2+)	3.031	0.387
					1488.51	1488.795	352-364	R.LAYIIPDLALEEK.H (2+)	3.887	0.238
					1870.76	1870.93	94-109	R.TYEEGLKHEAIIPQLK.E (3+)	3.909	0.525
					1888.05	1888.029	514-530	K.DPQALSEHLKIPVIAQK.I (3+)	4.071	0.558
					1889.51	1889.856	489-505	R.AMADPEVQQIMSDPAMR.L (2+)	5.236	0.541
					1905.59	1905.851	489-505	R.AMADPEVQQIM'SDPAMR.L (2+)	5.169	0.44
					1905.55	1905.851	489-505	R.AMADPEVQQIMSDPAM'R.L (2+)	4.813	0.403
					1905.55	1905.851	489-505	R.AM'ADPEVQQIMSDPAMR.L (2+)	5.53	0.526
					1913.61	1913.917	208-223	K.KETKPEPMEEDLPEHK.K (3+)	3.904	0.253
SH3-domain kinase binding protein 1 (Q96B97)	73082.7	6.24	5.8-5.5	25.86	1101.35	1101.616	290-299	K.EGDIVTLIHK.D (2+)	3.1	0.439
					1188.39	1188.611	248-259	K.LPATTATPDSSK.T (2+)	3.478	0.477
					1191.43	1191.622	536-546	K.TVTISQVSDIK.A (2+)	3.517	0.528
					1217.31	1217.656	651-660	R.LQMEVNDIKK.A (2+)	3.48	0.233
					1233.51	1233.651	6651-660	R.LQM'EVNDIKK.A (3+)	3.589	0.191
					1316.47	1316.706	247-259	K.KLPATTATPDSSK.T (2+)	3.403	0.51
					1320.47	1320.683	621-631	R.SIIETMKDQQK.R (2+)	3.06	0.359
					1336.47	1336.678	621-631	R.SIIETM'KDQQK.R (2+)	3.178	0.38
					1365.59	1365.76	408-420	K.TNSLSRPGALPPR.R (2+)	2.621	0.36
					1432.51	1432.744	585-598	R.AHSPSLFGTEGKPK.M (2+)	3.578	0.38
					1502.54	1502.837	649-660	R.LRLQM'EVNDIKK.A (3+)	3.638	0.287
					1588.89	1588.764	599-613	K.MEPAASSQAAVEELR.T (2+)	4.94	0.496
					1604.44	1604.759	599-613	K.M'EPAASSQAAVEELR.T (2+)	4.799	0.552
					1894.98	1894.944	274-289	K.VIFPYEAQHDDELTIK.E (2+)	3.08	0.492
					1907.66	1907.902	248-265	K.LPATTATPDSSKTEMDSR.T (3+)	3.522	0.449
					1941.63	1941.893	458-475	R.SNDIDLEGFDSVVSSTEK.L (2+)	6.088	0.576
					2035.71	2035.997	247-265	K.KLPATTATPDSSKTEMDSR.T (3+)	3.697	0.324
Deoxyuridine 5'-triphosphate nucleotidohydrolase precursor (P33316)	26690.7	9.65	5.8-5.5	17.86	968.27	968.4437	242-252	R.GSGGFGSTGKN.- (2+)	2.91	0.523
					1082.44	1082.559	119-128	R.LSEHATAPTR.G (2+)	2.864	0.426
					1284.52	1284.685	104-115	R.ARPAEVGGMQLR.F (2+)	3.583	0.342
					1300.41	1300.68	104-115	R.ARPAEVGGM'QLR.F (3+)	4.01	0.457
Nucleolar RNA helicase II (Q9NR30)	87303.4	9.32	5.8-5.5	12.52	1164.40	1164.711	260-270	R.APQVLVLAPTR.E (2+)	2.588	0.258
					1220.35	1220.704	210-215	R.GVTFLFPIQAK.T (2+)	4.057	0.473
					1224.27	1224.542	7-18	R.SDAGLESDTAMK.K (2+)	3.895	0.522
					1240.31	1240.537	7-18	R.SDAGLESDTAM'K.K (2+)	3.998	0.376
					1377.67	1377.833	258-270	R.GRAPQVLVLAPTR.E (2+)	3.851	0.496
					11431.44	1431.679	61-73	K.KAEPSEVDMIISPK.S (2+)	4.279	0.476
					1447.47	1447.674	61-73	K.KAEPSEVDM'IISPK.S (2+)	3.72	0.512
					1492.55	1493.727	5-18	K.LRSDAGLESDTAMK.K (2+)	3.203	0.521
					1137.39	1137.59	514-523	K.DPQALSEHLK.H (2+)	3.031	0.387
					1488.51	1488.795	352-364	R.LAYIIPDLALEEK.H (2+)	3.887	0.238
					1870.76	1870.93	94-109	R.TYEEGLKHEAIIPQLK.E (3+)	3.909	0.525
					1888.05	1888.029	514-530	K.DPQALSEHLKIPVIAQK.I (3+)	4.071	0.558
					1889.51	1889.856	489-505	R.AMADPEVQQIMSDPAMR.L (2+)	5.236	0.541
					1905.59	1905.851	489-505	R.AMADPEVQQIM'SDPAMR.L (2+)	5.169	0.44
					1905.55	1905.851	489-505	R.AMADPEVQQIMSDPAM'R.L (2+)	4.813	0.403
					1905.55	1905.851	489-505	R.AM'ADPEVQQIMSDPAMR.L (2+)	5.53	0.526
					1913.61	1913.917	208-223	K.KETKPEPMEEDLPEHK.K (3+)	3.904	0.253
SH3-domain kinase binding protein 1 (Q96B97)	73082.7	6.24	5.8-5.5	25.86	1101.35	1101.616	290-299	K.EGDIVTLIHK.D (2+)	3.1	0.439
					1188.39	1188.611	248-259	K.LPATTATPDSSK.T (2+)	3.478	0.477
					1191.43	1191.622	536-546	K.TVTISQVSDIK.A (2+)	3.517	0.528
					1217.31	1217.656	651-660	R.LQMEVNDIKK.A (2+)	3.48	0.233
					1233.51	1233.651	6651-660	R.LQM'EVNDIKK.A (3+)	3.589	0.191
					1316.47	1316.706	247-259	K.KLPATTATPDSSK.T (2+)	3.403	0.51
					1320.47	1320.683	621-631	R.SIIETMKDQQK.R (2+)	3.06	0.359
					1336.47	1336.678	621-631	R.SIIETM'KDQQK.R (2+)	3.178	0.38
					1365.59	1365.76	408-420	K.TNSLSRPGALPPR.R (2+)	2.621	0.36
					1432.51	1432.744	585-598	R.AHSPSLFGTEGKPK.M (2+)	3.578	0.38
					1502.54	1502.837	649-660	R.LRLQM'EVNDIKK.A (3+)	3.638	0.287
					1588.89	1588.764	599-613	K.MEPAASSQAAVEELR.T (2+)	4.94	0.496
					1604.44	1604.759	599-613	K.M'EPAASSQAAVEELR.T (2+)	4.799	0.552
					1894.98	1894.944	274-289	K.VIFPYEAQHDDELTIK.E (2+)	3.08	0.492
					1907.66	1907.902	248-265	K.LPATTATPDSSKTEMDSR.T (3+)	3.522	0.449
					1941.63	1941.893	458-475	R.SNDIDLEGFDSVVSSTEK.L (2+)	6.088	0.576
					2035.71	2035.997	247-265	K.KLPATTATPDSSKTEMDSR.T (3+)	3.697	0.324
Deoxyuridine 5'-triphosphate nucleotidohydrolase precursor (P33316)	26690.7	9.65	5.8-5.5	17.86	968.27	968.4437	242-252	R.GSGGFGSTGKN.- (2+)	2.91	0.523
					1082.44	1082.559	119-128	R.LSEHATAPTR.G (2+)	2.864	0.426
					1284.52	1284.685	104-115	R.ARPAEVGGMQLR.F (2+)	3.583	0.342
					1300.41	1300.68	104-115	R.ARPAEVGGM'QLR.F (3+)	4.01	0.457
Nucleolar RNA helicase II (Q9NR30)	87303.4	9.32	5.8-5.5	12.52	1164.40	1164.711	260-270	R.APQVLVLAPTR.E (2+)	2.588	0.258
					1220.35	1220.704	210-215	R.GVTFLFPIQAK.T (2+)	4.057	0.473
					1224.27	1224.542	7-18	R.SDAGLESDTAMK.K (2+)	3.895	0.522
					1240.31	1240.537	7-18	R.SDAGLESDTAM'K.K (2+)	3.998	0.376
					1377.67	1377.833	258-270	R.GRAPQVLVLAPTR.E (2+)	3.851	0.496
					11431.44	1431.679	61-73	K.KAEPSEVDMIISPK.S (2+)	4.279	0.476
					1447.47	1447.674	61-73	K.KAEPSEVDM'IISPK.S (2+)	3.72	0.512
					1492.55	1493.727	5-18	K.LRSDAGLESDTAMK.K (2+)	3.203	0.521

					1499.56	1499.771	79-91	K.KKEEPSQNDISPK.T (2+)	4.124	0.333
					1575.39	1575.769	60-73	K.KKAEPSEVDM*ISPK.S (3+)	3.573	0.316
					1668.47	1668.812	185-199	K.EGAFSIIFPISEETIK.L (2+)	4.795	0.448
					1711.51	1711.803	117-131	K.IIEEPSEEEIDAPKPK.K (2+)	3.849	0.27
					2039.67	2040.014	114-131	K.VTKIIEEPSEEEIDAPKPK.K (3+)	6.19	0.455
Lamin A/C (70 kDa lamin) (P02545)	74095.7	6.57	5.8-5.5	**17.47**	901.31	901.4995	172-180	K.LEAALGEAK.K (2+)	2.821	0.344
					1029.36	1028.574	241-249	R.LADALQELR.A (2+)	3.514	0.306
					1043.91	1043.549	124-133	K.EGDLIAAQAR.L (2+)	3.241	0.458
					1148.27	1148.58	63-72	R.ITESEEVVSR.E (2+)	3.624	0.532
					1165.24	1165.549	79-89	K.AAYEAELGDAR.K (2+)	2.517	0.16
					1187.48	1187.638	320-329	K.LRDLEDSLAR.E (2+)	2.742	0.145
					1275.40	1275.691	145-156	K.EAALSTALSEKR.T (2+)	3.418	0.488
					1349.43	1359.687	12-25	R.SGAQASSTPLSPTR.I (2+)	3.655	0.512
					1514.60	1514.829	120-133	R.IITKKEGDLIAAQAR.L (3+)	3.65	0.344
					1566.39	1566.751	628-644	R.SVGGSGGGSFGDNLVTR.S (2+)	4.367	0.601
Calcyclin (P06703)	10174.4	5.32	5.8-5.5	**22.22**	915.31	915.49	48-55	K.LQDAEIAR.L (2+)	2.72	0.343
					1358.55	1358.79	36-47	K.ELIQKELTIGSK.L (2+)	3.41	0.391
DNA-binding protein A (P16989)	40037	9.77	5.5-5.2	**17.2**	1783.01	1782.862	151-169	K.GAEAANVTGPDGVPVEGSR.Y (2+)	5.199	0.511
					1881.55	1881.847	354-372	K.AGEAPTENPAPPTQQSSAE.- (2+)	3.883	0.602
					2677.11	2677.282	301-326	R.PAPAVGEAEDKENQQATSGPNQPSVR.R (3+)	6.284	0.639
Annexin A2 (P07355)	38449.8	7.56	5.5-5.2	**31.36**	731.20	731.4303	81-87	K.ELASALK.S (1+)	2.047	0.301
					745.26	745.4646	169-175	K.LMVALAK.G (1+)	2.14	0.285
					1222.43	1222.596	104-114	K.TPAQYDASELK.A (2+)	4.456	0.543
					1225.23	1225.57	157-167	K.DIISDTSGDFR.K (2+)	3.591	0.523
					1244.42	1244.624	135-144	R.TIQELQEIIR.V (2+)	3.799	0.367
					1353.44	1353.665	157-168	K.DIISDTSGDFRK.L (2+)	3.314	0.475
					1421.35	1421.695	313-323	K.SLYYYIQQDTK.G (2+)	3.879	0.49
					1460.35	1460.673	233-244	K.SYSPYDMLESIR.K (2+)	3.576	0.548
					1542.55	1542.849	49-62	K.GVDEVTIVNILTNR.S (2+)	4.468	0.43
					1908.56	1908.883	179-195	R.AEDGSVIDYELIDQDAR.D (2+)	5.18	0.51
					1939.76	1939.961	152-168	K.TDLEKDIISDTSGDFRK.L (3+)	5.38	0.528
Phenylalanyl-tRNA synthetase alpha chain (Q9Y285)	57397.5	7.46	5.5-5.2	**6.9**	984.28	984.5002	177-186	K.GSAFSTSISK.Q (2+)	2.669	0.495
					1306.31	1306.595	127-137	R.VVDSMEDEVQR.R (2+)	3.967	0.425
					1322.35	1322.59	127-137	R.VVDSM*EDEVQR.R (2+)	3.518	0.524
					1491.48	1491.744	494-507	R.LDAEPRPPPTQEAA.- (2+)	3.474	0.577
Alpha enolase (P06733)	47009.3	6.99	5.5-5.2	**9.47**	904.24	904.46	412-419	R.IEEELGSK.A (2+)	2.69	0.04
					959.39	959.54	426-433	R.IIFRIIPLAK.- (2+)	2.68	0.21
					1088.47	1088.60	80-88	K.KLIIVTEQEKJ (2+)	2.82	0.25
					1316.47	1316.71	81-91	K.LIIVTEQEKIDK.L (2+)	3.35	0.28
					1406.39	1406.72	15-27	R.GIIPTVEVDLFTSK.G (2+)	3.59	0.28
					1444.64	1444.80	80-91	K.KLIIVTEQEKIDK.L (2+)	4.51	0.48
Eukaryotic translation initiation factor 3 subunit 4 (O75821)	35589.9	5.87	4.9-4.6	**32.19**	1031.43	1031.523	234-242	R.RADDIATIR.V (2+)	3.487	0.291
					1034.35	1034.512	243-251	R.VTIILSEDTR.E (2+)	3.129	0.33
					1074.39	1074.511	225-233	R.RGESMQPIIR.R (2+)	3.142	0.296
					1081.39	1081.558	281-289	K.GFAFISFHR.R (2+)	3.074	0.365
					1374.50	1374.712	181-193	K.ELAEQLGLSTGEK.E (2+)	3.784	0.519
					1523.55	1523.807	196-209	K.LPGELEPVQATQIK.T (2+)	4.3	0.464
					1525.51	1525.775	66-78	K.TVTEYKIDEDGKK.F (2+)	4.08	0.469
					1631.63	1631.849	181-195	K.ELAEQLGLSTGEKEK.L (2+)	4.632	0.395
					1780.71	1780.945	194-209	K.EKLPGELEPVQATQIK.T (2+)	5.11	0.539
					2464.12	2464.334	33-57	K.GIPLATGDTSPEPELLPGAPLPPPK.E (3+)	4.895	0.408

Table 3.3: List of proteins showing higher humoral response in cancer sera. Proteins were identified using LS ESI-MS/MS.

Protein Name	Acc#	Fraction pH	%Cov	MW
Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	> 7.9	16.43	37407.7
Macrophage migration inhibitory factor	P14174	> 7.9	9.65	12338.2
10 kDa heat shock protein, mitochondrial (Hsp10)	P61604	> 7.9	66.34	10794.8
Heterogeneous nuclear ribonucleoprotein A1	P09651	> 7.9	23.18	38692.1
Profilin-1	P07737	> 7.9	19.42	14914.5
Nuclease sensitive element binding protein 1	P67809	7.9-7.6	30.56	35903.7
60S ribosomal protein L13	P26373	7.6-7.3	34.76	24116.5
60S ribosomal protein L27	P61353	7.6-7.3	34.07	15657.7
60S ribosomal protein L21	P46778	7.6-7.3	40.88	18423
DNA-binding protein A (Cold shock domain protein A)	P16989	7.3-7.0	8.87	40037
Calgranulin B	P06702	7.3-7.0	44.74	13234.5
Histone H1.5 (Histone H1a)	P16401	7.3-7.0	15.11	22436.4
Cytochrome c oxidase polypeptide Vb, mitochondrial precursor	P10606	7.3-7.0	29.46	13688
Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	7.3-7.0	22.95	37407.7
Phosphoglycerate kinase 1	P00558	7.0-6.7	64.42	44456.1
Cyclin-dependent kinase inhibitor 1B	P46527	7.0-6.7	28.79	22060.7
28 kDa heat- and acid-stable phosphoprotein (PDGF-associated protein)	Q13442	6.4-6.1	44.75	20618.6
Splicing factor, arginine/serine-rich 9	Q13242	6.4-6.1	33.94	25527.4
Histone H4	P62805	6.4-6.1	64.71	11230.3
Isocitrate dehydrogenase	O75874	6.4-6.1	15.7	46630.5
26S protease regulatory subunit 8	P62195	6.4-6.1	40.15	45598.1
Seryl-tRNA synthetase, mitochondrial precursor	Q9NP81	6.4-6.1	18.34	58246.9
Heat-shock protein beta-1 (HspB1)	P04792	6.1-5.8	59.02	22769.5
Myotrophin (V-1 protein)	P58546	6.1-5.8	44.44	12756.6
Growth factor receptor-bound protein 2	P62993	6.1-5.8	27.65	25191.4
Superoxide dismutase [Cu-Zn]	P00441	6.1-5.8	35.29	15795.9
Stress-induced-phosphoprotein 1	P31948	5.8-5.5	22.28	62600.4
SH3-domain kinase binding protein 1	Q96B97	5.8-5.5	25.86	73082.7
Deoxyuridine 5'-triphosphate nucleotidohydrolase precursor	P33316	5.8-5.5	17.86	26690.7
Nucleolar RNA helicase II	Q9NR30	5.8-5.5	12.52	87303.4
Lamin A/C (70 kDa lamin)	P02545	5.8-5.5	17.47	74095.7
Calcyclin	P06703	5.8-5.5	22.22	10174.4
DNA-binding protein A	P16989	5.5-5.2	17.2	40037
Annexin A2	P07355	5.5-5.2	31.36	38449.8
Alpha enolase	P06733	5.5-5.2	9.47	47009.3
Phenylalanyl-tRNA synthetase alpha chain	Q9Y285	5.5-5.2	6.9	57397.5
Eukaryotic translation initiation factor 3 subunit 4	O75821	4.9-4.6	32.19	35589.9

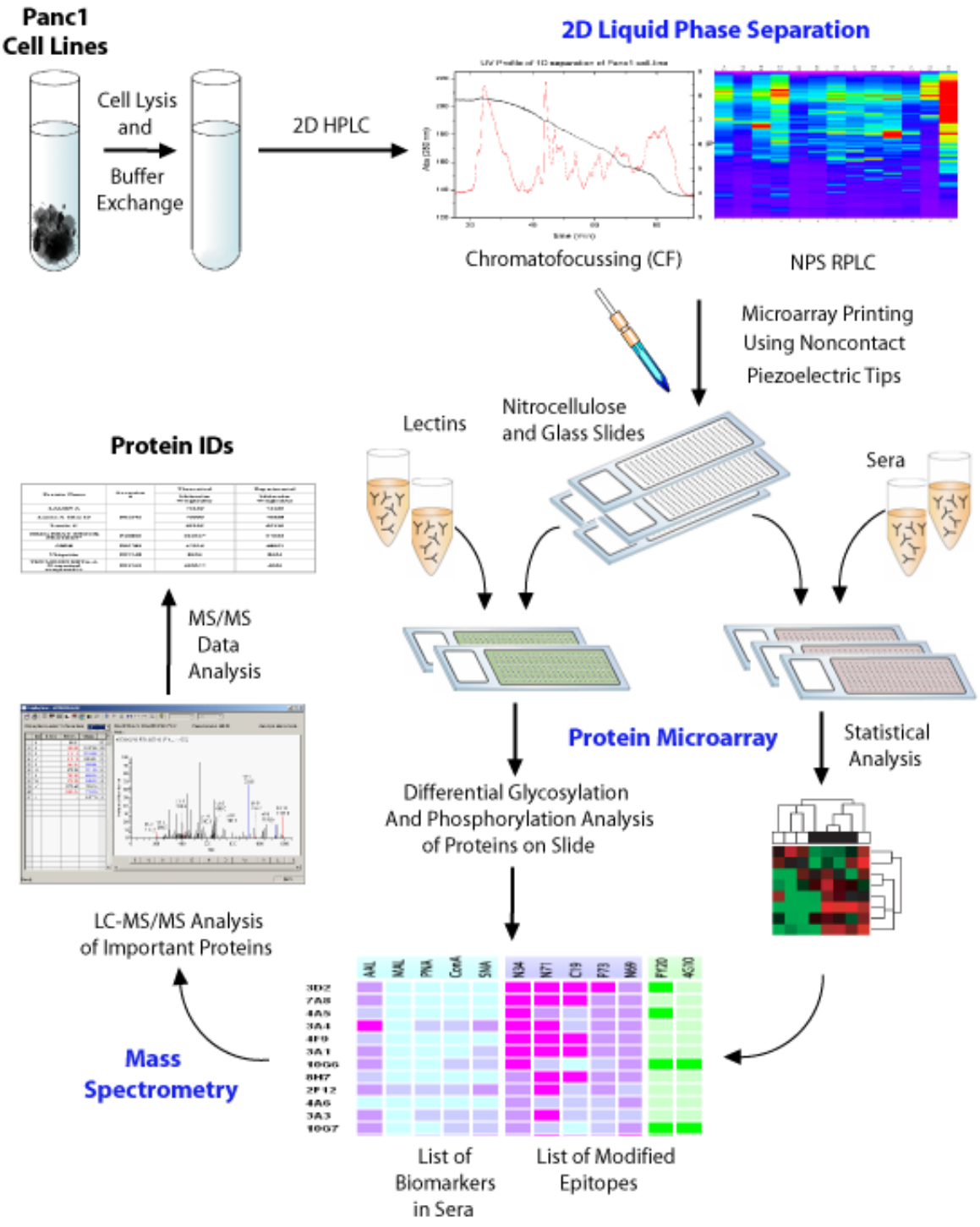

Figure 3.1: Analytical work-flow of the experiment.

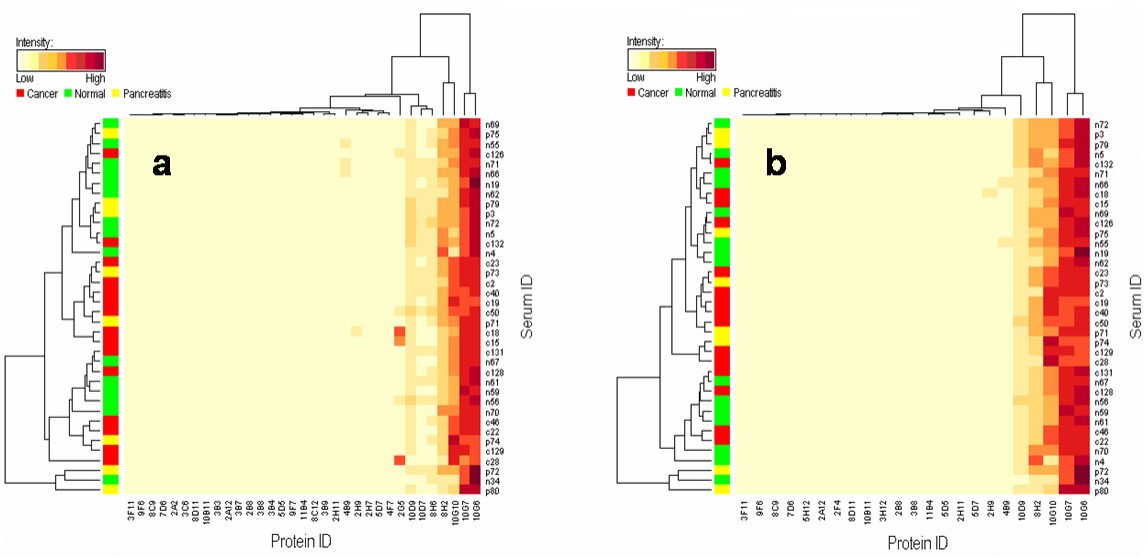

Figure 3.2: Heat maps depicting the list of proteins for distinguishing cancer and normal tissue. Analysis conducted after signature proteins distinguishing chronic pancreatitis and normal were removed (a) and included (b).

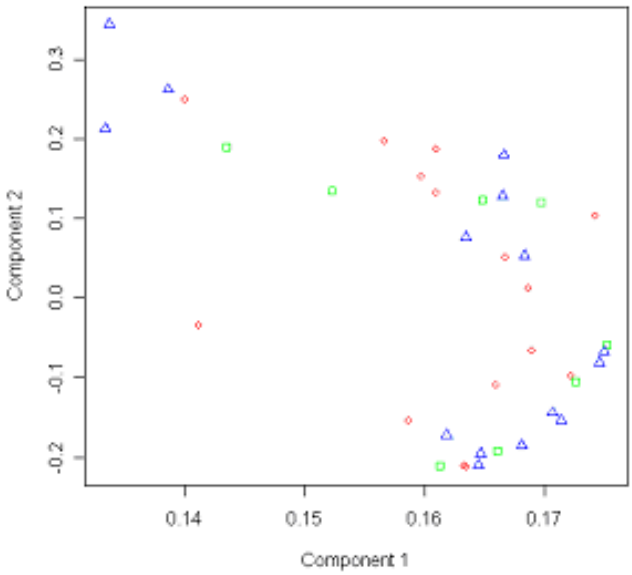

Figure 3.3: Plot of the first two components from PCA ($\circ$ = cancer, $\triangle$ = normal and $\square$ = chronic pancreatitis.

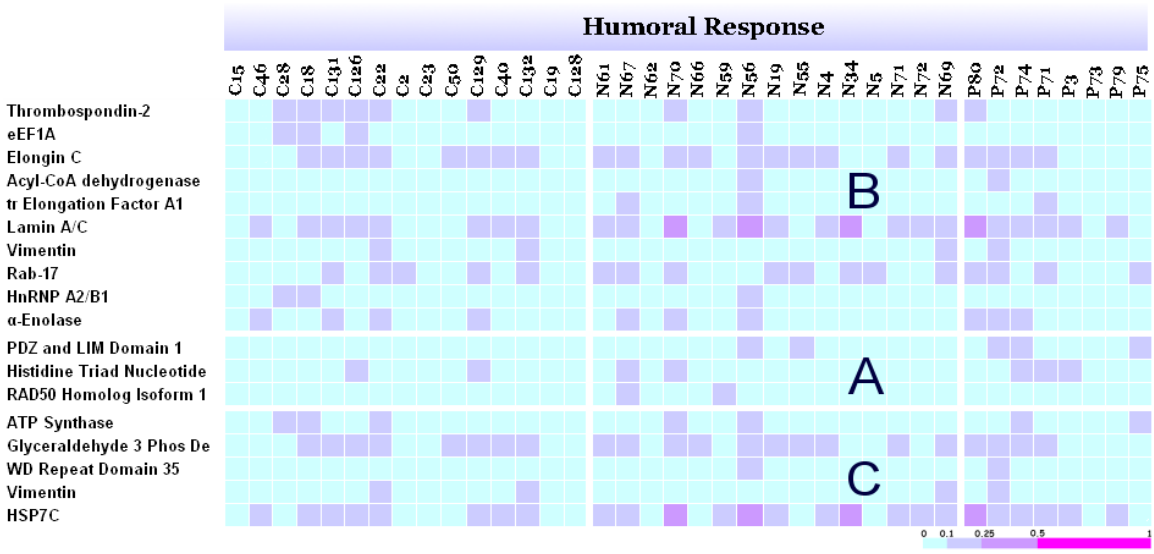

Figure 3.4: Response map showing humoral response and modifications on the identified Panc-1 proteins.

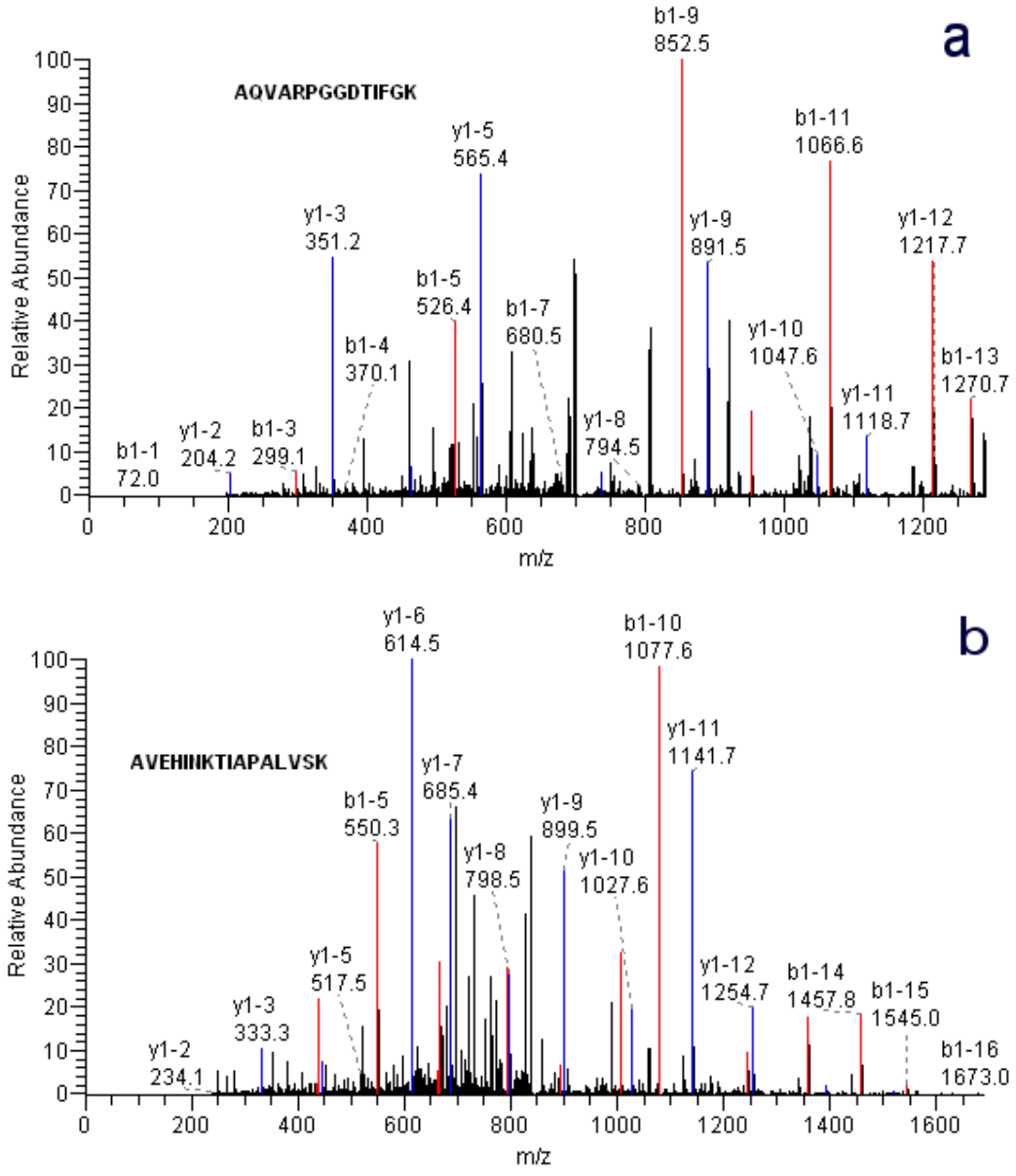

Figure 3.5: nESI-LC-MS/MS spectrum of (a) peptide AQVARPGGDTIFGK from Histidine Triad Nucleotide protein and (b) peptide AVEHINKTIAPALVSK from α-Enolase.

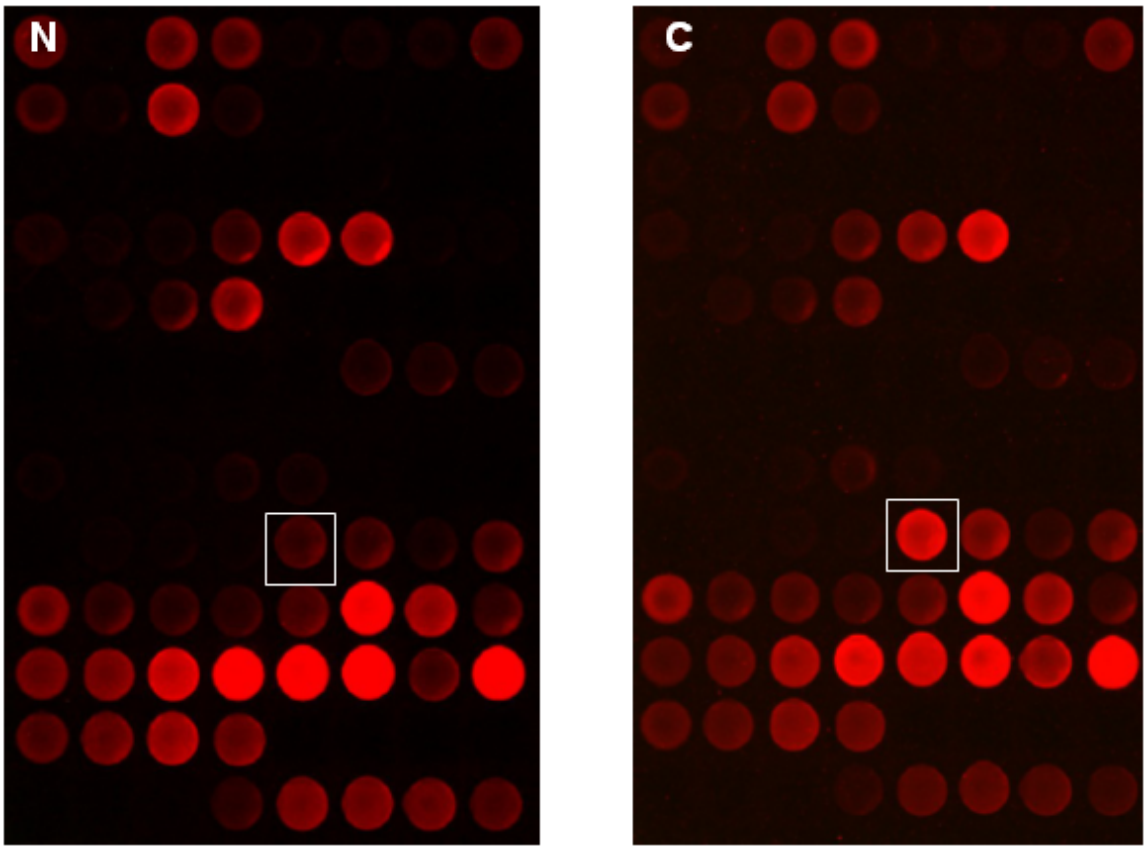

Figure 3.6: Slide image processed with normal serum (N) and with cancer serum (C) showing humoral response to the protein Thrombospondin.

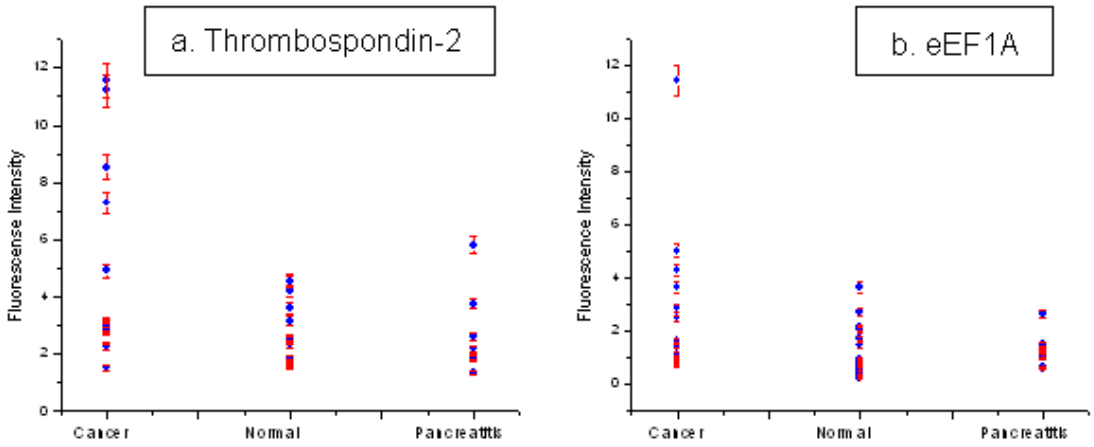

Figure 3.7: Spot fluorescence intensity for humoral response against (a) Thrombospondin-2 and (b) Elongation Factor 1A from all tested serum samples in the three groups.

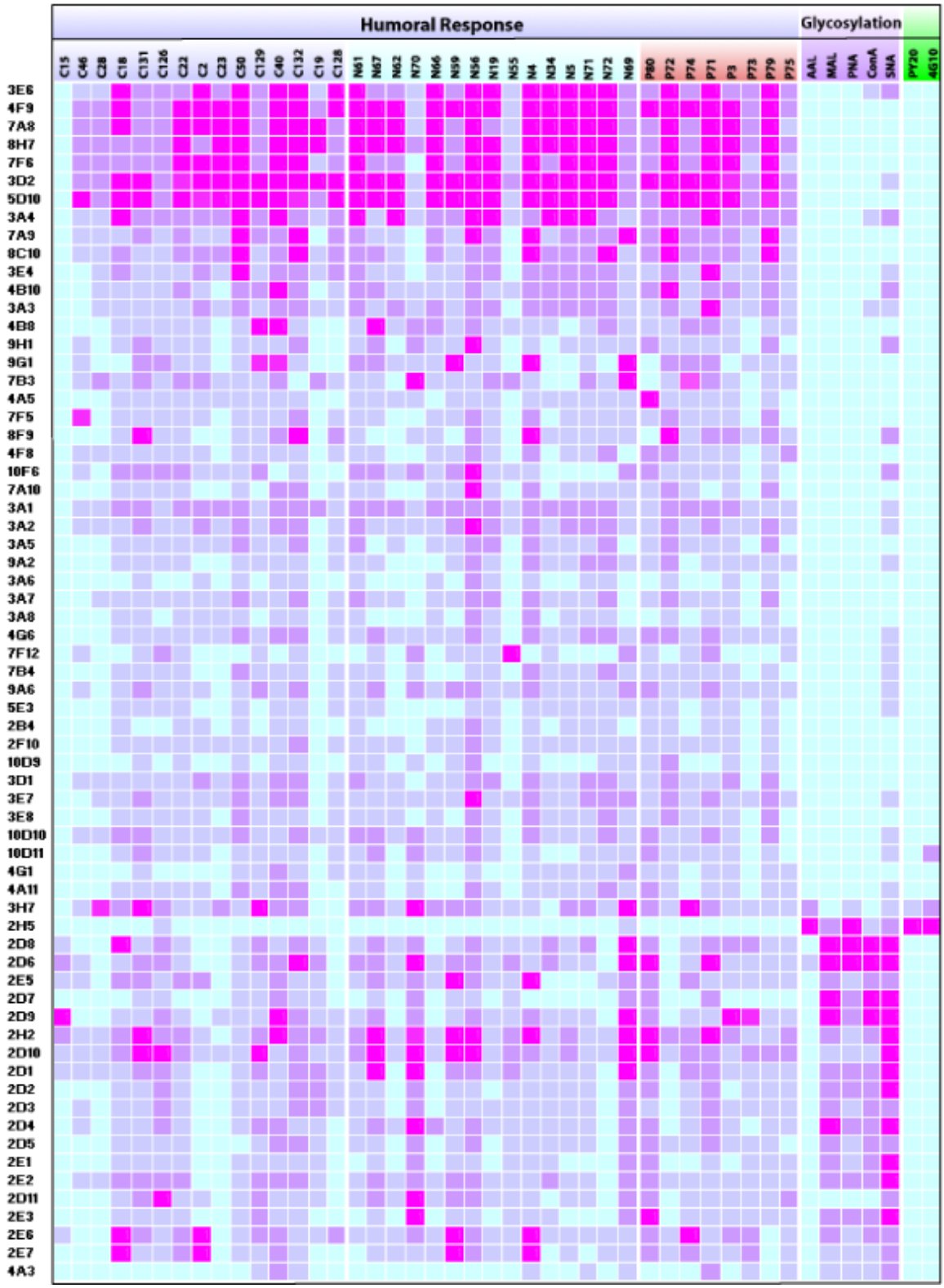

Figure 3.8: Comparison Map between Humoral Response and Modification among Panc-1 Proteins.

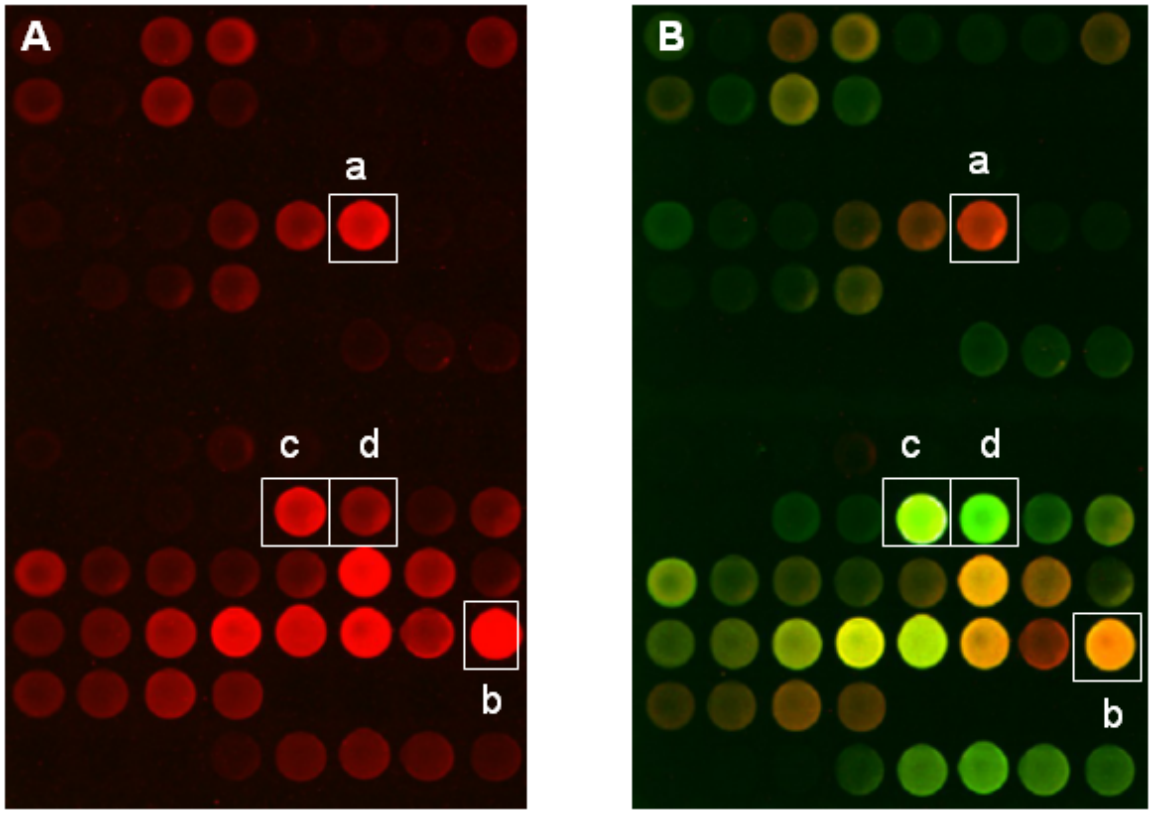

Figure 3.9: Slides processed with cancer sera (A) and SNA lectin (B) clearly distinguishes proteins which are not glycosylated (a) and glycosylated (c, d). The picture on the right is obtained by superimposing the picture on the left on top.

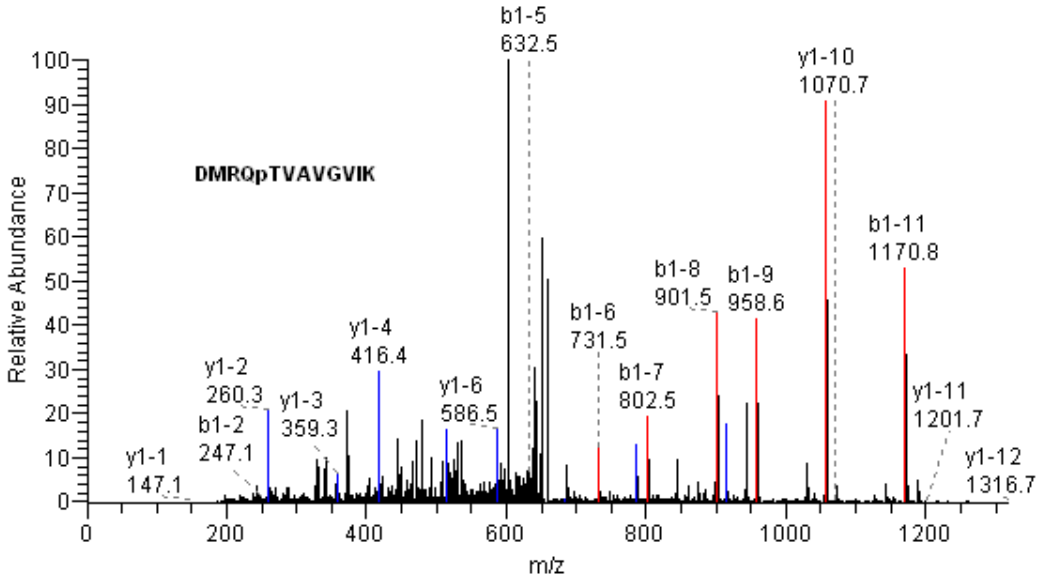

Figure 3.10: nESI-LC-MS/MS spectrum of phosphorylated peptide DMRQpTVAVGVIK from Elongation Factor 1A (eEF1A).

References

[1] Jemal A.; Siegel R.; Ward E.; Murray T.; Xu J.; Smigal C. and Thun M. J. *CA Cancer J. Clin.*, 56:106–130, 2006.

[2] Mann D. V.; Edwards R.; Ho S.; Lau W. Y. and Glazer G. *Eur. J. Surg. Oncol.*, 26:474–479, 2000.

[3] Ferrone C. R.; Finkelstein D. M.; Thayer S. P.; Muzikansky A.; Fernandez-del Castillo C. and Warshaw A. L. *J. Clin. Oncol.*, 24:2897–2902, 2006.

[4] Duffy M. J. *Ann. Clin. Biochem.*, 35:364–370, 1998.

[5] Boeck S.; Stieber P.; Holdenrieder S.; Wilkowski R. and Heinemann V. *Oncology*, 70:255–264, 2006.

[6] Ujiki M. B. and Talamonti M. S. *Semin. Radiation Oncol.*, 15:218–225, 2005.

[7] Fahy B. N.; Schlieman M. G.; Virudachalam S. and Bold R. *British Journal of Cancer*, 89:391–397, 2003.

[8] Ren Y. X.; Xu G. M.; Li Z. S. and Song Y. G. *World J. Gastroenterol*, 15:881–884, 2004.

[9] Almoguera C.; Shibata D.; Forrester K.; Martin J.; Arnheim N. and Perucho M. *Cell*, 53:549–554, 1998.

[10] Dabritz J.; Hanfler J.; Preston R.; Stieler J. and Oettle H. *British Journal of Cancer*, 92:405–412, 2005.

[11] Friess H.; Kleeff J.; Gumbs A. and Buchler M. W. *Digestion*, 58:557–63, 1997.

[12] Compton S. J.; Sandhu S.; Wijesuriya S. J. and Hollenberg M. D. *Biochemical Journal*, 368:495–505, 2002.

[13] Templin M. F.; Stoll D.; Schwenk J. M.; Potz O.; Kramer S. and Joos T. O. *Proteomics*, 3:2155–2166, 2003.

[14] Pal M.; Moffa A.; Sreekumar A.; Ethier S. P.; Barder T. J.; Chinnaiyan A. and Lubman D. M. *Anal. Chem.*, 78:702–710, 2006.

[15] Yan F.; Sreekumar A.; Laxman B.; Chinnaiyan A. M.; Lubman D. M. and Barder T. J. *Proteomics*, 3:1228–1235, 2003.

[16] Orchekowski R.; Hamelinck D.; Li L.; Gliwa E.; vanBrocklin M.; Marrero J. A.; Vande Woude G. F.; Feng Z.; Brand R. and Haab B. B. *Cancer Res.*, 65:11193–11202, 2005.

[17] Madoz-Gurpide J.; Wang H.; Misek D. E.; Brichory F. and Hanash S. M. *Proteomics*, 1:1279–1287, 2001.

[18] Nam M. J.; Madoz-Gurpide J.; Wang H.; Lescure P.; Schmalbach C. E.; Zhao R.; Misek D. E.; Kuick R.; Brenner D. E. and Hanash S. M. *Proteomics*, 3:2108–2115, 2003.

[19] Bouwman K.; Qiu J.; Zhou H.; Schotanus M.; Mangold L. A.; Vogt R.; Erlandson E.; Trenkle J.; Partin A. W.; Misek D. E.; Omenn G. S.; Haab B. B. and Hanash S. *Proteomics*, 3:2200–2207, 2003.

[20] Qiu J.; Madoz-Gurpide J.; Misek D. E.; Kuick R.; Brenner D. E.; Michailidis G.; Haab B. B.; Omenn G. S. and Hanash S. M. *J. Proteome Res.*, 3:261–267, 2004.

[21] Gao W. M.; Kuick R.; Orchekowski R. P.; Misek D. E; Qiu J.; Greenberg A. K.; Rom W. N.; Brenner D. E.; Omenn G. S.; Haab B. B. and Hanash S. M. *BMC Cancer.*, 5:110, 2005.

[22] Patwa T. H.; Zhao J.; Anderson M. A.; Simeone D. M. and Lubman D. M. *Anal. Chem.*, 78:6411–6421, 2006.

[23] Zhao J.; Patwa T. H.; Qiu W.; Shedden K.; Hinderer R.; Misek D. E.; Anderson M. A.; Simeone D. M. and Lubman D. M. *J. Proteome Research*, In Press.

[24] Keller A.; Nesvizhskii A. I.; Kolker E. and Aebersold R. *Anal. Chem.*, 74:5383–5392, 2002.

[25] Nesvizhskii A. I.; Keller A.; Kolker E. and Aebersold R. *Anal. Chem.*, 75:4646–4658, 2003.

[26] Madden M. E. and Sarras M. P. *Pancreas*, 3:512–528, 1988.

[27] Jelen F. O. A.; Smietana K. and Otlewski J. *Acta Biochimica Polonica*, 50:985–1017, 2003.

[28] Fanning A. S. and Anderson J. M. *Journal of Clinical Investigation*, 103:767–72, 1999.

[29] Vallenius T.; Luukko K. and Makela T. P. *J. Biol. Chem.*, 275:11100–11105, 2000.

[30] Su T.; Suzui M.; Wang L.; Lin C. S.; Xing W. Q. and Weinstein I. B. *Proc. Natl. Acad. Sci. USA*, 100:7824–7829, 2003.

[31] Heikkinen K.; Rapakko K.; Karppinen S. M.; Erkko H.; Knuutila S.; Lundan T.; Mannermaa A.; Borresen-Dale A. L.; Borg A.; Barkardottir R. B.; Petrini J. and Winqvist R. *Carcinogenesis*, 8:1593–1599, 2006.

[32] Shen J.; Person M. D.; Zhu J.; Abbruzzese J. and Li D. *Cancer Research*, 64:9018–9026, 2004.

[33] Jungert K.; Buck A.; von Wichert G.; Adler G.; Konig A.; Buchholz M.; Gress T. M. and Ellenrieder V. *Cancer Research*, 67:1563–1570, 2007.

[34] Hong S. H.; Misek D. E.; Wang H.; Puravs E.; Hinderer R.; Giordano T. J.; Greenson J. K.; Brenner D. E.; Simeone D. M.; Logsdon C. D. and Hanash S. M. *Biomarker Insights*, 2:175–183, 2006.

[35] Ditzel H. J.; Masaki Y.; Nielsen H.; Farnaes L. and Burton D. R. *Proc. Natl. Acad. Sci. USA*, 97:9234–9239, 2000.

[36] Porkka K.; Saramaki O.; Tanner M. and Visakorpi T. *Laboratory Investigation*, 82:629–637, 2002.

CHAPTER IV

Humoral Response Profiling Reveals Pathways to Prostate Cancer Progression

4.1 Introduction

Prostate carcinoma is the leading cancer diagnosis in American men where its early detection facilitates effective treatment modalities and improved mortality [1]. Although the advent of prostate specific antigen (PSA) screening has led to the detection of prostate cancer [2] much earlier, its lack of specificity for neoplasm has resulted in an increase in the number of subsequent prostate needle biopsies [3]. As the population of men 65 years and older is expected to increase from 14 million in the year 2000 to 31 million by 2030 [4] in the US, it will be increasingly important to distinguish men with benign prostatic hyperplasia from those having neoplastic disease requiring clinical intervention. There is a compelling need to define additional clinical markers for accurate detection of prostate cancers.

The scarcity of clinical markers has spawned a wide array of serum-based early detection methodologies leveraging protein microarrays among other platforms [5]. But this approach is complicated by the fact that potentially viable tumor biomarkers are embedded among a huge array of proteomic noise. This noise includes housekeeping and highly abundant proteins so the comparative low abundance of biomarker

candidates confounds their detection. Thus, effective detection methodologies must span as many as 10 orders of magnitude in dynamic range to reliably detect markers in complex biofluids like plasma or serum [6]. No existing technology or platform offers such a broad dynamic range without implementing pre-fractionation strategies. This however may result in the loss or suppression of important biomarkers as many high-abundant proteins removed by depletion act as carriers for low-abundant proteins.

In this work we utilize the immune system-driven amplification of the autoantibody response to intracellular antigens which promises higher sensitivity, specificity, predictive value and reproducibility in detecting low-abundant tumor markers [7]. Early efforts have identified many gene products eliciting the humoral response. Somatic alterations in p53 are observed in 30-40% of affected patients, which has been shown to be able to predate cancer diagnosis [8]. In other work, 60% of patients with lung adenocarcinomas exhibited humoral response to glycosylated annexins I and/or II, whereas none of the sera from non-cancer patients demonstrated such a response [9]. Similarly, autoantibodies to the prostasome and to such antigens as PSA, prostatic acid phosphatase (PAP), HER-2/neu, p53, alpha methylacyl-CoA racemase (AMACR) and GRP78 have been observed in the sera of prostate cancer patients [10–13]. Autoantibody signatures have also been identified using phage microarrays that can delineate prostate cancer patients from control individuals with >90% accuracy [14]. However, one of the major disadvantages of this platform lies in the fact that most of the humoral response targets identified may be mimitopes that resemble the target protein in either the amino acid sequence or structure [14]. Further, it is important to note that most of the proteins that elicit humoral response

are often differentiation antigens or antigens that are over-expressed or modified in cancer [12, 13]. Additionally, the humoral response elicited by cancers is heterogeneous. This is supported by studies from humoral response trials where among the large numbers of patients tested, only a subset of patients with a specific tumor type develop a response to a specific antigen. This heterogeneity in humoral response necessitates the use of a multiplex panel of protein targets as autoantibody biomarkers to be able to detect tumors with broad coverage for a large number of people. This motivates our strategy of coupling comprehensive two-dimensional liquid-phase fractionation of the prostate cancer tissue proteome to protein microarray analysis of patient sera. Mass spectrometry is used for the identification of proteins eliciting humoral response to prostate cancer antigens (Figure 4.1). In addition to using the humoral response signature for prostate cancer detection, we have attempted to highlight the power of autoantibody screening in delineating proteomic alterations and proposing altered pathways during prostate cancer development.

4.2 Experimental

4.2.1 Patient Population and Sample Selection

The study was approved by the Institutional Review Board (IRB) of the University of Michigan Medical School. Serum samples from patients who visited the Urology Clinic for prostate cancer screening were collected before biopsy. The sera were then stored at the University of Michigan Specialized Research Program in Prostate Cancer (SPORE) tissue/serum bank. A total of 34 serum samples from patients who visited the clinic on two successive days were sequentially used for the experiments. 18 patients were biopsy-positive for prostate cancer (PCa) and 16 were negative for neoplasm. The average age of all prostate cancer patients was 63.2 ± 12.8 years. For

patients who tested negative for neoplasm and were diagnosed with benign prostatic hyperplasia (BPH), the average age was 64.8 ± 10.7 years. PSA values for the PCa and BPH groups were 7.81 ± 5.34 and 6.79 ± 3.76 ng/mL respectively. Detailed clinical and pathology data for this study is available in Table 4.1.

4.2.2 Preparation of Reference Pools

Tissue samples obtained after surgery from clinically localized prostate cancer patients (n=5) and advanced prostate cancer patients (n=5) were used for 2D liquid phase fractionation as described below. All chemicals were purchased from Sigma (St Louis, MO) unless otherwise stated. For protein extraction, the tissue samples were re-suspended in lysis buffer consisting of 7 M urea, 2 M thiourea, 100 mM DTT, 0.5% biolyte ampholyte 3-10 (Pharmacia, Pitscataway, NJ), 2% OG and 1 mM PMSF. Samples were lysed at room temperature for 30 min, followed by centrifugation at 35,000 rpm at 4°C for 1 hr. The supernatant was then stored at -80°C for future use. Sample preparation for chromatofocusing (CF) included use of a PD10 column, equilibrated with 25 mM bis-tris in 6 M urea and 0.2% OG, which was used to exchange the tissue lysate from the lysis buffer to the above buffer.

4.2.3 Two-dimensional Protein Fractionation

Chromatofocusing (CF) on tissue lysate

Two buffers, a start buffer and elution buffer, were employed in the CF experiment. The start buffer was 25 mM bis-tris with pH 7.1, and the elution buffer consisted of a 10-fold dilution of polybuffer 96 and polybuffer 74 in a ratio of 3:7, the pH adjusted to 4.0. Both buffers were prepared in 6 M urea and 0.2% OG. Iminodiacetic acid

was used to adjust the pH of both buffers. A Beckman Coulter PF2D HPLC system was used for all the separations stages. A PS-HPCF 1D (250×2.1 mm) column was equilibrated with the start buffer until the pH of the effluent was 7.1. Sample was applied to the column with multiple injections. Once a stable baseline was achieved, the elution buffer was switched on to elute the proteins on the column in an isocratic mode. UV detection was performed at 280 nm and the pH of the effluent was monitored using a flow-through online pH probe. pH fractions were collected for every 0.2 pH intervals and 15 fractions in all were collected in the range of pH 7.0-4.0. The CF separation was stopped when the pH of the effluent reached 4.0.

Reverse phase HPLC on pI fractions

RP-HPLC was performed using PS-HPCF 2D (33×4.6 mm) columns. Solvent A was 0.1% TFA (JT Baker, Phillipsburg, NJ) in water and solvent B was 0.1% TFA in acetonitrile (Burdick and Jackson, Muskegon, MI). The Solvent B gradient was run from 5% to 15% in 1 min, 15% to 25% in 2 min, 25% to 31% in 2 min, 31% to 41% in 10 min, 41% to 47% in 6 min, 47% to 67% in 4 min, to 100% in 3 min, held at 100% for another 1 min, then back to 5% in 1 min. The flow rate was 1 mL/min, and the column temperature was 65°C. UV absorptions were monitored at 214 nm. RP fractions were collected using a FC204 fraction collector (Beckman Coulter). The fractions were dried down completely using a Speed-Vac at 75°C and stored at -80°C until further use.

4.2.4 Microarray Procedures

Printing the Fractionated Proteins

The fractionated proteins were re-suspended in 15 μL buffer containing 1XPBS (at pH 7.4) and protease inhibitors (Roche Biochemicals, Indianapolis, IN). The samples were transferred to a 96-well microtitre plate (MJ Research, Waltham, MA) and printed on nitrocellulose slides (Schleicher & Schuell, Keene, NH) using a GeSim Nanoplotter2, a non-contact ink jet printer. Each spot measured approximately 300 μm in diameter and was placed 1200 μm apart. The slides were dried for 1 hr at room temperature and were either used immediately or stored in a dessication chamber at -20°C.

Hybridization of slides

Nitrocellulose slides containing spotted proteins were hydrated in 1XPBS for 10 minutes and blocked in 1XPBS containing 1% BSA (Sigma, St Louis, USA) and 0.1% Tween-20 (Sigma, St Louis, USA) overnight at 4°C. The slides were then incubated with either serum from prostate cancer patients or benign individuals (1:400 dilution) in probe buffer (PBS, pH 7.4 containing 1% BSA, 5 mM $MgCl_2$, 0.5 mM DTT, 0.05% Triton X-100 and 5% glycerol) for 2 hr at 4°C. Slides were then washed six times with probe buffer, each for 5 minutes. They were then incubated with Alexa Fluor 647 conjugated anti-human IgG (1:2000 dilution, Invitrogen, Carlsbad, CA) for 1 hr at 4°C, and then washed with probe buffer as above, dried through centrifugation at 8000 rpm, and analyzed using a GenePix 4000B microarray scanner (Axon Instruments Inc., Foster City, CA).

4.2.5 Data Analysis

Primary analysis, including scanning and quantification of slides, was executed with GenePix 6.0 (Axon Instruments Inc.); gridding was completed per manufacturer instructions. The single-channel Alexa Fluor 647 values were calculated for each individual fraction spot. An initial round of spot check was performed using GenePix default parameters. This was followed by a second round of curation where spots having any of the following characteristics were manually flagged: a diameter of less than 300 μm, irregular outline, spots localized in region of high local background or spot localized in areas of the array with obvious defects. Flagged spots were seeded to -1 in raw intensity units in the subsequent analysis. The background corrected median spot intensity values of the 'Cy5 channel' was extracted from each array and normalized. The data set was filtered for dominantly negative clones, retaining only those with non-negative raw intensity in $\geq$50% of samples in the cohort. Intra-array standardization entailed median centering and scaling by their respective median absolute deviations. Quantile normalization was then executed to obtain the same empirical distribution across all arrays. Two-way average-linkage hierarchical clustering of an un-centered Pearson correlation similarity matrix was executed and figures were generated using Cluster and TreeView [15].

4.2.6 Development of a Predictor

A supervised analysis was conducted to coalesce around a subset of fractions from the 2,016-element humoral response arrays which were most predictive for class distinction across the serum samples. Array data, normalized as previously described, were applied to a test statistic-based feature selection procedure calculating the F-statistic

between cancer and BPH samples across all 2016 clones (fractions). Different counts of the best ranking fractions by F-statistic (every count of best clones from 5 to 100) were used to build a Support Vector Machine (SVM) prediction model. The SVM over multiple kernel test permutations was embedded in a finite grid-search of paired values of exponentially growing sequences of cost (C) and gamma (γ). A linear kernel produced the best prediction, whose accuracy and error were calculated using leave-one out cross validation (LOOCV) to evaluate the performance of the models. The top-ranked 20 fractions were ultimately selected from the fraction predictor based on their best performance in specificity and sensitivity and with highest stability of recurrence at the top ranks over all the left-out iterations. These 20 fractions were additionally tolerant to repeated testing and small changes to model parameters. All statistical analyses were performed in R 2.3.0 and SPSS.

4.2.7 Mass Spectrometry

Targeted UV peaks in the second dimension RP-HPLC chromatogram were collected and aliquoted into two fractions. The protein content of fractions identified by best classification performance as well as 27 fractions demonstrating no differential response between classes were digested using Porcine Trypsin (1:50, Promega, Madison, WI) in 1 M ammonium bicarbonate, pH 9.0. The digestion was performed for 16 hr at 37°C. At the end of 24 hr, the trypsin activity was stopped using 3% formic acid. The peptide mixtures were separated by reverse-phase chromatography using a 0.075×150 mm C18 column attached to a Paradigm HPLC pump (Michrome BioResources Inc, Auburn, CA). Peptides were eluted using a 45-min gradient from 5 to 95% B (0.1% FA/95% ACN), where solvent A was 0.1% FA/2% ACN. A Finnigan LTQ mass spectrometer (Thermo Fisher Corporation, Waltham, MA) was used to ac-

quire mass spectra, the instrument operating in data-dependent mode with dynamic exclusion enabled. The MS/MS spectra for the three most abundant peptide ions in full MS scan were obtained. The spectra were searched using the Mascot algorithm (MatrixSciences, Boston, MA) against the composite, NCBI human RefSeq database (downloaded on 2005-06-28). The search was done using a mass tolerance of 2 Da for the precursor and 0.5 Da for the daughter fragments. All searches were performed using monoisotopic peptide mass with methionine oxidation (M+16) as the variable modification. Confidence in peptide assignment accuracy and protein identifications were assigned via the open source Trans-Proteomic Pipeline (Institute for Systems Biology) implementing PeptideProphet and ProteinProphet, which validates peptides assigned to MS/MS spectra and protein identifications respectively [16–18]. The resulting protein lists were filtered using a ProteinProphet probability of $\geq$0.90, corresponding to an error rate of less than 1% as estimated by ProteinProphet. All proteins identified by a single peptide having a charge state of +1 were also removed due to reduced confidence in these identifications (where the raw data is obtained using an ion trap mass spectrometer). Keratins were manually removed from the filtered list of proteins. All validated identifications, peptide sequences and annotations were stored in a relational database for downstream analysis.

4.2.8 MCM Analysis

The statistical model used for testing and storing the results of associations between independent molecular concepts has been previously described [19]. Protein identifications from the humoral response signature were converted to Entrez Gene identifiers and batch loaded to MCM (`www.molecularconcepts.org`) for analysis. Concept-to-concept enrichment extensions as described in the text were performed

with the same. To generate an immune program under STAT control for meta-analysis with public gene expression studies, we downloaded the IRIS repository of 1622 genes expressed in, and classified by multiple immune cell lineages [20]. There is a 179-gene overlap between the IRIS compendium and those genes under STAT control (the union of either STATx, STAT1, STAT3 or STAT5B homodimer). This seeds the metamap analysis described in the text above.

Additionally, 27 fractions lacking reactivity were selected from the array platform for sequencing. Identical thresholds and standards were used to cull the subsequent negative control protein list from 14 of those fractions for the MCM analysis. The unique count of high-probability identifications from the negative control protein content was 415, a similar number as the humoral response signature, eliminating potential two-group offset issues between the putative signature and its negative control. The gene identifiers representing these 415 encoded proteins were uploaded to MCM as a negative control enrichment analysis. Any concept enriching both the real signature of increased humoral response in prostate cancer and the negative control, were removed for further analysis.

4.3 Results

4.3.1 Development of arrays via proteome fractionation

An overview of the approach we took in identifying humoral response targets in prostate cancer is depicted in Figure 4.1. To generate protein microarrays for prostate cancer, we independently fractionated proteins from clinically localized and hormone-refractory metastatic cancer tissues (n=5 each) in two dimensions using chromato-focusing and reverse-phase chromatography. The fractionated proteins were spotted

on nitrocellulose coated glass slides and served as bait for capturing potential autoantibodies found in serum. Proteins that reacted with prostate cancer sera but not with the control were identified using mass spectrometry, database search and downstream protein informatics (see Methods). The list of proteins obtained was used to both characterize the predictor, but also in a 'molecular concept' analysis for their involvement in disease processes (Figure 4.8, Table 4.5-4.8).

Approximately 2,300 fractions were used to generate the protein microarrays. Using this 2,300-feature protein microarray, we evaluated sera from prostate cancer patients and controls. A one-color system with a red fluorescent Alexa Fluor 647 dye was used to measure the levels of bound IgG. Therefore, increased intensity represented varying levels of immune reactivity. In this discovery approach we evaluated 34 serum samples consisting of 18 sera from prostate cancer patients (biopsy-positive, high PSA) and 16 from individuals with BPH (biopsy-negative for cancer, high PSA). Critically, these samples constitute the clinically challenging distinction between cancer-negative (benign hyperplastic condition) and cancer-positive needle biopsy findings in the setting of elevated levels of circulating prostate-specific antigen.

4.3.2 Identification and Validation of the 20-fraction Predictor

The primary aim of the microarray experiment was to obtain the pattern of differential autoantibody response that could discriminate between the benign and prostate cancer groups. Cross-validated supervised analysis implementing the non-parametric Support Vector Machine (SVM) was performed using the 34 samples as a training set, looking for humoral response correlates of the two-class distinction between BPH and PCa (see Methods). Of the 1522 features remaining after filtering for dominantly

negative fractions as the result of hybridization, a subset of proteins demonstrated differential reactivity patterns. Embedded feature re-selection during LOOCV produced a 20-fraction predictor having 75% specificity (4 of 16 BPH samples were misclassified) and 78% sensitivity (4 of 18 prostate-cancer samples were misclassified) in discriminating between the group with BPH and that with PCa (Table 4.3). This was an especially encouraging result, as this is a particularly challenging classification; all members of the cohort have a baseline inflammatory physiological condition that is eliciting a humoral response accompanying their biopsy-proven clinical condition. The 20-fraction predictor was chosen as the minimum fraction count producing maximum accuracy in classification without over-fitting the predictor and with highest recurrence over left-out iterations during cross validation, otherwise interpreted as highest tolerance in statistical significance to the leaving out of any given sample in the cohort (Figure 4.2a, Methods and Table 4.3, 4.4). A heatmap of the reactivity profile generated by the 20-fraction predictor reveals a distinct bipartite pattern as would be expected by real reactivity to class-dependent antigens, in lieu of silence in reactivity (Figure 4.2b, 4.3a). The heterogeneity of the reactivity profile is likely attributable to many causes including loss of humoral response. The latter has previously been reported in patients diagnosed with breast cancer where loss of autoantibodies to mucin have been shown to be indicators of poor prognosis [21]. Additionally, in this study BPH is hyperplastic and already inflammatory, representing a heterogeneous population consisting of a mixture of biopsy-negative and potentially PCa-positive individuals. We are interested in identifying and quantifying the increase using BPH as the baseline condition. Finally, some instability in reactivity is observed certainly due to the above mentioned issues of the consistency of response across patients to specific intracellular antigens.

We next calculated receiver-operating characteristic curves for the 20-fraction predictor and measured PSA levels in the sample cohort (Figure 4.3b). A variety of cutoff values of the SVM decision scores were used as thresholds to plot the true- against false-positive rates for the prediction model. The ability of the 20 fraction predictor to discriminate PCa serum samples from BPH samples was significant (p=0.013) with an area under the curve of 0.75 (95% confidence interval, 0.58 to 0.92, Figure 4.3b). On the other hand, the area under the curve for measured PSA level was 0.49 (p=0.94; 95% confidence interval, 0.29 to 0.70). This result is significant as all benign patients in this cohort have elevated PSA levels associated with an early inflammatory condition, again a more challenging clinical distinction than with a healthy normal control. Further, a permutated rank-based test of significance of the difference in AUC between the SVM-derived predictor and that of measured PSA was also significant (p=0.05, Table 4.2 for association to additional pathological parameters).

4.3.3 Characterization of the 20-fraction Predictor

Though demonstrating promising results and certainly superior to the current PSA-based clinical standard, this result falls short of the accuracy of a desired diagnostic platform. Nevertheless, as these tumor-associated antigens are derived from the cancer proteome, and we expect them to be perturbed in cancer development given their loss of immunological tolerance, we hypothesized that their identification could form part of a larger alteration underlying a systematic biological process. To arrive at such a systems perspective, it was critical to identify these humoral response targets. Consequently, each of the 20 informative high-stability fractions from classification

was taken for mass spectrometry (see Methods and Table 4.5). In addition, 27 non-informative fractions were sequenced as a negative control for downstream analysis. A total of 359 unique proteins were identified from sequencing the 20-fraction predictor at probabilities of correct assignment greater than 0.90 [16]. A given fraction's immuno-reactivity profile is a complex combination of potential interactions. It may be the case that a fraction is composed of as few as 5 to 8 proteins, all of which elicit an autoantibody response to patient serum. Alternatively, fractions of equal or higher protein content with similar reactivity may be a combination of isolated proteins eliciting response, a cooperative humoral response between complexes of proteins, 'bleed-over' of proteins common to adjacent fractions, and those proteins that may appear in several fractions and represent no more than cellular machinery having nothing to do with eliciting the class-dependent immuno-reactivity of the fraction. We subsequently created a compendium of proteins from this sequencing that met criteria making them the most likely subset of proteins eliciting the humoral response signature. This included several rounds of subtraction for proteins considered either non-specific or noise to the reactivity profile (see Methods, Table 4.6, 4.7, Figure 4.9). The final compendium of likely humoral response targets numbered 248 (Table 4.8).

Regulated proteins such as prostatic acid phosphatase (ACPP, Figure 4.10a) and hypoxia up-regulated 1 (HYOU1), proteins involved in FGFR signaling pathway like synaptotagmin binding and cytoplasmic RNA interacting protein (SYNCRIP), regulators of actin cytoskeletal reorganization including calponin 1 (CNN1), Was/wasl interacting protein family member 2 (WICH2), and valosin containing protein (VCP), and finally tumor suppressors including a novel ring finger B box, coiled-coil family

member (HLS5), synaptopodin (SYNPO2), and trypsinogen IV (PRSS3) were identified.

To determine the ability of our PCa-specific humoral signature to identify pathways that are deregulated during prostate cancer development and progression, we performed a 'molecular concept' analysis (MCM) on the group of proteins that were identified to have cancer-specific autoantibody repertoire (see Methods, Figure 4.4 and Table 4.8). Recent work in the enrichment analysis of gene sets biologically related in a meaningful way allows for the discovery of patterns of shared behavior over a vast database of high-throughput experimental data and biological annotation [19]. Of the 27 fractions selected as a non-differential negative control and sequenced as described, the subsequent protein content of 14, culled with the same criteria operated as a control during concept enrichment. Any concept enriching both the differential predictor and the negative control were subtracted from the analysis. The MCM analysis of the 'increased humoral response in prostate cancer' signature identified an enrichment network containing metabolism concepts, including the KEGG pathway for nitrogen metabolism ($p=4.6\times10^{-5}$) as well as the mitochondrion cellular localization ($p=3.4\times10^{-4}$), and multiple drug compounds, including a gene set up-regulated upon cyclosporin treatment (p=0.006), which is an immuno-suppressive, and a gene set down-regulated upon treatment of rosiglitazone (p=0.006), having a potent anti-inflammatory effect (Figure 4.4). Additional concepts enriched by the PCa-specific humoral targets included four promoter binding sites implicating the enrichment of a STAT-regulating transcriptome, and HPRD interaction networks for three proteins, which in concert, play a role in mRNA processing. Among these, several concepts were taken for further analysis (Figure 4.4 and 4.6).

Among the various aforementioned concepts, the nitrogen metabolism concept was intriguing in the context of earlier results that revealed a shift in the metabolism paradigm during prostate cancer development driven by increased protein biosynthesis [19]. This in turn was thought to be regulated by a combination of androgen and over-expression of the fusion isoform of the transcription factor ERG [19]. Critically, the observed humoral response to components of the nitrogen pathway correlate well with increased utilization of protein or its derivatives (amino acids) by prostate tumors as a prime source of energy generation. Furthermore, it is conceivable that increased protein biosynthesis would lead to a linear increase in the generation of nitrogen which would first be directed into the nitrogen metabolism pathway in advance of downstream entrance into the urea cycle. Closer inspection of the nitrogen metabolism concept revealed three pathway components eliciting differential humoral response. These include two functionally similar glutamate dehydrogenase enzymes GLUD1 and 2 (Figure 4.10b), and carbonic anhydrase II. The former mediate the coupled conversion of alpha-ketoglutarate and ammonia to glutamate and the transamination of glutamate to alpha-ketoglutarate and ammonia. Additionally, glutamate participates in a second transamination reaction producing aspartate, a nitrogen donor to urea synthesis, and a regenerated alpha-ketoglutarate (Figure 4.5). The up-regulation of glutamate as a metabolite coupled to the enhanced enzymatic mediation via humorally reactive GLUD1 shifts chemical equilibrium toward increased ammonia output. The latter, coupled to the production of the nitrogen donor aspartate, represents up-regulated nitrogen production directed to downstream urea synthesis as well as activating a potential positive feedback mechanism in regeneration of alpha-ketoglutarate. Interestingly, an independent metabolomic dataset

generated using 16 benign prostate tissues and 12 localized prostate cancer samples shows significantly increased levels of glutamate and aspartate in prostate cancer specimens (Figure 4.5).

Thus, using MCM we were able to directly extend our interrogation beyond single-molecule targets to examine global changes in biological pathways. Nevertheless, there existed a nested set of significant concepts including multiple STAT-family promoter binding sites as well as c-Ets 1(68) sites, which in isolation, were not revealing. This was especially noteworthy as both have been shown to play an important role during prostate cancer development/progression [22–26]. This is not unexpected, as measuring autoantibody response, which unlike protein expression, is an indirect measure of protein abundance or modification state and is highly variable depending on the immunogenicity of targets and their exposure to the immune system. Given this situation, one could imagine a scenario where a change in a group of proteins under the control of a common regulator can be attributed to a holistic change in the programming controlled by the regulator. In other words, change in levels of the master regulator could lead to a cascade of expression changes of its targets, only some of which will subsequently elicit a humoral response. Thus, it was important for us to understand the fate that results from deregulation of such regulatory components (in this case, the five concepts enriched in our MCM analysis). In an attempt to interrogate this aspect, we executed individual MCM enrichment extensions for five concepts highlighted in Figure 4.4 and 4.6.

The analysis was effectively extended by one level. Instead of enriching for concepts from the original humoral target list, each of the five concepts from the original analy-

sis seeded a subset MCM enrichment. The five resulting enrichment networks were sequentially merged into a single common network (Figure 4.6). Orphaned concepts from a single-concept extension were removed during this merge stage. Interestingly, extension of the four STAT concepts and c-Ets 1(68) binding site concept, revealed a common and systematic theme of immune modulation. The extended concepts reveal a high count of immuno-modulatory events that included acute phase response, complement and coagulation cascades, chemokine activity, and more. There was a significant overlap of the STAT controlled-gene program with those regulated by ETS-family transcription factors and AR, all of which have been critically implicated in prostate cancer [22, 23]. The extension analysis also identified an overlap with gene expression profiles that were down-regulated upon treatment with a variety of anti-inflammatory drugs, further confirming the existence of an underlying immune modulation theme in the humoral signature [25].

To confirm the existence of STAT-induced immune programming, we completed a meta-analysis, seeded by a STAT-regulated immune signature (see Methods), with six prostate gene expression signatures profiling different comparisons between normal prostate, BPH or prostate carcinoma using Oncomine (Figure 4.7a, 4.7b, `www.oncomine.org`) [27–29]. There is a dominant pattern of over-expression for this STAT-regulated immune signature in clinically defined carcinoma relative to the benign condition in each study (Figure 4.7b). Additionally, a subset of immune-regulatory genes under STAT control was significantly enriched in BPH compared to normal tissue, and the immune activity of PCa was found to be higher than BPH (Figure 4.7b). It is important to note that our study design compares the immune response profile between two immunologically active cohorts, and does not include

profiling of normal sample due the ambiguities of defining such a cohort of age-matched individuals for prostate disease. In an attempt to address the scarcity of humoral targets in the cancer-specific signature and appreciating the disadvantage of low concordance between gene and protein expression, our aim was to identify any difference in expression among a set of immuno-modulatory targets across a broad set of study data.

4.3.4 Clinical associations of the humoral response signature

To evaluate whether the 20-faction predictor is useful as a supplement to measured PSA, logistic regression analysis was performed on the sample cohort. Disease state (cancer or non-cancer) was assigned to the response variable and univariate logistic regression was executed for both the standardized SVM decision scores and measured PSA, independently at first, and then in combination. It was found that only the SVM-derived decision score test was statistically significant (odds ratio, OR for standardized decision scores=3.74, 95% CI=1.23-11.36, p=0.02; OR for PSA=1.019, 95% CI=0.87-1.18, p=0.80). In the multivariate logistic regression analysis with disease as the response and fitting both SVM decision scores and measured PSA as covariates, we found that the effect of the decision scores was significant even after adjusting for the effect of PSA (OR=4.13, 95% CI=1.28-13.34, p=0.0177). We then calculated the likelihood ratio test statistic to test whether the addition of the added covariate improves the fit over the univariate logistic regression model of PSA. The addition of decision scores to PSA was significant in fit improvement (p=0.0104). This indicates that the 20-fraction predictor provides additional predictive value over measured PSA alone.

This was further validated with a rank-based permutation test of significance in difference between the SVM-derived and measured PSA AUCs. The real difference in AUC between the SVM model and PSA was calculated. After converting both the numeric measurements of PSA and the SVM decision scores to ranks, the sample class labels were randomly permutated and the AUC difference between the two metrics were recomputed over 10,000 iterations. The relative p-value of this test is equivalent to a one-sided test of the alternative hypothesis that there is a larger difference in AUC between the two models and the null hypothesis of equal AUCs. The calculated p-value, the count of AUC differences greater than the master difference over the 10,000 iterations was significant (p=0.05).

4.4 Discussion

By coupling multidimensional protein fractionation with protein microarrays this proof-of-principle study demonstrates the power of immune system-driven autoantibody response for detection of prostate cancer. Through its ability to detect proteomic alterations, the autoantibody response can reveal deregulated biological processes during cancer development and progression. Further, this study utilizes a clinically challenging population in which both the benign and localized cancer patients have high amounts of circulating PSA, the current clinical standard for prostate cancer detection. Importantly, only biopsy results were used to cull the cancer subgroup in this patient cohort. Needle biopsy by itself, in addition to being invasive, has a well-documented false negative rate, missing 28% of existing prostate cancers [30]. This adds to the complexity of the study cohort as a subset of benigns (as defined by negative biopsy) may very well harbor neoplasm that went undetected. Given

this fallacy, it was interesting that the 20 fraction tumor-specific humoral signature could classify PCa with 75% specificity and 78% sensitivity, respectively. In addition to known autoantibody targets like ACPP and PSA, the PCa-specific humoral signature contained proteins that could play a role during tumor development and progression. For instance, the humoral signature included two proteins in the FGFR signaling which has been implicated in prostate cancer development and progression [31–33]. Different predicted molecular phenotypes of increased FGF signaling include increased motility, invasiveness, proliferation and androgen independence, all of which are ultimately thought to promote tumor progression [33]. Notably, elevated machinery regulating motility and invasion was evident in the PCa-specific compendium of humoral targets that included known regulators of actin cytoskeletal reorganization, namely calponin 1 (CNN1), Was/wasl interacting protein family member 2 (WICH2) and valosin containing protein (VCP). Also, FGF is known to potentiate tumor progression by signaling through various pathways that include the STAT pathway [33], components of which were present in the cancer-specific humoral signature.

In addition to being sensitive to changes in protein levels or modifications, the humoral response generation could be driven by other factors like the presence of anti-idiotypic antibodies that constitute the causal mechanism behind the observation of humoral response against the autoantigen [34]. Such is proven to be the case for Wegener granulomatosis autoantigen or PRTN3, autoantibodies to which are found in inflammatory vascular disease [34]. These studies indicate that, during vascular inflammation, there occur complementary, or antisense, peptides that can bind to PRNT3, termed cPR-3, which are the initial targets for generation of

autoantibodies [34]. These cPR-3-specific idiotypic antibodies then elicit the anti-idiotypic response, antibodies originating from which can bind to the parent antigen PRNT3 [34]. In our case, a similar concept may govern the presence of autoantibodies to GLUD1/GLUD2, for which HSP70-like protein has been shown to be the antisense gene pair in moulds [35]. Further, HSP70 has been shown to elicit autoantibodies to itself [36, 37]. Accordingly, one might predict that reactivity of prostate cancer serum to GLUD1/2 may potentially follow a mechanism similar to the one reported for PRNT3 [34]. Alternately it is worth noting that both GLUD1 and GLUD2 are key regulators of nitrogen metabolism and downstream urea cycle activity. This gains importance in the context of our earlier gene expression-based analyses that revealed increased protein biosynthesis in localized prostate cancer [19]. The breakdown of the resulting proteins could lead to increased accumulation of elemental nitrogen that will have to be eliminated through nitrogen breakdown and urea cycle pathways. Our humoral response data coupled with independent assessments showing increased levels of glutamate, aspartate, and constituent urea cycle metabolites in localized prostate cancer corroborate our gene expression-based hypothesis. Of course the metabolic flux of both GLUD1 and GLUD2 as well their potential anti-idiotypic antibody response in prostate carcinoma requires further investigation.

In addition to drawing direct correlates between humoral targets and known pathway alterations in tumors, it was intriguing to observe proteins in the STAT pathway, which by themselves did not enrich for any known tumor-associated processes. Specificity of these STAT pathway-associated proteins to prostate cancer is validated by their absence in a random humorally-nonreactive signature (see Methods). More-

over, it has been observed that STATs are regulated by FGF, the activity of which, according to our data and previous reports, is elevated in prostate cancer [33]. Also, elevation of STAT signalling has been reported in prostate cancer [38]. Moreover, it is known that proteins differ in their ability to generate an antibody response [39]. Accordingly, it was tempting to speculate that the existence of a group of proteins in a given pathway in our prostate cancer-specific humoral signature may signify global reprogramming of that pathway, in this case, the global perturbation of the STAT pathway. We further interrogated the dataset for enrichment of the STAT-regulated proteome with MCM, which revealed multiple immuno-modulatory proteins as defined by an earlier gene expression study [20]. This STAT-induced deregulation of immuno-modulatory components was validated by meta-analyses of independent prostate cancer data sets. The meta-analysis reveals existence of high levels of basal immuno-activity in BPH, which is further inflated during tumor development. This is not only consistent with reports of neoplastic development occurring on a background of focal inflammation [40], but also explains the challenge associated with discerning the two classes, namely BPH and localized cancer based on their response profiles as reported in this study. Thus, we highlight the ability of autoantibody repertoires to uncover alterations in biological processes that might otherwise not be revealed by standard protein profiling platforms owing to the vast dynamic range of the proteome.

4.5 Conclusion

The ability to correlate the humoral signature with actual cellular processes emphasizes the importance of employing a screening platform containing proteins extracted

from tumor, itself reflecting a physiologically realistic swath of the prostate cancer proteome. Such correlations to tumor function would be difficult to establish using proteins fractionated from cell lines [41, 42] or using phage array platforms [14, 43]. In addition, the multi-dimensional protein fractionation-coupled microarray retains post-translational modifications that are most indicative of the cellular phenotype, and in most cases better reflects the reality of humoral response to cancer antigens. This is best illustrated by the observation of phosphorylated PSA as a target antigen in our dataset (data not shown).

In summary, this study for the first time uses autoantibodies generated against the tumor proteome to classify a clinically challenging cohort of patients comprised of BPH and localized PCa and, in the process, reveals multiple alterations in tumor-associated functional pathways that would otherwise not be discernable by conventional profiling strategies.

Table 4.1: Clinical and pathology information for the 34 benign prostatic hyperplasia and clinically localized prostate cancer patient serum samples used in the training/validation set. *(1) Plus-minus value are mean ± SD. PSA denotes prostate-specific antigen. (2) Data were available for 34 patients.

Characteristic*	**Value (Cancer Patients)**	**Value (BPH Patients)**
No. of patients	18	16
Age (yr)	63.2 ± 12.77 (35-81)	64.8 ± 10.47 (43-80)
PSA level		
Mean (ng/ml)	7.18 ± 5.34 (2.9-20.4)	6.79 ± 3.76 (2.1-14.1)
0-4 ng/ml (%)	38.9	25.0
4.1-10 ng/ml (%)	38.9	50.0
>10 ng/ml (%)	22.2	25.0
Gleason grade (%)		
Minor		
3	44.4	-
4	50.0	-
5	5.6	-
Major		
3	77.8	-
4	22.2	-
≤6	33.3	-
≥7	66.67	-
Race		
White (non-Hispanic origin) (%)	89	87.5
Black (non-Hispanic origin) (%)	11	6.25
Asian or Pacific Islander (%)	-	6.25

Table 4.2: Associations between the 20-fraction humoral response signature and various clinical and pathological parameters.

Variable	**Analysis**	**PSA**	**Total Gleason**	**Major Gleason**	**Minor Gleason**
20-fraction	Pearson correlation	-0.204	-0.214	-0.334	-0.041
signature	P-value	0.248	0.394	0.175	0.871

Table 4.3: Summary of class predictions for the sample set. A prediction model was built using the 20 fractions selected from a leave-one-out cross validation strategy implementing a Support Vector Machine. The column 'Call' is the prediction of a sample by the model along with its numeric decision value, which is the signed distance from the hyperplane constructed by the trained model. Error indicates misclassified samples.

Sample Name	**Pathology**	**Decision Value**	**Call**	**Error**
Serum-1	Cancer	-1.793	Cancer	
Serum-2	BPH	-1.56278	BPH	
Serum-3	BPH	0.030892	Cancer	*
Serum-4	BPH	-2.1205	BPH	
Serum-5	BPH	-0.149	BPH	
Serum-6	BPH	-0.57144	BPH	
Serum-7	Cancer	0.717425	Cancer	
Serum-8	Cancer	0.418136	Cancer	
Serum-9	Cancer	0.357485	Cancer	
Serum-10	BPH	-2.55515	BPH	
Serum-11	Cancer	1.644764	Cancer	
Serum-12	Cancer	3.678237	Cancer	
Serum-13	Cancer	-1.33488	BPH	*
Serum-14	BPH	-0.55419	BPH	
Serum-15	Cancer	1.151819	Cancer	
Serum-16	BPH	0.115368	Cancer	*
Serum-17	BPH	0.829295	Cancer	*
Serum-18	Cancer	5.249221	Cancer	
Serum-19	BPH	-0.28224	BPH	
Serum-20	Cancer	-0.91374	BPH	*
Serum-21	BPH	-2.92816	BPH	
Serum-22	Cancer	1.803605	Cancer	
Serum-23	Cancer	0.997157	Cancer	
Serum-24	BPH	-0.36761	BPH	
Serum-25	Cancer	1.320617	Cancer	
Serum-26	BPH	-0.18682	BPH	
Serum-27	BPH	-0.80733	BPH	
Serum-28	BPH	1.176218	Cancer	*
Serum-29	BPH	-0.67047	BPH	
Serum-30	Cancer	4.546478	Cancer	
Serum-31	Cancer	3.274788	Cancer	
Serum-32	Cancer	-1.3201	BPH	*
Serum-33	Cancer	0.770342	Cancer	
Serum-34	Cancer	-1.68823	BPH	*

Table 4.4: Comprehensive list of fractions used during the 20-fraction best classification in sensitivity and specificity. Recurrence is the count of samples that contributed that clone to the top 20 fractions during leave-one out cross validation (LOOCV). High-stability fractions are defined as those persistent in high rank regardless of sample left out. The F-statistic for each fraction is a representative value from a single of the 34 iterations of LOOCV calculated on that iteration's 33 training samples.

Fraction	**Recurrence**	**F-Statistic**	**Fraction**	**Recurrence**	**F-Statistic**
3A8PCA3	34	7.793006	22D3Met8	4	6.185279
1B5PCA1	34	11.46584	6F9PCA5	3	5.071975
2F2PCA2	34	8.370373	7A6PCA6	2	6.293373
4H4PCA4	34	9.212151	21E9Met7	2	5.246445
14A8Met2	34	8.2016	21F11Met7	2	5.521059
13B3Met1	34	11.15364	4E11PCA4	1	4.17968
23C5Met9	34	12.90059	6E7PCA5	1	5.659875
19C12Met6	34	7.847889	6E4PCA5	1	4.979168
13D4Met1	34	11.60105	6E12PCA5	1	2.646888
19D12Met6	34	13.66645	3B12PCA3	1	5.743078
23H3Met9	30	6.637657	7B8PCA6	1	4.475109
3H11PCA3	28	7.684081	2C5PCA2	1	5.811833
13H10Met1	28	6.685937	3C8PCA3	1	4.902343
6D9PCA5	27	5.819973	9D7PCA8	1	4.49316
18F3Met5	27	7.442341	4D6PCA4	1	3.782356
2B3PCA2	26	10.26944	3H6PCA3	1	6.13266
22C5Met8	25	5.653969	13A1Met1	1	6.00282
7A12PCA6	22	6.690722	21E7Met7	1	6.222062
3F5PCA3	21	7.154193	24A4Met10	1	6.356494
19C11Met6	15	6.423083	22E12Met8	1	2.533672
3B6PCA3	10	4.878724	24F7Met10	1	4.572809
18G8Met5	10	5.856973	19B10Met6	1	4.746501
7G7PCA6	7	7.162571	23C11Met9	1	2.417708
4E6PCA4	6	7.380854	19C6Met6	1	5.724045
26H3Met11	5	5.859942	13D9Met1	1	4.375554
21D12Met7	5	5.303875	17D7Met4	1	3.899244
2A10PCA2	4	4.955912	17H3Met4	1	3.603949
4B5PCA4	4	4.643863	26H5Met11	1	2.970845
4D11PCA4	4	5.882793			

Table 4.5: List of fractions taken for mass spectrometry. The fractions listed in the first two columns are those clones presented in the heatmap from Figure 4.2 and are most stable in LOOCV iterations during classification. Fractions are grouped by their reactivity pattern membership.

Fractions taken for mass spectrometry		
BPH	**PCa**	**Negative Control**
2B3PCA2 18F3Met5	1B5PCA1	1A9PCA1 1C5PCA1 14F2MET2 16B5MET3
4H4PCA4 19C11Met6	2F2PCA2	1C3PCA1 1C10PCA1 16B3MET3 17B5MET4
6D9PCA5 19C12Met6	3A8PCA3	1C8PCA1 3G11PCA3 17B3MET4 17G4MET4
13B3Met1 19D12Met6	3F5PCA3	3G9PCA3 7E5PCA6 17G2MET4 22A11MET8
13D4Met1 22C5Met8	3H11PCA3	7E3PCA6 7E11PCA6 22A9MET8 22B12MET8
13H10Met1 23C5Met9	7A12PCA6	7E9PCA6 11A7PCA9 26C12MET11 26D12MET11
14A8Met2 1A11PCA1	23H3MET9	11A5PCA9 14F4MET2

Table 4.6: Non-overlapping adjacent fraction protein content removed from the final protein compendium of suspected targets of autoantibody response.

Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)	Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)
gi6382060	27	ABL2	19D12MET6(1, 1)	gi51460532	54454	KIAA1240	19D12MET6(1, 1)
gi4557231	34	ACADM	19C11MET6(1, 11), 19C12MET6(0.99, 1)	gi14149742	57658	KIAA1536	19D12MET6(0.9, 1)
gi47078295	100	ADA	19C11MET6(0.99, 1), 19C12MET6(0.98, 1)	gi4504981	3956	LGALS1	19C11MET6(1, 2)
gi32484973	132	ADK	19C11MET6(0.94, 1)	gi33636766	64077	LHPP	19C11MET6(0.99, 1)
gi4502013	204	AK2	19C11MET6(0.99, 1)	gi51471054	121006	LOC121006	19D12MET6(1, 1)
gi24497577	10327	AKR1A1	19C11MET6(1, 8), 19C12MET6(0.99, 1)	gi51466547	154761	LOC154761	19D12MET6(1, 1)
gi41327764	8574	AKR7A2	19C11MET6(1, 4), 19C12MET6(0.99, 1)	gi23943872	64744	LOC64744	19D12MET6(1, 1)
gi4757756	302	ANXA2	19C12MET6(1, 2), 19D12MET6(1, 5)	gi14211889	84661	LOC84661	19C11MET6(1, 3)
gi4502109	309	ANXA6	19C11MET6(1, 2)	gi45827809	9064	MAP3K6	19D12MET6(1, 1)
gi40068462	57492	ARID1B	19D12MET6(1, 1)	gi21264363	10747	MASP2	19D12MET6(1, 1)
gi32526896	25852	ARMC8	19D12MET6(1, 1)	gi11034825	27430	MAT2B	19C11MET6(1, 5), 19C12MET6(0.99, 1)
gi13435129	23400	ATP13A2	19D12MET6(0.97, 1)	gi5174533	4160	MC4R	19D12MET6(1, 1)
gi21361181	476	ATP1A1	19D12MET6(1, 1)	gi7661788	29079	MED4	19C12MET6(0.98, 1)
gi4757834	9532	BAG2	19C12MET6(1, 3)	gi13129068	79050	MGC3162	19D12MET6(1, 1)
gi46430495	54796	BNC2	19D12MET6(1, 1)	gi22749137	153733	MGC39633	19D12MET6(1, 1)
gi36031016	55088	C10orf118	19D12MET6(1, 1)	gi33457311	145553	MGC5987	19C11MET6(1, 4)
gi51468578	338616	C10orf52	19D12MET6(0.95, 1)	gi8922701	55750	MULK	19C12MET6(0.99, 1)
gi27734837	255352	C10orf93	19D12MET6(0.98, 1)	gi32307128	23054	NCOA6	19D12MET6(1, 1)
gi51474886	23211	C19orf7	19D12MET6(0.99, 1), 22C5MET8(1, 1)	gi4505357	4697	NDUFA4	19D12MET6(1, 1)
gi11037061	56683	C21orf59	19C11MET6(1, 5)	gi37693993	4832	NME3	19C11MET6(0.93, 1)
gi4502501	721	C4B	19C12MET6(0.99, 1)	gi51475065	343578	None	19D12MET6(1, 1)
gi4502517	759	CA1	19C11MET6(1, 2), 19C12MET6(1, 5), 4H4PCA4(0.99, 1)	gi51471384	387946	None	19D12MET6(1, 1)
gi10716563	821	CANX	19D12MET6(0.97, 1)	gi51475265	388819	None	19D12MET6(0.99, 1)
gi16507962	64072	CDH23	19D12MET6(1, 1)	gi51460480	388922	None	19D12MET6(1, 1)
gi38158011	1149	CIDEA	19D12MET6(1, 1)	gi51474106	401887	None	19D12MET6(1, 1)
gi10334859	1159	CKMT1	19C12MET6(0.99, 1)	gi42656911	402152	None	19D12MET6(0.98, 1)
gi6912310	22802	CLCA4	19D12MET6(1, 1)	gi51467792	439997	None	19D12MET6(1, 1)
gi33457344	57511	COG6	19D12MET6(0.97, 1)	gi51464685	441098	None	19D12MET6(1, 1)
gi7656985	1282	COL4A1	19D12MET6(1, 1)	gi51466769	442400	None	19D12MET6(1, 2)
gi18105030	8292	COLQ	19D12MET6(1, 1)	gi34932414	4841	NONO	19C11MET6(0.96, 1), 19C12MET6(0.99, 1)
gi4502989	1347	COX7A2	19C11MET6(1, 5)	gi20127624	79400	NOX5	19D12MET6(0.99, 1)
gi25121972	144402	CPNE8	19D12MET6(1, 1)	gi4505431	4863	NPAT	19D12MET6(1, 1)
gi41327708	23418	CRB1	19D12MET6(1, 1)	gi38570156	221294	NT5C2L1	19D12MET6(1, 1)
gi4758056	1387	CREBBP	19D12MET6(0.98, 2)	gi13259372	4969	OGN	19C11MET6(1, 5)
gi47080099	1393	CRHBP	19D12MET6(0.94, 1)	gi4501955	142	PARP1	19C12MET6(1, 5), 19D12MET6(1, 3)
gi29570791	1457	CSNK2A1	19C12MET6(1, 3)	gi4505615	5080	PAX6	19D12MET6(1, 1)
gi14702162	7461	CYLN2	19D12MET6(1, 1)	gi11761615	5091	PC	19D12MET6(0.95, 1)
gi31542331	3491	CYR61	19D12MET6(1, 1)	gi16933555	65217	PCDH15	19D12MET6(0.98, 1)
gi4758128	9201	DCAMKL1	19D12MET6(1, 1)	gi19923440	50940	PDE11A	19D12MET6(1, 1)
gi38201710	10521	DDX17	13D4MET1(1, 2), 19C11MET6(0.99, 1)	gi6912582	23578	PEF	19D12MET6(1, 1)
gi24308179	55917	DKFZp547A023	19C11MET6(0.97, 1)	gi11863130	5277	PIGA	19D12MET6(1, 1)
gi4557525	1738	DLD	19D12MET6(0.98, 1)	gi23397648	84720	PIGO	19D12MET6(1, 1)
gi4503381	1810	DR1	19C12MET6(0.99, 1)	gi25777665	5314	PKHD1	19D12MET6(0.9, 1)
gi46195707	1954	EGFL4	19D12MET6(1, 1)	gi4826914	8681	PLA2G4B	19D12MET6(1, 1)
gi13324677	5167	ENPP1	19D12MET6(1, 1)	gi5032223	10154	PLXNC1	19D12MET6(1, 2)
gi11034849	59084	ENPP5	19D12MET6(1, 1)	gi4506193	5689	PSMB1	19D12MET6(1, 1)
gi13654274	81889	FAHD1	19C12MET6(1, 7), 19D12MET6(1, 1)	gi33457332	125950	RAVER1	19C12MET6(0.99, 1)
gi42661041	85302	FBF1	19D12MET6(1, 1)	gi41281444	9699	RIMS2	19D12MET6(1, 1)
gi13435350	2232	FDXR	19C12MET6(1, 10)	gi8923844	55811	SAC	19D12MET6(1, 1)
gi40255039	55236	FLJ10808	19D12MET6(1, 1)	gi4502261	462	SERPINC1	19C12MET6(0.97, 1)
gi41150761	55036	FLJ20753	19D12MET6(1, 1)	gi19923315	6472	SHMT2	19C11MET6(1, 18), 19C12MET6(1, 10)
gi27734761	286053	FLJ32440	19D12MET6(0.98, 1)	gi27436873	257218	SHPRH	19D12MET6(0.99, 1)
gi20149680	79672	FN3KRP	19D12MET6(1, 4)	gi18699722	57619	ShrmL	19D12MET6(0.98, 1)
gi24497501	3169	FOXA1	19C11MET6(0.99, 1), 19C12MET6(0.93, 1)	gi51474764	400709	SIGLECP16	19C12MET6(0.91, 1)
gi4826730	2475	FRAP1	19D12MET6(1, 1)	gi42661622	23094	SIPA1L3	19C12MET6(0.9, 1)
gi4503905	8811	GALR2	19D12MET6(1, 1)	gi30520363	133482	SLCO6A1	19D12MET6(1, 1)
gi6996010	2617	GARS	19D12MET6(1, 1)	gi21071052	6596	SMARCA3	19D12MET6(1, 1)
gi9558729	26088	GGA1	19D12MET6(1, 1)	gi30315658	6711	SPTBN1	19D12MET6(1, 1)
gi13518231	2736	GLI2	19D12MET6(1, 1)	gi38679884	6717	SRI	19C11MET6(1, 2)
gi7657128	29998	GLTSCR1	19D12MET6(1, 1)	gi5454086	8406	SRPX	19D12MET6(1, 1)
gi10567816	2775	GNAO1	19D12MET6(1, 1)	gi10835067	6741	SSB	19C11MET6(1, 11), 19C12MET6(1, 3)
gi4504053	2784	GNB3	19D12MET6(1, 1)	gi28557788	6744	SSFA2	19D12MET6(1, 1)
gi29826294	26003	GORASP2	19C11MET6(0.95, 1)	gi44921606	8867	SYNJ1	19D12MET6(1, 1)
gi29789277	84435	GPR123	19D12MET6(1, 1)	gi4507337	6861	SYT5	19D12MET6(1, 1)
gi41406084	2876	GPX1	18F3MET5(1, 8), 18F3MET5(0.99, 3), 19C11MET6(0.99, 1)	gi19913369	10607	TBL3	19D12MET6(1, 1)
gi24308295	80273	GRPEL1	19C11MET6(1, 2)	gi21536371	7011	TEP1	19D12MET6(1, 1)
gi4504351	3045	HBD	19C11MET6(0.99, 1)	gi23238188	26136	TES	19D12MET6(1, 1)
gi6715607	3048	HBG2	19C12MET6(0.99, 1)	gi13569901	81628	THG-1	19C12MET6(0.98, 1)
gi23097260	92521	HCMOGT-1	19D12MET6(1, 1)	gi13236587	79188	TMEM43	19D12MET6(0.91, 1)
gi20336761	50865	HEBP1	19C11MET6(0.92, 1)	gi22547116	8795	TNFRSF10B	19D12MET6(1, 1)
gi27436929	3305	HSPA1L	19C11MET6(1, 3), 19C12MET6(1, 3), 22C5MET8(0.98, 1), 23C5MET9(1, 4)	gi11321607	9319	TRIP13	19D12MET6(1, 1)

Continued on next page

Continued from last page

Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)	Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)
gi34419635	3310	HSPA6	19C11MET6(0.99, 1), 19C12MET6(0.99, 1), 23C5MET9(1, 1)	gi20143914	7273	TTN	19D12MET6(1, 1)
gi4504517	3315	HSPB1	18F3MET5(1, 4), 19C11MET6(0.92, 1), 4H4PCA4(1, 3)	gi13376539	80086	TUBA4	19C12MET6(0.99, 1)
gi31542947	3329	HSPD1	19C11MET6(1, 4), 23C5MET9(1, 6)	gi14389309	84790	TUBA6	19D12MET6(0.96, 1), 22C5MET8(1, 2)
gi28178825	3417	IDH1	19C12MET6(1, 2), 19D12MET6(1, 8)	gi13562114	81027	TUBB1	19C11MET6(0.98, 1)
gi10800142	3444	IFNA7	19C12MET6(0.99, 1)	gi9507221	53347	UBASH3A	19D12MET6(0.91, 1)
gi4557882	3561	IL2RG	19D12MET6(1, 1)	gi33188427	7404	UTY	19D12MET6(1, 1)
gi31317249	3570	IL6R	19D12MET6(1, 1)	gi21614499	7430	VIL2	19C11MET6(1, 2), 22C5MET8(1, 1)
gi5803115	10989	IMMT	19C11MET6(0.99, 1)	gi40068485	64856	WARP	19C12MET6(1, 7)
gi51460530	114818	KBTBD9	19D12MET6(1, 1)	gi4507909	7454	WAS	19D12MET6(1, 1)
gi4504825	8514	KCNAB2	19D12MET6(1, 1)	gi31543021	22911	WDR47	19D12MET6(1, 1)
gi26051271	3757	KCNH2	19D12MET6(1, 1)	gi4507931	7479	WNT8B	19D12MET6(1, 1)
gi51463940	26032	KIAA0527	19D12MET6(1, 1)	gi13386506	7490	WT1	19D12MET6(1, 1)
gi41281469	9786	KIAA0586	19D12MET6(1, 1)	gi10863945	7520	XRCC5	19C12MET6(1, 2)
gi39930349	23231	KIAA0746	19D12MET6(1, 1)	gi28274709	7789	ZXDA	19D12MET6(1, 1)

Table 4.7: Protein content removed as non-specific noise from the final protein compendium.

Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)	Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)
gi4501867	50	ACO2	18F3MET5(1,9), 3F5PCA3(0.99,2), 4H4PCA4(1,21), 18F3MET5(0.99,3), 22C5MET8(1,10), 23C5MET9(1,5), 23H3MET9(1,2), 6D9PCA5(1,2)	gi4557014	847	CAT	18F3MET5(1,8), 3A8PCA3(1,3), 3F5PCA3(1,6), 4H4PCA4(1,16), 19C11MET6(1,2), 19C12MET6(1,4), 19D12MET6(1,3), 7A12PCA6(0.98,1)
gi4502027	213	ALB	13B3MET1(0.99,1), 22C5MET8(1,2), 23C5MET9(1,1), 4H4PCA4(1,3), 7A12PCA6(1,5)	gi34577110	226	ALDOA	13D4MET1(1,6), 13H10MET1(1,9), 18F3MET5(0.95,1), 3A8PCA3(0.99,1)
gi4502101	301	ANXA1	19C11MET6(0.98,1), 19C12MET6(1,2), 7A12PCA6(1,3)	gi32189394	506	ATP5B	23C5MET9(1,4), 23H3MET9(0.99,1)
gi15011913	1291	COL6A1	19C11MET6(0.99,1), 19C12MET6(0.99,1), 3A8PCA3(0.91,1), 3F5PCA3(0.99,1), 4H4PCA4(0.99,1)	gi4501885	60	ACTB	22C5MET8(1,4), 23C5MET9(1,2), 23H3MET9(0.92,1)
gi24308169	55567	DNAH3	14A8MET2(0.99,1), 1B5PCA1(1,1), 23H3MET9(0.99,1), 2F2PCA2(1,1), 6D9PCA5(1,1)	gi13876382	1770	DNAH9	19D12MET6(0.99,1), 1B5PCA1(0.97,1), 23H3MET9(0.95,1), 6D9PCA5(0.97,1)
gi10800138	3017	HIST 1H2BD	13B3MET1(1,5), 13H10MET1(1,9), 2B3PCA2(0.99,1), 3A8PCA3(1,1), 3F5PCA3(0.99,1), 4H4PCA4(1,2)	gi4503471	1915	EEF1A1	13B3MET1(1,2), 13D4MET1(1,2), 18F3MET5(0.99,8), 19C11MET6(1,6), 23C5MET9(0.93,1), 2F2PCA2(0.98,1), 3A8PCA3(1,5), 3F5PCA3(1,5), 4H4PCA4(1,7)
gi11761629	2243	FGA	13D4MET1(0.93,1), 13H10MET1(0.96,1), 18F3MET5(1,3), 19C11MET6(0.93,1), 2F2PCA2(0.9,1), 3A8PCA3(0.96,1), 3F5PCA3(0.98,1), 4H4PCA4(0.98,1)	gi11761631	2244	FGB	13H10MET1(1,2), 18F3MET5(1,11), 19C11MET6(1,3), 19C12MET6(0.99,1), 4H4PCA4(1,7)
gi7661968	9685	ENTH	23C5MET9(1,1), 23H3MET9(1,1)	gi41327741	23474	ETHE1	18F3MET5(0.9,1), 3F5PCA3(0.99,1), 4H4PCA4(0.99,1)
gi11761633	2266	FGG	19C12MET6(1,2), 7A12PCA6(1,3)	gi16554592	2027	ENO3	18F3MET5(0.99,1), 19C11MET6(0.99,1), 19C12MET6(0.99,1), 23C5MET9(0.99,1), 4H4PCA4(0.99,1)
gi7669492	2597	GAPD	13D4MET1(1,1), 18F3MET5(1,4), 3A8PCA3(1,8), 3F5PCA3(1,8), 4H4PCA4(1,4)	gi4504183	2950	GSTP1	19C11MET6(1,6), 4H4PCA4(0.99,1), 7A12PCA6(1,2)
gi4504345	3040	HBA2	13B3MET1(0.98,1), 13D4MET1(1,5), 18F3MET5(0.97,3), 19C11MET6(1,6), 19C12MET6(1,9), 19D12MET6(1,6), 23C5MET9(1,3), 23H3MET9(0.99,1), 3A8PCA3(1,2), 3F5PCA3(0.98,1)	gi4504349	3043	HBB	13D4MET1(1,8), 18F3MET5(1,12), 19C11MET6(1,11), 19C12MET6(1,11), 19D12MET6(1,8), 23C5MET9(1,4), 23H3MET9(0.99,1), 3A8PCA3(1,12), 3F5PCA3(1,19), 4H4PCA4(1,4), 7A12PCA6(1,2)
gi28302131	3047	HBG1	13D4MET1(0.99,1), 19C11MET6(0.97,1), 19C12MET6(0.99,1), 19D12MET6(0.99,1), 3F5PCA3(0.91,1)	gi10645195	3012	HIST 1H2AE	13B3MET1(1,3), 13D4MET1(0.99,1), 18F3MET5(1,3), 19C11MET6(1,2), 22C5MET8(1,3), 2B3PCA2(1,4), 3A8PCA3(1,7), 3F5PCA3(1,7), 4H4PCA4(1,6), 7A12PCA6(1,2)
gi11415030	8363	HIST 1H4J	13B3MET1(1,3), 13H10MET1(1,3), 2B3PCA2(0.99,1), 3F5PCA3(0.98,1)	gi4885371	3005	H1F0	1B5PCA1(0.92,1), 2B3PCA2(1,1)
gi24638446	8338	HIST 2H2AC	18F3MET5(0.9,1), 3A8PCA3(1,2), 3F5PCA3(1,2)	gi5031753	3187	HNRPH1	18F3MET5(1,2), 19C11MET6(1,3), 19C12MET6(0.99,1), 23C5MET9(0.99,1), 4H4PCA4(0.99,1), 7A12PCA6(1,2)
gi4885431	3304	HSPA1B	19C11MET6(1,5), 19C12MET6(1,3), 23C5MET9(1,3), 23H3MET9(0.98,1), 3A8PCA3(0.96,1)	gi28178832	3418	IDH2	13D4MET1(0.96,1), 3A8PCA3(1,10), 3F5PCA3(1,9)
gi51460541	22979	KIAA0953	19D12MET6(1,1), 1B5PCA1(0.97,1), 22C5MET8(0.98,2), 23H3MET9(0.96,1), 6D9PCA5(0.98,1)	gi8922712	55752	NA	18F3MET5(1,4), 3F5PCA3(1,3), 4H4PCA4(1,6)
gi12056473	54187	NANS	19C11MET6(1,3), 19C12MET6(1,15), 19D12MET6(1,4), 3F5PCA3(0.99,1), 7A12PCA6(1,4)	gi37655183	10397	NDRG1	23C5MET9(1,7), 23H3MET9(1,3)
gi51465474	346085	None	18F3MET5(0.95,1), 3A8PCA3(1,2), 3F5PCA3(0.94,1), 4H4PCA4(0.99,1)	gi51464772	389268	None	14A8MET2(0.96,1), 1B5PCA1(0.98,1), 2F2PCA2(0.99,1), 6D9PCA5(0.97,1)
gi10835121	5313	PKLR	19C11MET6(0.95,1), 3A8PCA3(0.98,1)	gi33286418	5315	PKM2	13D4MET1(1,3), 18F3MET5(1,7), 19C11MET6(1,7), 3A8PCA3(1,20), 3F5PCA3(1,28), 4H4PCA4(1,5)
gi5453549	10549	PRDX4	19C11MET6(1,6), 19C12MET6(1,3), 7A12PCA6(1,2)	gi4758638	9588	PRDX6	18F3MET5(1,3), 19C11MET6(1,6), 3F5PCA3(1,6)
gi4506605	9349	RPL23	13B3MET1(0.92,1), 2B3PCA2(0.99,1), 3F5PCA3(0.99,1)	gi4506753	8607	RUVBL1	19C11MET6(1,7), 7A12PCA6(1,8)
gi4506773	6280	S100A9	19C11MET6(1,9), 19C12MET6(1,7), 4H4PCA4(0.93,1), 7A12PCA6(1,2)	gi4507813	7358	UGDH	19C11MET6(1,3), 19C12MET6(0.99,1), 3F5PCA3(1,2)
gi46593007	7384	UQCRC1	19C11MET6(1,11), 19C12MET6(1,7), 19D12MET6(1,3), 7A12PCA6(1,5)	gi4503571	2023	ENO1	18F3MET5(1,6), 19C11MET6(1,6), 23C5MET9(1,3), 3A8PCA3(0.96,1), 3F5PCA3(0.99,1), 7A12PCA6(0.97,1)

Table 4.8: The final post-subtraction compendium of proteins suspected of eliciting the humoral response profiles of the 20-fraction predictor. Proteins with increased humoral response in prostate cancer are highlighted.

Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)	Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)
gi7656849	27034	**ACAD8**	3F5PCA3(0.97,1)	gi47132622	4507	MTAP	18F3MET5(0.97,1)
gi6382064	55	**ACPP**	7A12PCA6(1,8)	gi21956645	136319	MTPN	22C5MET8(1,2)
gi4502157	341	**APOC1**	3A8PCA3(1,6) 3F5PCA3(1,5)	gi33620747	50804	MYEF2	18F3MET5(0.99,2)
gi33946285	330	**BIRC3**	2F2PCA2(0.9,1)	gi21361478	23157	NA	4H4PCA4(1,6)
gi4557395	760	**CA2**	3A8PCA3(1,6) 3F5PCA3(1,6)	gi21945064	151011	NA	4H4PCA4(0.99,1)
gi26787970	1041	**CDSN**	3H11PCA3(0.97,1)	gi27262628	4678	NASP	22C5MET8(1,1)
gi13994236	27439	**CECR6**	1B5PCA1(0.91,1)	gi4505365	4712	NDUFB6	13H10MET1(1,4)
gi38045913	4830	**NME1**	3F5PCA3(0.99,1) 7A12PCA6(1,2)	gi10835025	4729	NDUFV2	18F3MET5(0.95,2)
gi7705676	51050	**PI15**	3A8PCA3(1,2) 3F5PCA3(0.97,1)	gi4885063	230	ALDOC	13H10MET1(1,2) 18F3MET5(1,9) 4H4PCA4(1,7)
gi4506031	5538	**PPT1**	7A12PCA6(0.96,1)	gi8923444	55651	NOLA2	13B3MET1(0.99,1)
gi4557363	639	**PRDM1**	1B5PCA1(0.98,1) 2F2PCA2(0.98,1)	gi41150140	388076	None	13B3MET1(0.96,1)
gi5453629	10540	**DCTN2**	23H3MET9(1,5)	gi30090003	52	ACP1	22C5MET8(0.99,2)
gi21536452	5646	**PRSS3**	3A8PCA3(1,5)	gi41058276	388642	None	13D4MET1(1,1) 18F3MET5(0.98,1) 19C11MET6(1,5) 19C12MET6(1,9) 19D12MET6(1,6)
gi4503643	2153	**F5**	1B5PCA1(0.93,1)	gi4826972	9939	RBM8A	22C5MET8(0.99,1)
gi4502961	1294	**COL7A1**	3F5PCA3(0.99,1)	gi6005848	11201	POLI	22C5MET8(0.98,1)
gi21361120	1264	**CNN1**	3A8PCA3(1,2) 3F5PCA3(0.99,1)	gi5453974	5586	PKN2	22C5MET8(1,2)
gi51464135	401124	**FLJ16686**	1B5PCA1(0.91,1)	gi15431290	6135	RPL11	13B3MET1(1,4)
gi21314644	10818	**FRS2**	1B5PCA1(0.91,1)	gi16753227	6128	RPL6	13B3MET1(1,2)
gi7382458	5657	**PRTN3**	2F2PCA2(0.93,1)	gi4503379	1809	DPYSL3	2B3PCA2(0.98,1)
gi7657116	26330	**GAPDS**	3A8PCA3(0.94,1) 3F5PCA3(0.9,1)	gi4506675	6184	RPN1	18F3MET5(1,2
gi4885281	2746	**GLUD1**	3F5PCA3(1,2)	gi4506685	6207	RPS13	13B3MET1(1,6)
gi31377775	2747	**GLUD2**	3F5PCA3(0.99,1)	gi5032051	6208	RPS14	13B3MET1(1,6) 2B3PCA2(1,2)
gi5031755	10236	**HNRPR**	3A8PCA3(1,2)	gi5902102	6632	SNRPD1	13B3MET1(0.99,1)
gi40254873	29062	**HSPC049**	1B5PCA1(0.93,1) 2F2PCA2(0.9,1)	gi4507155	6652	SORD	13D4MET1(0.97,1)
gi38202209	55777	**MBD5**	2F2PCA2(0.9,1)	gi29788785	203068	TUBB	22C5MET8(1,1)
gi9910244	56945	**MRPS22**	3F5PCA3(1,9)	gi20127460	7508	XPC	13B3MET1(1,3)
gi27734781	286514	**MGC33889**	2F2PCA2(0.9,1)	gi12711674	65109	UPF3B	13B3MET1(1,3)
gi5453832	10525	**HYOU1**	23H3MET9(1,2)	gi21361495	28970	PTD012	4H4PCA4(1,5)
gi42822880	153364	**LOC153364**	3F5PCA3(0.99,1)	gi7524354	23564	DDAH2	19C11MET6(1,7) 19C12MET6(1,11) 19D12MET6(1,6)
gi21362062	84328	**LZIC**	23H3MET9(0.98,1)	gi4507645	7167	TPI1	13D4MET1(1,3) 18F3MET5(1,2) 19C11MET6(1,2) 19C12MET6(1,6) 19D12MET6(1,6)
gi31543657	10948	**STARD3**	1B5PCA1(0.92,1)	gi4503423	1854	DUT	13B3MET1(1,2)
gi4506181	5683	**PSMA2**	7A12PCA6(0.98,1)	gi25777602	5708	PSMD2	22C5MET8(1,2)
gi14165464	5725	**PTBP1**	3F5PCA3(0.99,1)	gi30240932	10938	EHD1	22C5MET8(1,7)
gi5360210	5819	**PVRL2**	2F2PCA2(0.99,1)	gi15150811	92259	MRPS36	13B3MET1(1,2)
gi7657486	27089	**QP-C**	3A8PCA3(1,3) 3F5PCA3(1,3)	gi13994257	54460	MRPS21	13B3MET1(0.99,1)
gi16753233	7094	**TLN1**	3A8PCA3(1,3) 3F5PCA3(1,3)	gi27436901	6182	MRPL12	23C5MET9(1,2)
gi23397427	10492	**SYNCRIP**	3F5PCA3(0.93,1)	gi13489091	4357	MPST	4H4PCA4(0.99,1)
gi22035665	83660	**TLN2**	1B5PCA1(0.93,1) 2F2PCA2(0.9,1)	gi11125772	4082	MARCKS	13B3MET1(0.99,1) 22C5MET8(1,1)
gi25777692	23087	**TRIM35**	1B5PCA1(0.9,1)	gi24308009	23173	METAP1	4H4PCA4(1,3)
gi11024714	7314	**UBB**	3A8PCA3(1,2) 3F5PCA3(0.98,2)	gi21361809	494115	LOC494115	13B3MET1(1,8) 22C5MET8(1,2)
gi51458553	23358	**USP24**	1B5PCA1(0.92,1)	gi4557657	3396	ICT1	13B3MET1(1,2)
gi6005942	7415	**VCP**	23H3MET9(1,8)	gi33413400	2098	ESD	4H4PCA4(0.99,1)
gi18959210	147179	**WIRE**	2F2PCA2(0.9,1)	gi4757702	9470	EIF4E2	4H4PCA4(0.99,1)
gi41151132	90649	**ZNF486**	1B5PCA1(1,1)	gi4758862	9521	EEF1E1	13D4MET1(1,1)
gi7661646	25996	**DKFZP566E144**	7A12PCA6(0.93,1)	gi23110942	5686	PSMA5	23C5MET9(1,2)
gi20070255	26286	ARFGAP3	22C5MET8(0.99,1)	gi51475270	388823	None	3H11PCA3(0.96,1)
gi4885079	509	ATP5C1	23C5MET9(0.99,1)	gi41151603	389842	None	22C5MET8(0.99,1)
gi4502297	513	ATP5D	23C5MET9(1,10)	gi51493118	390407	None	2B3PCA2(0.95,1)
gi6005717	521	ATP5I	22C5MET8(1,4)	gi51472614	390531	None	4H4PCA4(1,2)
gi18644883	522	ATP5J	13B3MET1(1,7) 22C5MET8(1,2)	gi41150478	390712	None	4H4PCA4(0.99,1)
gi4502303	539	ATP5O	13D4MET1(1,4)	gi42657272	401206	None	13B3MET1(0.91,1)
gi32171238	55971	BAIAP2L1	22C5MET8(1,1)	gi51475407	440786	None	13B3MET1(0.99,1)
gi37574726	637	BID	23C5MET9(0.98,1)	gi7706254	51602	NOP5/NOP58	2B3PCA2(0.98,1)
gi23200008	83451	ABHD11	13D4MET1(1,2)	gi4505449	4852	NPY	13B3MET1(1,6) 22C5MET8(1,7)
gi4502419	645	BLVRB	18F3MET5(0.98,3)	gi34098946	4904	NSEP1	13B3MET1(1,4)
gi5802976	10974	C10orf116	13D4MET1(1,1)	gi7657033	30833	NT5C	4H4PCA4(0.99,1)
gi5803013	10961	C12orf8	22C5MET8(0.97,1)	gi21264365	4928	NUP98	22C5MET8(1,1)
gi4502503	722	C4BPA	22C5MET8(0.99,1)	gi8923427	54940	OCIAD1	18F3MET5(1,8) 4H4PCA4(1,5)
gi11641247	152007	C9orf19	13B3MET1(1,2)	gi45238849	5042	PABPC3	18F3MET5(1,4)
gi4502009	9049	AIP	4H4PCA4(0.95,1)	gi4504715	8761	PABPC4	18F3MET5(1,3)
gi13124881	865	CBFB	22C5MET8(1,1)	gi4557042	5064	PALM	22C5MET8(1,2)
gi19923356	10602	CDC42EP3	22C5MET8(0.91,1)	gi31543380	11315	PARK7	13D4MET1(1,1) 18F3MET5(0.99,1)
gi11095441	4329	ALDH6A1	13D4MET1(1,1)	gi4505621	5037	PBP	13B3MET1(1,2)

Continued on next page

Continued from last page

Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)	Accession Number	Entrez Gene	Symbol	Fraction (Probability, Peptide Count)
gi40354205	229	ALDOB	18F3MET5(1,4)	gi14141166	5094	PCBP2	13B3MET1(0.98,1)
gi5031635	1072	CFL1	13B3MET1(1,4) 22C5MET8(1,4)	gi4885539	5110	PCMT1	18F3MET5(1,4)
gi14719392	1073	CFL2	13B3MET1(1,2) 22C5MET8(0.99,1)	gi4505651	5833	PCYT2	4H4PCA4(1,6)
gi25092725	27341	CGI-96	22C5MET8(0.97,1)	gi42476169	54623	PD2	22C5MET8(0.99,1)
gi4502801	1104	CHC1	4H4PCA4(0.99,1)	gi7656883	27295	PDLIM3	13B3MET1(0.99,1)
gi4502805	1113	CHGA	22C5MET8(1,1)	gi12408675	5202	PFDN2	18F3MET5(0.99,2)
gi19920317	10970	CKAP4	23C5MET9(1,2)	gi4826898	5216	PFN1	13B3MET1(0.99,1)
gi21536286	1152	CKB	13D4MET1(1,1) 22C5MET8(1,2)	gi4505753	5223	PGAM1	13D4MET1(1,4)
gi8923900	55907	CMAS	4H4PCA4(0.94,1)	gi22165431	311	ANXA11	13D4MET1(1,3)
gi17402875	1292	COL6A2	18F3MET5(0.99,2) 4H4PCA4(0.95,1)	gi21359873	5347	PLK1	14A8MET2(0.92,1)
gi5902134	11151	CORO1A	4H4PCA4(1,6)	gi5453930	5438	POLR2I	22C5MET8(1,1)
gi5174675	10321	CRISP3	13B3MET1(0.99,1) 22C5MET8(0.94,1)	gi13699256	5511	PPP1R8	22C5MET8(1,5) 22C5MET8(1,5)
gi20070160	8531	CSDA	13B3MET1(1,3)	gi31083236	5527	PPP2R5C	22C5MET8(0.98,1)
gi4503143	1509	CTSD	19C11MET6(1,10) 19C12MET6(1,13) 19D12MET6(1,16)	gi4502133	325	APCS	18F3MET5(0.99,2)
gi41281768	1528	CYB5	22C5MET8(1,7)	gi18375501	328	APEX1	22C5MET8(1,1)
gi11128019	54205	CYCS	13B3MET1(0.97,1)	gi32189392	7001	PRDX2	19C11MET6(1,4) 19C12MET6(1,12) 19D12MET6(1,4)
gi21361670	28988	DBNL	22C5MET8(1,3)	gi41349454	5549	PRELP	18F3MET5(0.99,4) 4H4PCA4(1,16)
gi7657056	30845	EHD3	22C5MET8(1,2)	gi40068475	5756	PTK9	4H4PCA4(1,2)
gi4503607	2108	ETFA	13D4MET1(1,1)	gi4506413	5906	RAP1A	4H4PCA4(0.97,1)
gi4557581	2171	FABP5	13B3MET1(0.97,1)	gi40354214	5948	RBP2	22C5MET8(0.99,1)
gi7661714	10447	FAM3C	13D4MET1(1,1)	gi6005854	11331	REA	13D4MET1(1,1)
gi4758356	2237	FEN1	13D4MET1(0.93,1)	gi24307923	6120	RPE	18F3MET5(1,11) 4H4PCA4(1,5)
gi4503745	2316	FLNA	4H4PCA4(1,4)	gi4506621	6154	RPL26	2B3PCA2(0.94,1)
gi5031699	10211	FLOT1	13D4MET1(1,10)	gi4506649	6122	RPL3	2B3PCA2(0.99,1)
gi34577057	23769	FLRT1	14A8MET2(0.94,1)	gi16579885	6124	RPL4	13B3MET1(0.99,1)
gi16933542	2335	FN1	22C5MET8(1,4)	gi14591909	6125	RPL5	13B3MET1(1,2)
gi17402900	8880	FUBP1	13B3MET1(1,3)	gi4506661	6130	RPL7A	13B3MET1(1,2)
gi4826734	2521	FUS	13B3MET1(1,3)	gi15431306	6132	RPL8	2B3PCA2(0.99,1)
gi38202257	23193	GANAB	13B3MET1(1,2)	gi15431303	6133	RPL9	22C5MET8(1,6)
gi21361657	2923	GRP58	22C5MET8(1,3)	gi4506687	6209	RPS15	13B3MET1(1,4)
gi7705704	373156	GSTK1	13D4MET1(1,2)	gi4506693	6218	RPS17	22C5MET8(1,6)
gi20357599	94239	H2AFV	4H4PCA4(0.99,1)	gi11968182	6222	RPS18	13B3MET1(1,3)
gi4504327	3032	HADHB	18F3MET5(0.93,2) 4H4PCA4(1,3)	gi4506695	6223	RPS19	13B3MET1(1,2)
gi4885403	3054	HCFC1	13B3MET1(1,2)	gi15055539	6187	RPS2	13H10MET1(1,2)
gi4885413	3094	HINT1	13B3MET1(0.99,1)	gi4506697	6224	RPS20	13B3MET1(1,2)
gi4885375	3006	HIST 1H1C	13B3MET1(1,6) 2B3PCA2(1,2)	gi4506725	6191	RPS4X	13B3MET1(1,3) 13D4MET1(0.99,1)
gi20544168	3010	HIST 1H1T	13B3MET1(1,4)	gi4506741	6201	RPS7	22C5MET8(1,1)
gi10800140	3018	HIST 1H2BB	13H10MET1(1,2) 2B3PCA2(0.96,1)	gi4759068	6341	SCO1	22C5MET8(1,1)
gi20270186	9324	HMGN3	13D4MET1(0.95,1)	gi14141195	6388	SDF2	13B3MET1(1,3)
gi5803036	10949	HNRPA0	13B3MET1(0.99,1)	gi4759080	6389	SDHA	4H4PCA4(0.99,1)
gi14043070	3178	HNRPA1	13B3MET1(1,4)	gi32454741	871	SERPINH1	18F3MET5(0.98,3) 22C5MET8(1,4)
gi14043072	3181	HNRPA2B1	13B3MET1(1,35) 22C5MET8(1,7)	gi4506891	6418	SET	22C5MET8(0.99,1)
gi34740329	220988	HNRPA3	13B3MET1(0.96,1)	gi4506901	6428	SFRS3	13B3MET1(0.98,1) 22C5MET8(0.99,1)
gi14110428	3183	HNRPC	13B3MET1(1,7) 13B3MET1(0.99,1) 22C5MET8(1,1)	gi4506903	8683	SFRS9	22C5MET8(0.98,1)
gi14110414	3184	HNRPD	13B3MET1(1,2) 22C5MET8(1,7)	gi13775198	83442	SH3BGRL3	23C5MET9(0.97,1)
gi14110407	9987	HNRPDL	22C5MET8(0.98,1)	gi23397666	25942	SIN3A	22C5MET8(0.99,1)
gi4826760	3185	HNRPF	22C5MET8(1,3)	gi7657431	27044	SND1	22C5MET8(1,2)
gi14141157	3189	HNRPH3	18F3MET5(1,4)	gi4759156	6626	SNRPA	13D4MET1(1,16) 14A8MET2(0.97,1)
gi14165435	3190	HNRPK	13B3MET1(0.99,1) 22C5MET8(1,3)	gi38149981	6629	SNRPB2	13D4MET1(0.97,1)
gi14141161	3192	HNRPU	13B3MET1(0.99,1)	gi4759160	6634	SNRPD3	13B3MET1(1,2)
gi40018640	79577	HRPT2	18F3MET5(1,6)	gi14741936	414153	SNRPEL1	13D4MET1(0.99,1)
gi4504505	3295	HSD17B4	18F3MET5(1,2)	gi13027644	6638	SNRPN	13B3MET1(1,3)
gi13676857	3306	HSPA2	19C11MET6(1,4) 19C12MET6(1,2) 19D12MET6(1,3) 23C5MET9(1,2)	gi4507149	6647	SOD1	22C5MET8(1,2) 6D9PCA5(1,1)
gi24234686	3312	HSPA8	19C11MET6(1,9) 19C12MET6(1,4) 19D12MET6(1,2)	gi10835187	6648	SOD2	13D4MET1(1,2)
gi7657015	51493	HSPC117	4H4PCA4(1,21)	gi31543653	6727	SRP14	13B3MET1(0.96,1)
gi4557663	3476	IGBP1	23C5MET9(0.98,1)	gi4507357	8407	TAGLN2	13D4MET1(1,1)
gi31542984	3700	ITIH4	4H4PCA4(1,2)	gi4557871	7018	TF	4H4PCA4(1,13)
gi4504865	8570	KHSRP	13B3MET1(1,3)	gi4758152	1678	TIMM8A	22C5MET8(1,2)
gi4502173	354	KLK3	22C5MET8(1,2) 3A8PCA3(1,2) 3F5PCA3(1,2) 4H4PCA4(1,3)	gi42518065	9414	TJP2	22C5MET8(0.99,1)
gi5031887	4026	LPP	22C5MET8(0.99,1) 6D9PCA5(1,1)	gi39725636	54732	TMED9	19C11MET6(1,3) 19C12MET6(1,2) 19D12MET6(0.98,1)
gi4505047	4060	LUM	4H4PCA4(1,2)	gi19913406	7153	TOP2A	18F3MET5(0.9,1)
gi14195618	4133	MAP2	22C5MET8(1,1)	gi5032179	10155	TRIM28	18F3MET5(1,8) 22C5MET8(1,2)
gi47519639	4134	MAP4	2B3PCA2(1,5)	gi17402907	23650	TRIM29	4H4PCA4(1,10)
gi21735621	4191	MDH2	13D4MET1(0.97,1)	gi21361322	10382	TUBB4	22C5MET8(1,10)
gi4826830	4204	MECP2	2B3PCA2(1,2)	gi14249348	84817	TXNL5	22C5MET8(1,2)
gi21362050	84317	MGC12981	13D4MET1(0.99,1)	gi21361091	7345	UCHL1	4H4PCA4(0.93,1)
gi29788768	347733	MGC8685	22C5MET8(1,3)	gi4827050	9097	USP14	22C5MET8(1,1)
gi15809016	103910	MRLC2	2B3PCA2(1,2)	gi4507855	8078	USP5	22C5MET8(0.94,1)
gi22547129	51264	MRPL27	13B3MET1(0.99,1)	gi17865802	9525	VPS4B	22C5MET8(1,3)

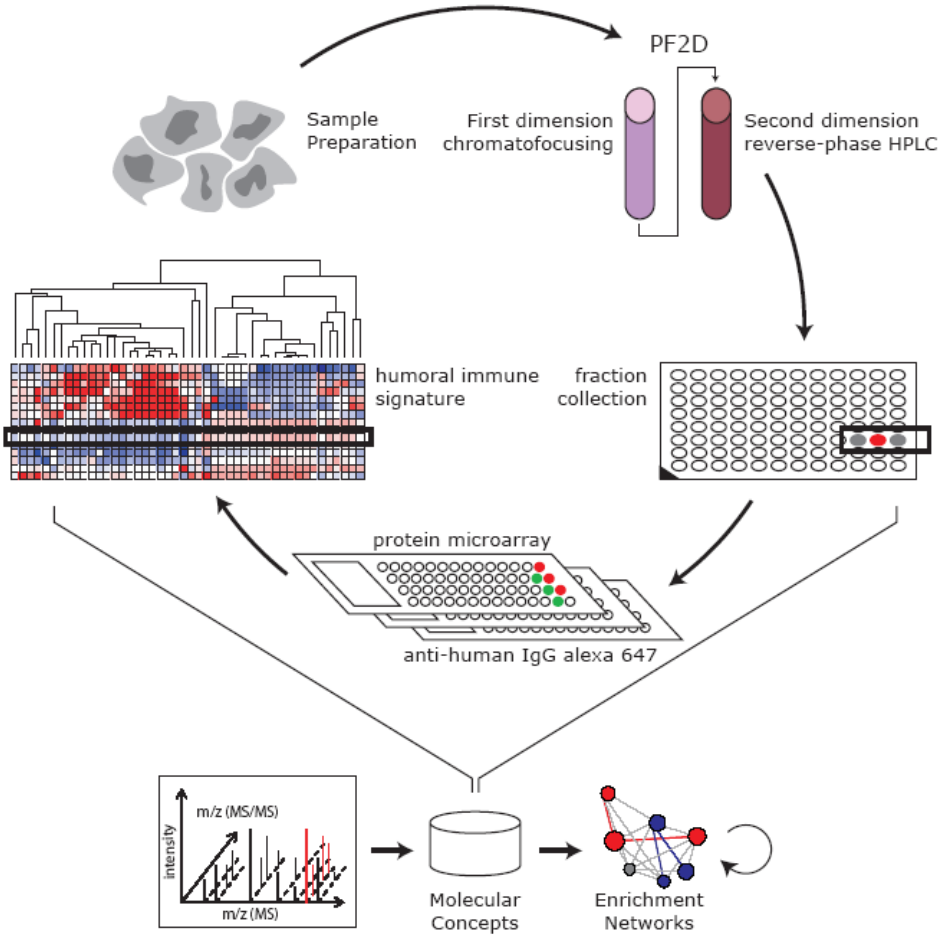

Figure 4.1: Overview of the experimental approach.

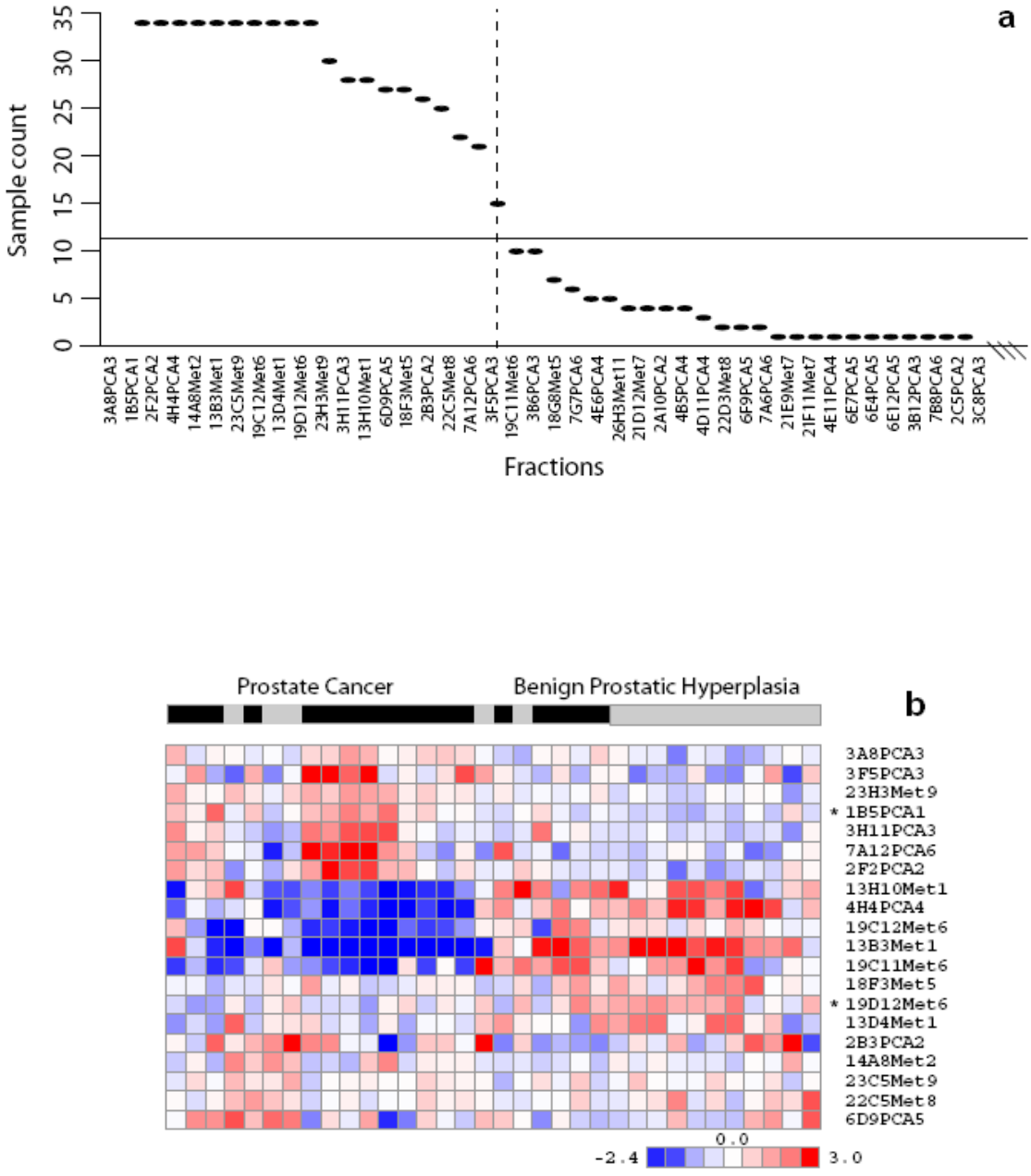

Figure 4.2: The 20-fraction predictor (a) chosen as the minimum fraction count producing maximum accuracy in classification without over-fitting the predictor and a heatmap of the reactivity profile (b) generated by the 20-fraction predictor revealing a distinct bipartite pattern.

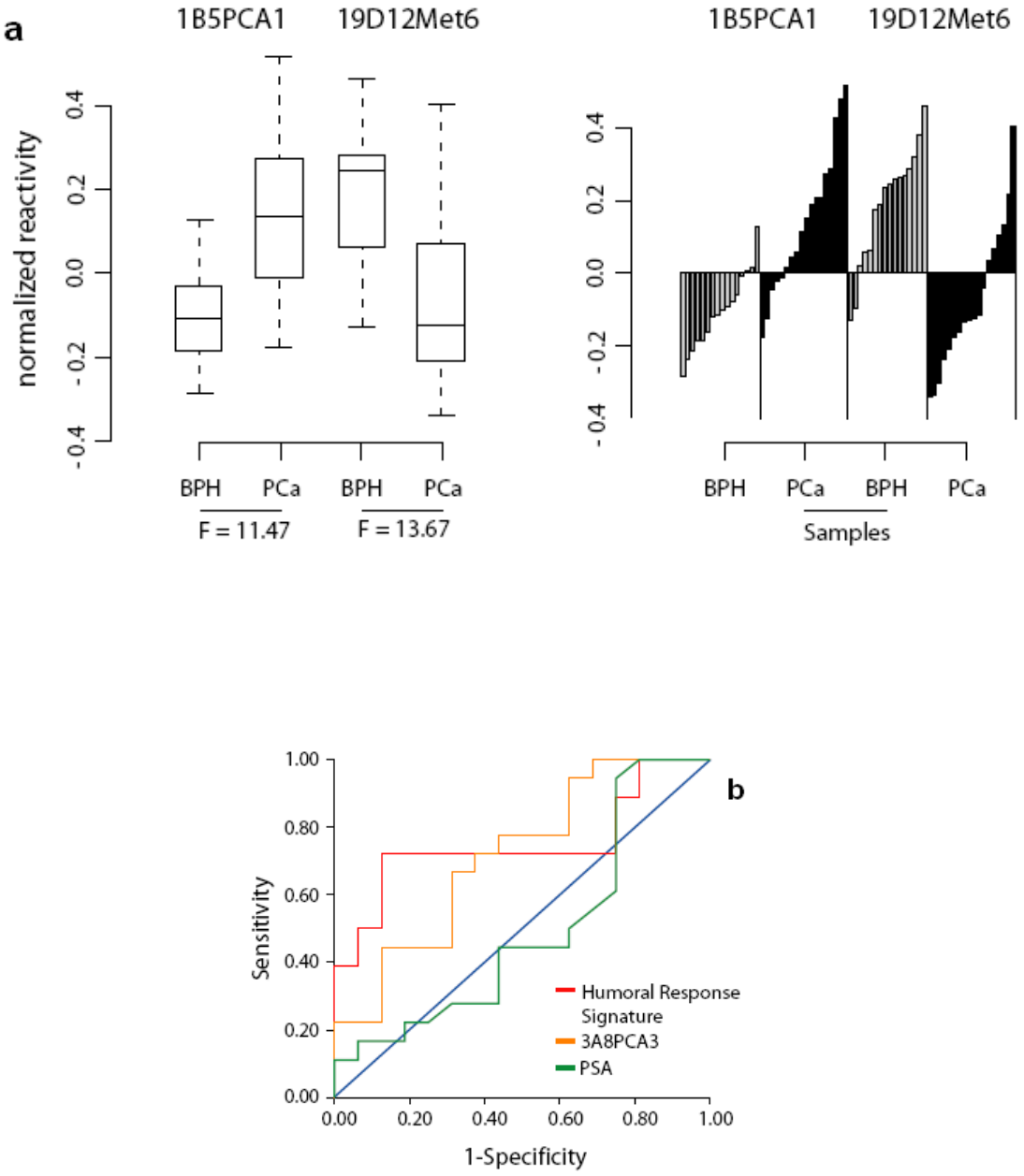

Figure 4.3: Reactivity profile of individual protein markers towards BPH and PCa (a), and (b) shows the receiver-operating characteristic (ROC) curves for the 20-fraction predictor and measured PSA levels in the sample cohort.

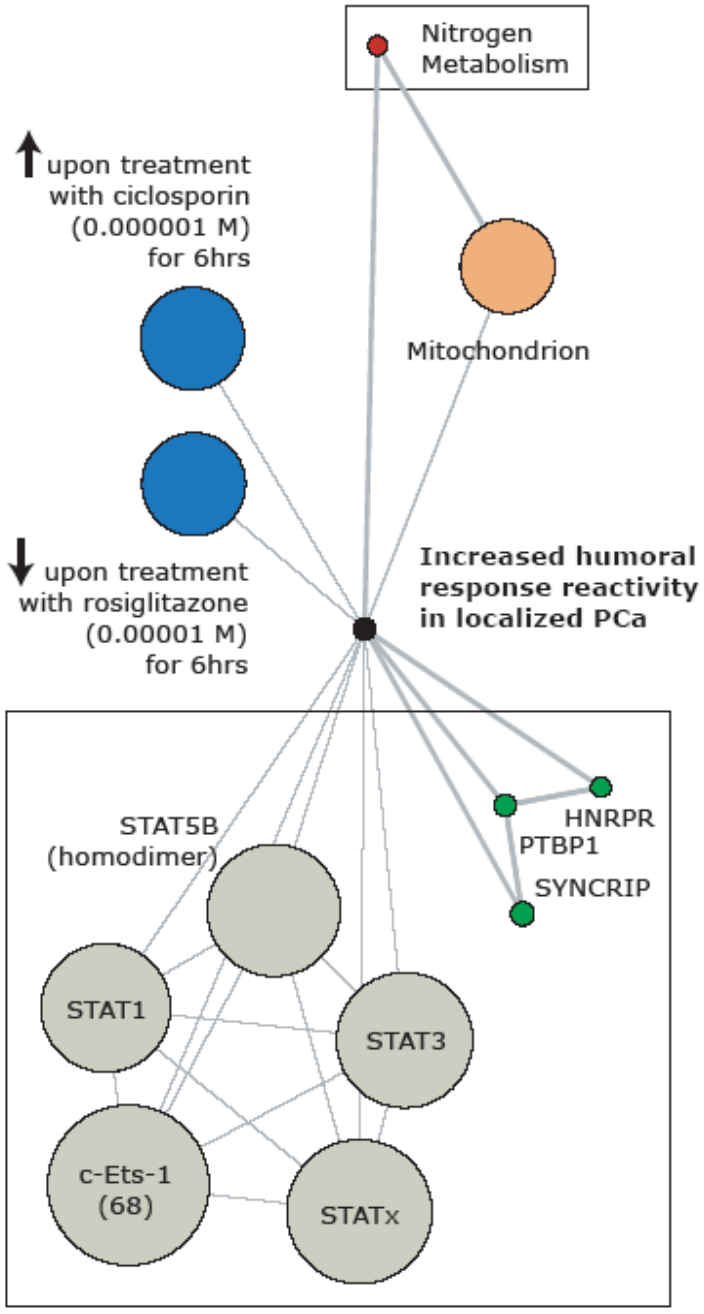

Figure 4.4: Molecular concept analysis (MCM) on the group of five proteins, c-Etc-1(68), STAT1, STAT3, STATx and STAT5B, which were identified to have cancer-specific autoantibody repertoire.

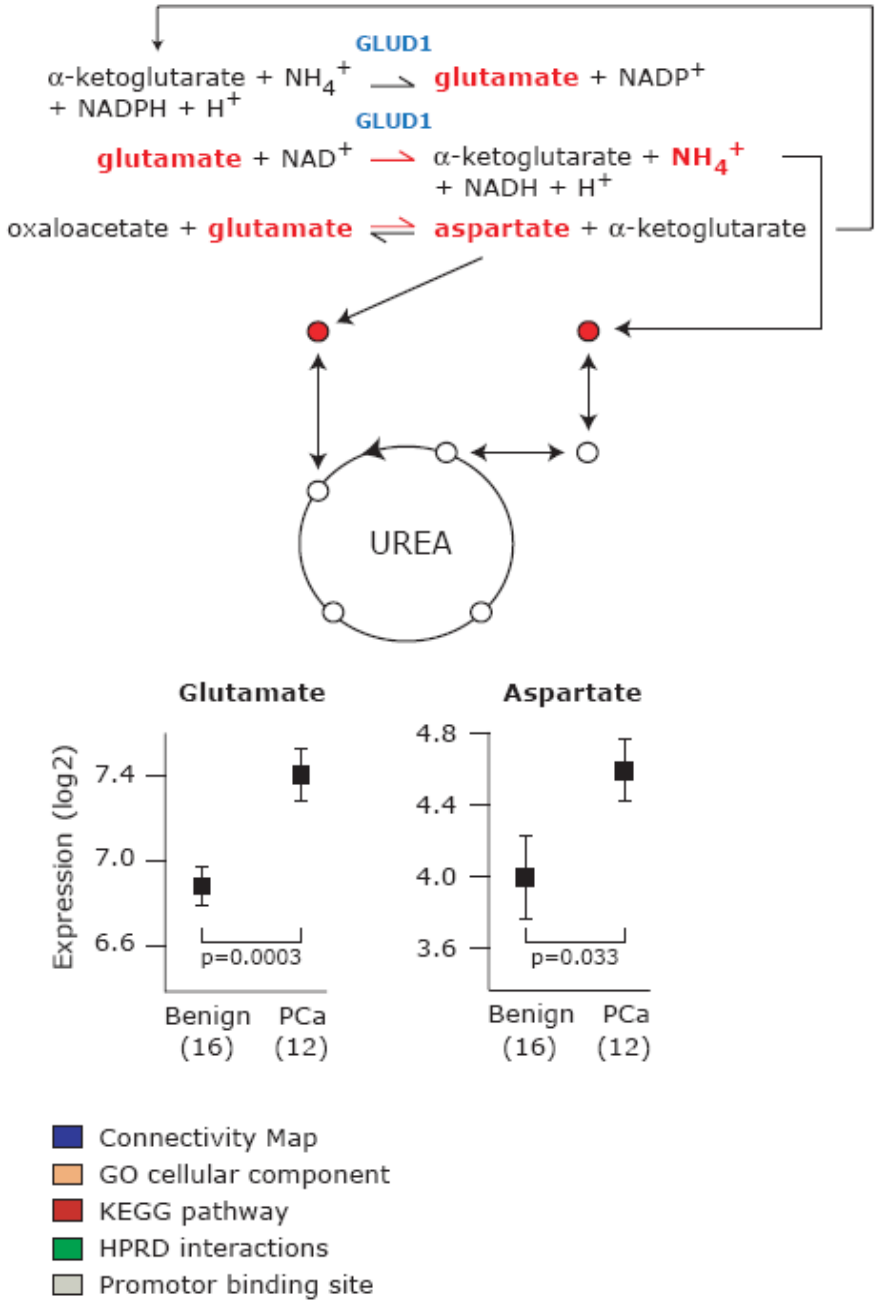

Figure 4.5: Figure showing the nitrogen metabolism concept where increased protein biosynthesis is observed during prostate cancer development. The up-regulation of glutamate as a metabolite coupled to the enhanced enzymatic mediation via humorally reactive GLUD1 shifts chemical equilibrium toward increased ammonia output. Independent metabolomic data shows significantly increased levels of glutamate and aspartate in prostate cancer specimens.

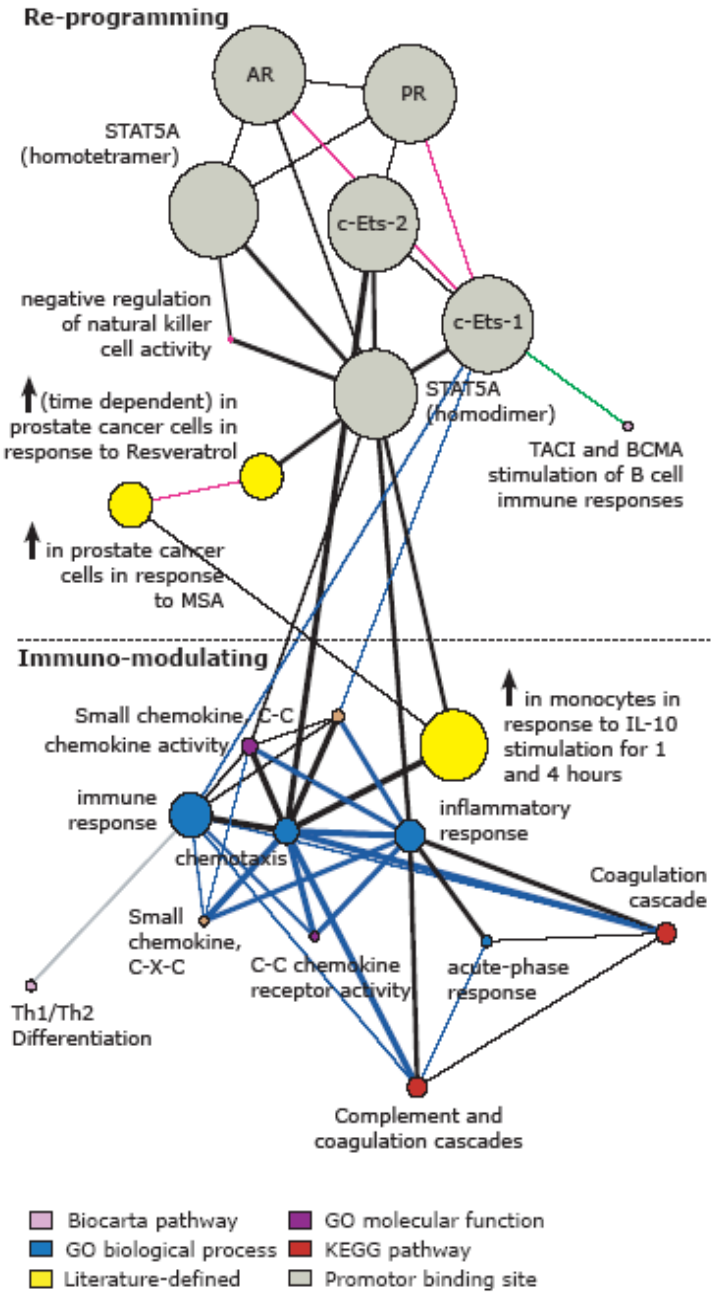

Figure 4.6: Figure showing concepts enriched by the PCa-specific humoral targets including four promoter binding sites implicating the enrichment of a STAT-regulating transcriptome. The five enrichment networks obtained from the earlier analysis were sequentially merged into a single common network shown here.

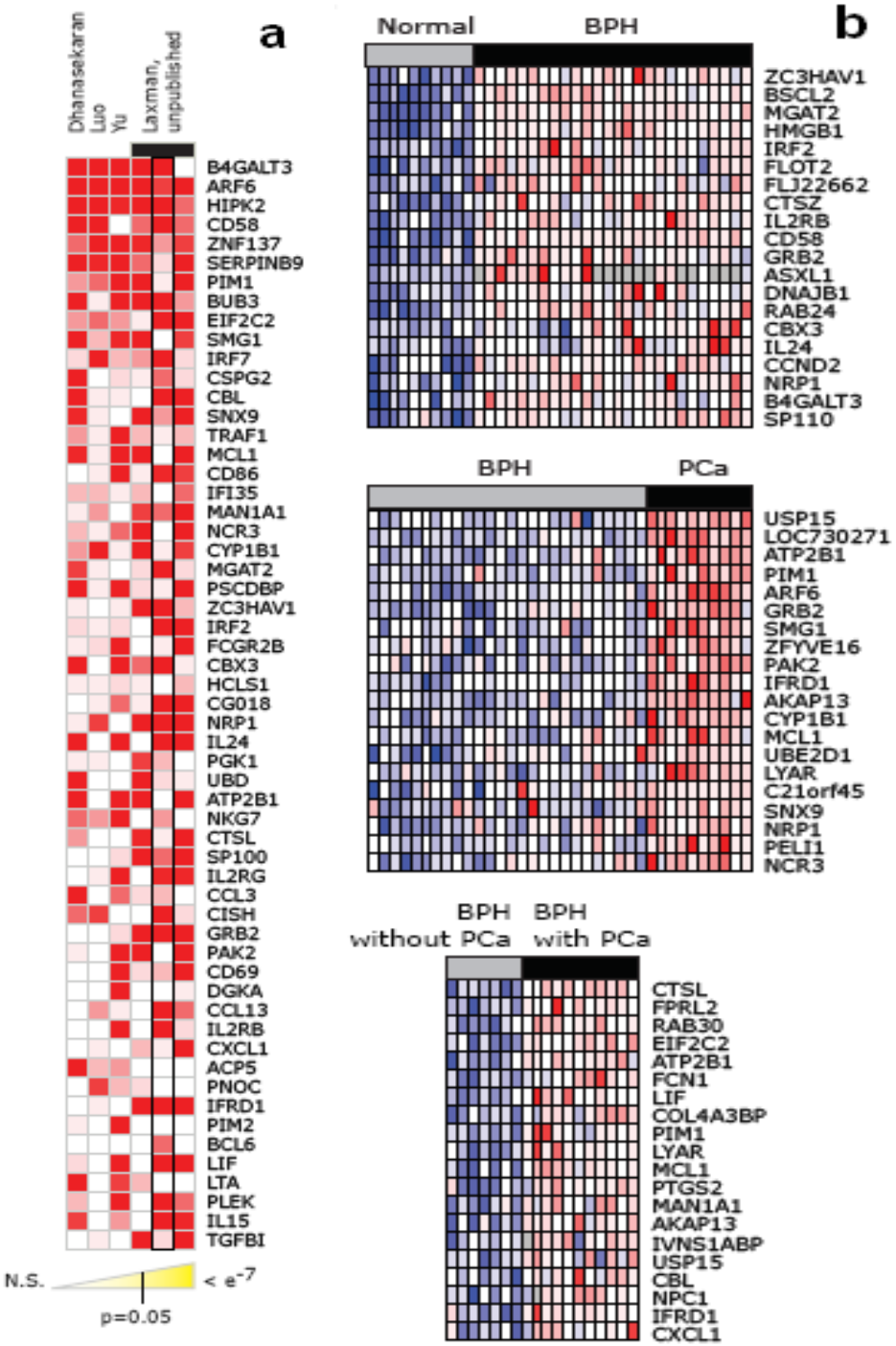

Figure 4.7: Figure showing (a) STAT-regulated immune signature for six prostate cancer gene expression profiling and (b) different humoral response comparisons between normal prostate, BPH or prostate cancer.

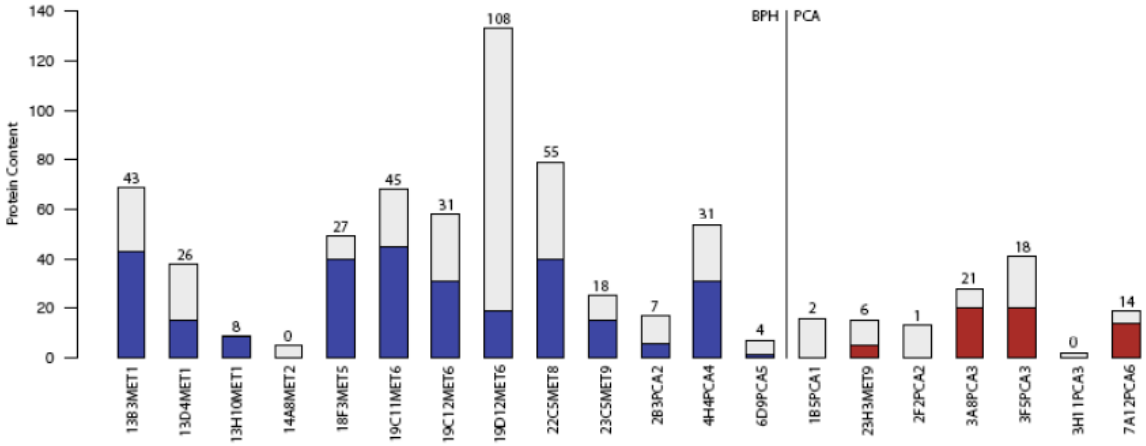

Figure 4.8: Summary of unique protein content across the 20-fraction predictor as sequenced by mass spectrometry. Fractions are stratified and colored by their membership in either of the two dominant humoral response reactivity patterns. The first is increased reactivity in BPH samples relative to PCa (blue), and the second are those with increased reactivity in PCa relative to BPH (red). The colored fraction of each bar represents the percent of total proteins per fraction sequenced with two or more peptides and the integer above each reflects the count of proteins identified with a probability of correct assignment of one.

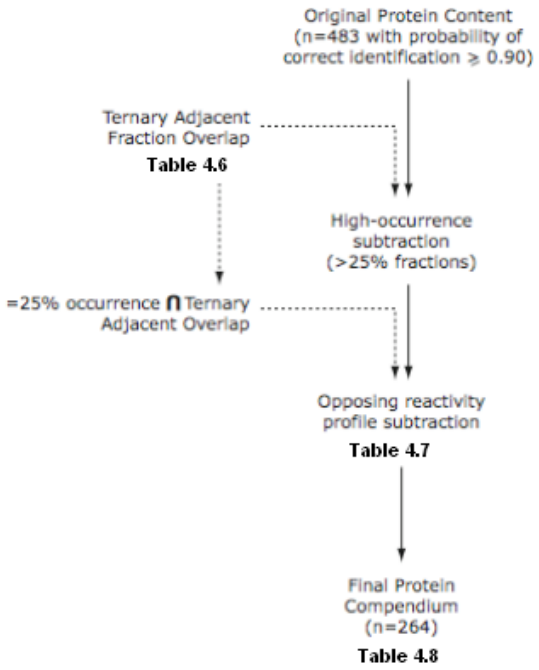

Figure 4.9: Workflow for non-specific protein content removal undertaken in the formulation of the final protein compendium produced from initial mass spectrometry sequencing.

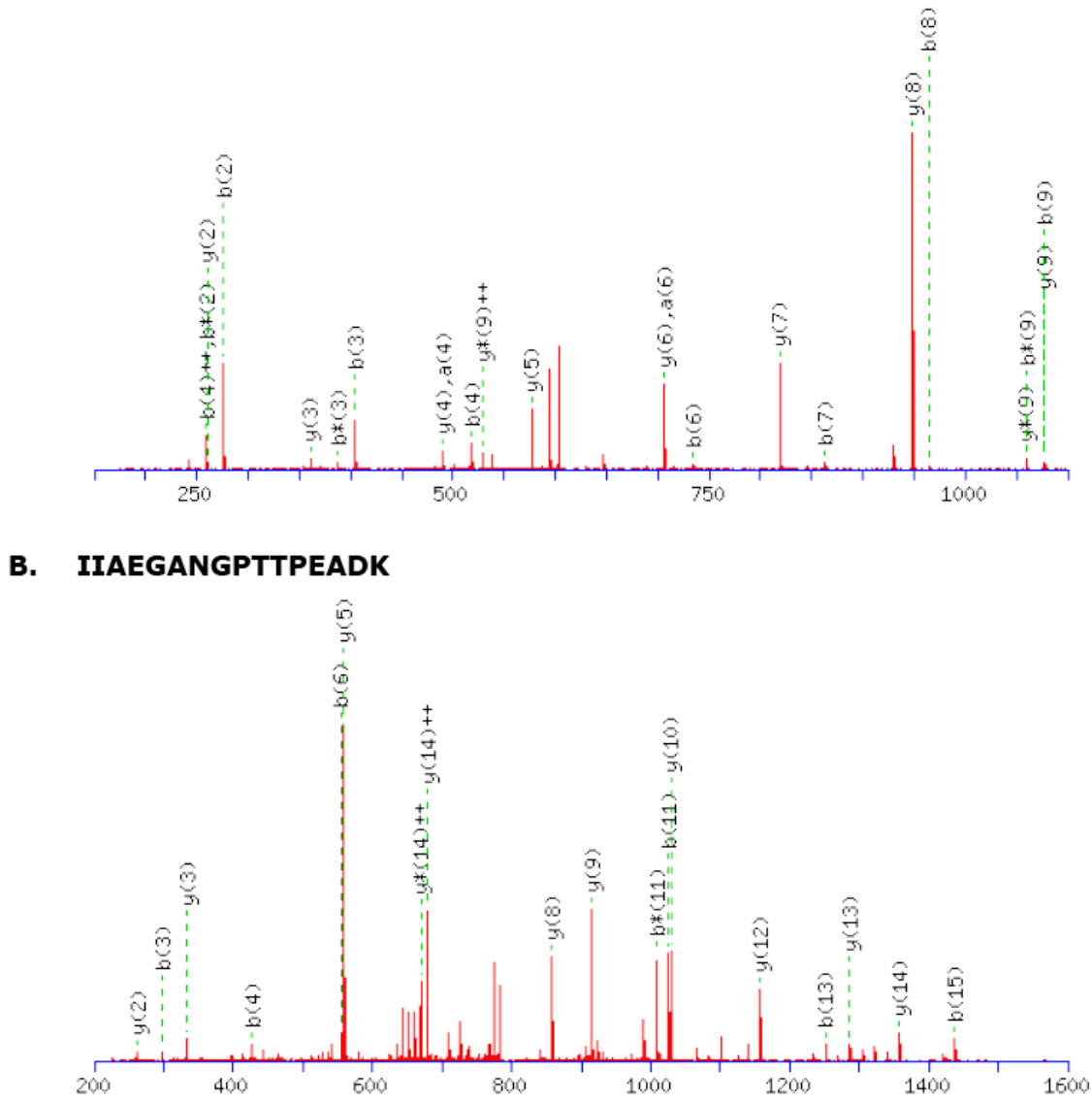

Figure 4.10: MS/MS spectra for identified peptides (A) FQELESETLK (Prostatic acid phosphatase, ACPP; Swissprot ID P15309) from fraction 7A12PCA6 and (B) IIAEGANGPTTPEADK (Glutamate dehydrogenase 1, mitochondrial precursor, GLUD1; Swissprot ID P00367) from fraction 3F5PCA3.

References

[1] Jemal A.; Murray T.; Ward E.; Samuels A.; Tiwari R. C.; Ghafoor A.; Feuer E. J.; and Thun M. J. *CA Cancer J. Clin.*, 55(1):10–30, 2005.

[2] Catalona W. J. *N. Engl. J. Med.*, 331(15):996–1004, 1994.

[3] Jacobsen S. J.; Katusic S. K.; Bergstralh E. J.; Oesterling J. E.; Ohrt D.; Klee G. G.; Chute C. G.; and Lieber M. M. *Jama*, 274(18):1445–9, 1995.

[4] Brown C.; Sauvageot J.; Kahane H.; and Epstein J. I. *Mod. Pathol.*, 9(3):205–9, 1996.

[5] Wulfkuhle J. D.; Liotta L. A.; and Petricoin E. F. *Nat. Rev. Cancer*, 3(4):267–75, 2003.

[6] Anderson N. L.; Polanski M.; Pieper R.; Gatlin T.; Tirumalai R. S.; Conrads T. P.; Veenstra T. D.; Adkins J. N.; Pounds J. G.; Fagan R.; and Lobley A. *Mol. Cell. Proteomics*, 3(4):311–26, 2004.

[7] Brown D. M.; Fisher T. L.; Wei C.; Frelinger J. G.; and Lord E. M. *Immunology*, 102(4):486–97, 2001.

[8] Soussi T. *Cancer Res.*, 60(7):1777–88, 2000.

[9] Brichory F. M.; Misek D. E.; Yim A. M.; Krause M. C.; Giordano T. J.; Beer D. G.; and Hanash S. M. *Proc. Natl. Acad. Sci. USA*, 98(17):9824–9, 2001.

[10] Mintz P. J.; Kim J.; Do K. A.; Wang X.; Zinner R. G.; Cristofanilli M.; Arap M. A.; Hong W. K.; Troncoso P.; Logothetis C. J.; Pasqualini R.; and Arap W. *Nat. Biotechnol.*, 21(1):57–63, 2003.

[11] Nilsson B. O.; Carlsson L.; Larsson A.; and Ronquist G. *Ups. J. Med. Sci.*, 106(1):43–9, 2001.

[12] Sreekumar A.; Laxman B.; Rhodes D. R.; Bhagavathula S.; Harwood J.; Giacherio D.; Ghosh D.; Sanda M. G.; Rubin M. A.; and Chinnaiyan A. M. *J. Natl. Cancer Inst.*, 96(11):834–43, 2004.

[13] McNeel D. G.; Nguyen L. D.; Storer B. E.; Vessella R.; Lange P. H.; and Disis M. L. *J. Urol.*, 164(5):1825–9, 2000.

[14] Wang X.; Yu J.; Sreekumar A.; Varambally S.; Shen R.; Giacherio D.; Mehra R.; Montie J. E.; Pienta K. J.; Sanda M. G.; Kantoff P. W.; Rubin M. A.; Wei J. T.; Ghosh D.; and Chinnaiyan A. M. *N. Engl. J. Med.*, 353(5):1825–9, 2005.

[15] Eisen M. B.; Spellman P. T.; Brown P. O.; and Botstein D. *Proc. Natl. Acad. Sci. USA*, 95(25):14863–8, 1998.

[16] Keller A.; Nesvizhskii A. I.; Kolker E.; and Aebersold R. *Anal. Chem.*, 74(20):5383–92, 2002.

[17] Keller A.; Eng J.; Zhang N.; Li X.-j.; and Aebersold R. *Mol. Syst. Biol.*, 1(2005), 2005.

[18] Keller A.; Kolker E.; Nesvizhskii A. I. and Aebersold R. *Anal. Chem.*, 75(17):4646–58, 2003.

[19] Tomlins S. A.; Mehra R.; Rhodes D. R.; Cao X.; Wang L.; Dhanasekaran S. M.; Kalyana-Sundaram S.; Wei J. T.; Rubin M. A.; Pienta K. J.; Shah R. B.; and Chinnaiyan A. M. *Nat. Genet.*, 39(1):41–51, 2007.

[20] Abbas A. R.; Baldwin D.; Ma Y.; Ouyang W.; Gurney A.; Martin F.; Fong S.; van Lookeren Campagne M.; Godowski P.; Williams P. M.; Chan A. C.; and Clark H. F. *Genes Immun.*, 6(4):319–31, 2005.

[21] von Mensdorff-Pouilly S.; Verstraeten A. A.; Kenemans P.; Snijdewint F. G.; Kok A.; Van Kamp G. J.; Paul M. A.; Van Diest P. J.; Meijer S.; and Hilgers J. *J. Clin. Oncol.*, 18(3):574–83, 2000.

[22] Tomlins S. A.; Mehra R.; Rhodes D. R.; Smith L. R.; Roulston D.; Helgeson B. E.; Cao X.; Wei J. T.; Rubin M. A.; Shah R. B.; and Chinnaiyan A. M. *Cancer Res.*, 66(7):3396–400, 2006.

[23] Tomlins S. A.; Rhodes D. R.; Perner S.; Dhanasekaran S. M.; Mehra R.; Sun X. W.; Varambally S.; Cao X.; Tchinda J.; Kuefer R.; Lee C.; Montie J. E.; Shah R. B.; Pienta K. J.; Rubin M. A.; and Chinnaiyan A. M. *Science*, 310(5748):644–8, 2005.

[24] Kazansky A. V.; Spencer D. M.; and Greenberg N. M. *Cancer Res.*, 63(24):8757–62, 2003.

[25] Aoyagi K.; Shima I.; Wang M.; Hu Y.; Garcia F. U.; and Stearns M. E. *Clin. Cancer Res.*, 4(9):2153–60, 1998.

[26] Lee S. O.; Lou W.; Qureshi K. M.; Mehraein-Ghomi F.; Trump D. L.; and Gao A. C. *Prostate*, 60(4):303–9, 2004.

[27] Dhanasekaran S. M.; Barrette T. R.; Ghosh D.; Shah R.; Varambally S.; Kurachi K.; Pienta K. J.; Rubin M. A.; and Chinnaiyan A. M. *Nature*, 412(6849):822–6, 2001.

[28] Luo J. H.; Yu Y. P.; Cieply K.; Lin F.; Deflavia P.; Dhir R.; Finkelstein S.; Michalopoulos G.; and Becich M. *Mol. Carcinog.*, 33(1):25–35, 2002.

[29] Yu Y. P.; Landsittel D.; Jing L.; Nelson J.; Ren B.; Liu L.; McDonald C.; Thomas R.; Dhir R.; Finkelstein S.; Michalopoulos G.; Becich M.; and Luo J. H. *J. Clin. Oncol.*, 22(14):2790–9, 2004.

[30] Bak J. B.; Landas S. K.; and Haas G. P. *Clin. Prostate Cancer*, 2(2):115–8, 2003.

[31] Yasumoto H.; Matsubara A.; Mutaguchi K.; Usui T.; and McKeehan W. L. *Prostate*, 61(3):236–42, 2004.

[32] Hinsby A. M.; Olsen J. V.; Bennett K. L.; and Mann M. *Mol. Cell. Proteomics*, 2(1):29–36, 2003.

[33] Kwabi-Addo B.; Ozen M.; and Ittmann M. *Endocr. Relat. Cancer*, 11(4):709–24, 2004.

[34] Pendergraft W. F.; 3rd Preston; G. A. Shah; R. R. Tropsha; A. Carter; C. W. Jr.; Jennette J. C.; and Falk R. J. *Nat. Med.*, 10(1):72–9, 2004.

[35] LeJohn H. B.; Cameron L. E.; Yang B.; and Rennie S. L. *J. Biol. Chem.*, 269(6):4523–31, 1994.

[36] Menoret A.; Chandawarkar R. Y.; and Srivastava P. K. *Immunology*, 101(3):364–70, 2000.

[37] Takashima M.; Kuramitsu Y.; Yokoyama Y.; Iizuka N.; Harada T.; Fujimoto M.; Sakaida I.; Okita K.; Oka M.; and Nakamura K. *Proteomics*, 6(13):3894–900, 2006.

[38] Gao B.; Shen X.; Kunos G.; Meng Q.; Goldberg I. D.; Rosen E. M.; and Fan S. *FEBS Lett.*, 488(3):179–84, 2001.

[39] Sela M.; Fuchs S.; and Arnon R. *Biochem. J.*, 85:223–35, 1962.

[40] Zheng S. L.; Liu W.; Wiklund F.; Dimitrov L.; Balter K.; Sun J.; Adami H. O.; Johansson J. E.; Sun J.; Chang B.; Loza M.; Turner A. R.; Bleecker E. R.; Meyers D. A.; Carpten J. D.; Duggan D.; Isaacs W. B.; Xu J.; and Gronberg H. *Prostate*, 66(14):1556–64, 2006.

[41] Yan F.; Sreekumar A.; Laxman B.; Chinnaiyan A. M.; Lubman D. M.; and Barder T. J. *Proteomics*, 3(7):1228–35, 2003.

[42] Bouwman K.; Qiu J.; Zhou H.; Schotanus M.; Mangold L. A.; Vogt R.; Erlandson E.; Trenkle J.; Partin A. W.; Misek D.; Omenn G. S.; Haab B. B.; and Hanash S. *Proteomics*, 3(11):2200–7, 2003.

[43] Chatterjee M.; Mohapatra S.; Ionan A.; Bawa G.; Ali-Fehmi R.; Wang X.; Nowak J.; Ye B.; Nahhas F. A.; Lu K.; Witkin S. S.; Fishman D.; Munkarah A.; Morris R.; Levin N. K.; Shirley N. N.; Tromp G.; Abrams J.; Draghici S.; and Tainsky M. A. *Cancer Res.*, 66(2):1181–90, 2006.

CHAPTER V

Toward high sequence coverage of proteins in human breast cancer cells using on-line monolith-based HPLC-ESI-TOF MS compared to CE MS

5.1 Introduction

An important problem in current proteomics is searching for biomarkers of various diseases [1]. This involves searching for proteins that undergo highly dynamic changes during the course of a disease such as cancer progression. These changes may involve sequence modifications including truncations, deletions, splice variants, and sequence substitutions, as well as posttranslational modifications (PTMs), such as phosphorylation, glycosylation, acetylation, and methylation, and a host of other modifications that may affect the function of the proteins [2-4] and play a significant role in the pathways leading to carcinogenesis [5]. In addition to this complexity, many important proteins involved in regulation are often present in low abundance, presenting further challenges in proteomic studies. Traditionally, 2D-GE [6] has been used to profile protein expression and to search for changes in such expression levels of disease states, including tumor cells.

Despite the tremendous contribution 2D-GE method has made to the study of various proteomes, it has a number of fundamental limitations. These include the

labor-intensive and time-consuming nature of the technique, poor reproducibility, and under-representation of certain classes of proteins, so that truly comprehensive analysis is impossible. Furthermore, it cannot provide accurate M_r information and it still remains difficult to interface 2D-GE directly to MS analysis [7,8]. More recently, a method utilizing 2-D liquid-phase separations based on pI and hydrophobicity of proteins has been developed for mapping of proteins and applied to large-scale study of several different types of human cancer cells [9-11]. This method allows for direct interfacing of chromatographic separation for ESI-TOF MS analysis to obtain an accurate M_r value. Alternatively, the proteins contained in selected pH ranges can be collected, enzymatically digested, and analyzed by MALDI-TOF MS to obtain a peptide map or sequencing information by tandem mass spectrometric analysis. The protein identification can then be readily obtained by database searching.

The use of peptide mapping by MALDI-TOF MS provides very limited sequence coverage, typically, 40-50% at the most, due to the ionization efficiency often depending on the choice of matrix [12], peptide composition [13,14], ion suppression, and suppression or loss of PTMs. Thus, the issue becomes obtaining sufficiently high sequence coverage to prevent false identification of unknown proteins and then to identify the presence of important PTMs or other variations that are often particular to different stages of cancer and other diseases.

In recent work, a method combining CE/ESI-TOF and MALDI-TOF MS was used to improve the sequence coverage of peptide maps [15]. A sequence coverage of over 90% could be obtained in most cases from tryptic digests by combining the peptides

detected from both techniques. The sequence could be used with the M_r value determined by ESI-TOF MS and the known M_r value from the database to determine the sequence coverage and where modifications or isoforms were present. Using this method, various isoforms of proteins found in breast cancer cells could be identified and sequence variations could be detected.

An alternate form of capillary separation involves the use of monolithic columns, which are formed of a single, rigid, and porous polymer. The use of these monolithic columns has been described in many studies involving biological molecules, including nucleic acids, ribonucleic acids, proteins, and peptides [16-22]. It has recently become a stationary phase of choice in LC with several advantages over conventional packed columns. Due to the lack of interstitial space, these monoliths can provide a very fast separation with high resolution. Also, as analyte does not dwell at the resin, it is capable of producing high recovery, hence increased sensitivity, and can separate very hydrophobic analytes that might otherwise be difficult to elute with conventional packed columns. Compared to CE-MS, the method is very rugged where large numbers of separations can be run and the enhanced loadability allows much improved sensitivity.

In the present work, proteins isolated from 2-D liquid separations of breast cancer cell lysates were analyzed by combining several methods to obtain high sequence coverage and to search for isoforms and modified proteins. The method involved the use of capillary monolith LC separation of tryptic digests of proteins isolated from the liquid separations with on-line detection by ESI-TOF MS and linear IT MS. The

sequence coverage obtained by this method was combined with that from MALDI-TOF MS to yield >85% coverage in most cases relative to the known sequence for the protein. Also, a comparison of sequence coverage to previous CE-MS work is shown for several proteins. The combined MALDI and ESI-MS procedure was performed for ten proteins from a malignant breast cancer cell line in the pH 7.5 fraction. Also, tandem mass spectrometric analysis confirmed the presence of a unique peptide in different isoforms of lamin.

5.2 Materials and Methods

The experimental overview of the work presented here is shown in Figure 5.1. Proteins from lysed cells are subjected to 2-D liquid-phase separations according to their pI and hydrophobicity. The isolated proteins are then enzymatically digested and analyzed by MALDI-TOF MS and monolith-based LC-MS for protein identification by PMF. The sequence coverage results from CE-MS analysis from previous work are also compared for several proteins. The intact M_r of the proteins is obtained using nonporous (NPS)-RP-HPLC separations interfaced on-line with ESI-TOF MS. The results from PMF analysis using the two different MS methods are combined to yield high sequence coverage and a comparison between the experimental and theoretical intact M_r is used to provide highly reliable protein identification and to identify the presence of isoforms. The sequencing information of the selected tryptic peptides is provided by the monolith-based LC directly interfaced with linear IT ESI-MS to confirm their presence.

5.2.1 Cell lines

The cells used in this work are CA1a.cl1 and CA1d.cl1, cloned and fully malignant lines derived from the MCF10A human breast cancer cell line, which are maintained and prepared by the Barbara Ann Karmanos Cancer Institute (Wayne State University, Detroit, MI) as previously described [23].

5.2.2 Preparative Liquid-Phase IEF

The first-dimensional separation based on pI was performed using the Mini-Rotofor (BioRad, Hercules, CA) to separate the cell extracts as previously described [9,15]. Briefly, cell extracts were mixed with IEF running buffer containing 8 M urea, 2 M thiourea, and 2% Biolyte ampholyte at pH 310 (BioRad). The Rotofor chamber was loaded with 18 mL of the buffer, and the separation was controlled at 12 W for 3.5 hr. The separated pI fractions were harvested into 20 tubes for pH measurements using an Orion pH meter (model 250A, Allometrics, Baton Rouge, LA) and Accumet combination electrodes (Fischer, Pittsburgh, PA). Each fraction was subjected to a colorimetric based Lowry assay (RC Protein Assay, BioRad) for relative quantitation and stored at -80°C until use.

5.2.3 NPS-RP-HPLC Separation

The second-dimensional separation was performed using NPS-silica columns. The NPS-RP-HPLC column (33×4.6 mm) packed with 1.5 μm C18 NPS ODSIIIE silica beads (Eprogen, Darien, IL) was used at a flow rate of 0.5 mL/min. Approximately 250 μg of protein obtained from each Rotofor fraction was loaded for separation with

the HPLC System Gold equipped with UV detector (Beckman Coulter, Fullerton, CA). The solvent system comprised of solvent A: helium-degassed deionized (DI) water (Millipore, Billerica, MA) with 0.1% TFA (TFA, 99.5%; Sigma, St. Louis, MO) and solvent B: ACN (ACN, 99.93% HPLC grade; Sigma) with 0.1% TFA. The column was maintained at 65°C (model 7971 column heater, Jones Chromatography, Resolution Systems, Holland, MI) to improve resolution and increase the separation speed by using a gradient elution profile as follows: 5-15% B in 1 min; 15-25% B in 2 min; 25-31% B in 3 min; 31-41% B in 10 min; 41-47% B in 3 min; 47-67% B in 4 min; 67-100% B in 1 min; 100% B for 2 min; and 100-5% B in 1 min. The proteins separated by NPS-RP HPLC were monitored at 214 nm and collected into 1.5 mL Eppendorf tubes using a Beckman SC-100 fraction collector controlled by semiautomated acquisition program. Protein collection was performed off-line according to the peaks detected from the HPLC separation where 40-50 fractions, each containing a volume in the range of 100-500 μL, were obtained.

5.2.4 NPS-RP-HPLC/ESI-TOF MS

Intact M_r analysis was performed by analyzing eluent from NPS-RP HPLC for on-line ESI-TOF MS (LCT, Waters-Micromass, Manchester, UK). The separation was performed under the same experimental conditions as in the previous section except that TFA was substituted with 0.3% formic acid (Sigma) in both mobile phases to improve ESI efficiency. A splitter system was used so that 40% of eluent from the HPLC was delivered to the LCT. The capillary voltage for electrospray was set at 3200 V, sample cone at 40 V, extraction cone at 3 V, and reflection lens at 750 V. Desolvation was enhanced by controlling the desolvation temperature at 300°C

and source temperature at 120°C. The nitrogen gas flow was maintained at approximately 650 L/h. One mass spectrum was acquired per second. The intact M_r value was obtained by deconvoluting the combined ESI spectra of the protein utilizing the MaxEnt1 feature of MassLynx software version 4.0 (Waters-Micromass).

5.2.5 Protein Digestion

The proteins collected off-line by NPS-RP-HPLC separation were completely dried down using a SpeedVac (Labconco, Kansas City, MO). 50 μL of 50 mM NH_4HCO_3 at pH 7.8, filtered through 0.22 μm and 0.5 μg of TPCK-modified sequencing-grade trypsin (Promega, Madison, WI), was added for digestion at 37°C for 18 hr and stopped thereafter by adding 1 μL of 10% v/v TFA. The digestion mixture was divided into 10 μL for MALDI-TOF MS analysis and the rest dried completely with a SpeedVac. Prior to monolith-based LC-MS and LC-MS/MS analysis, peptides were reconstituted in 5 μL of deionized water.

5.2.6 MALDI-TOF MS

Each digested protein was analyzed by MALDI-TOF MS. Prior to spotting, each of the digested samples were desalted and concentrated by using C18 ZipTip (Millipore, Bedford, MA) and eluted into 0.5% v/v TFA/60% v/v ACN. The MALDI-matrix solution was prepared by diluting saturated CHCA (Sigma) with 0.5% v/v TFA/60% v/v ACN at 1:4 ratio v/v. The internal standards included angiotensin I, adrenocorticotropic hormone (ACTH) fragments 1-17 and 18-39 (all from Sigma). Internal standard peptides were added so that a final concentration of 50 fmol each was

reached in each spot of the MALDI plate.

Peptide masses were measured by the TofSpec2E (Waters-Micromass) with delayed extraction in reflectron mode with positive polarity using a nitrogen laser (337 nm). The operating voltage was 20 kV and reflectron voltage was 24.5 kV. The pulse voltage used for delayed extraction was set at 2300 V at a 520 ns delay time. The sampling rate was 2 GHz. Peptide mass spectra were internally calibrated resulting in a mass accuracy of 50 ppm or less. The calibrated spectra were processed using MassLynx version 4.0 to obtain monoisotopic experimental masses for submission to MS-Fit available from `http://prospector.ucsf.edu` to search the Swiss-Prot and NCBI databases for protein identification under the species of Homo sapiens with no restrictions on M_r and pI range. A maximum of two missed cleavages was allowed, and cysteine was unmodified. The possible modifications including oxidation of methionine, N-terminal acetylation, and phosphorylation at S, T, and, Y were allowed.

5.2.7 Monolith-based HPLC Separation and On-line Interfacing with ESI-TOF MS

The Ultra-Plus II MD Capillary Pump module (Micro-Tech Scientific, Vista, CA) with a home-built column heater utilizing a variable autotransformer (Staco Energy Product, Dayton, OH) was used for all chromatographic experiments with the monolithic-capillary column. Monolithic capillary columns of dimension of 360 μm od $\times$ 200 μm id $\times$ 60 mm L were prepared by co-polymerizing styrene and divinylbenzene according to the protocol described elsewhere [19]. The capillary column was directly mounted to a microinjector (model C4-1004-.5, Valco Instruments, Houston, TX) with a 500 nL internal sample loop and a microtight union with 5 nL swept

volume (Upchurch Scientific, Oak Harbor, WA) was used to connect a capillary tubing to the LCT. The flow rate of the solvent delivery pump was set at 0.5 mL/min, which was split precolumn to produce a flow rate of 2.5 μL/min at 60°C through the monolithic-capillary column. A mobile-phase system of two solvents was used, where solvents A and B are composed of 0.05% formic acid (Sigma) in DI water and ACN, respectively. A linear gradient of 0-100% B in 18 min was applied, immediately followed by column equilibration.

Each digested protein was analyzed using monolith-based HPLC interfaced on-line with the LCT. The desolvation temperature was maintained at 120°C and source temperature at 80°C. The nitrogen gas flow was controlled at 90 L/h. One mass spectrum was acquired per second. The monoisotopic peptide M_r values were obtained by utilizing MaxEnt3 feature of MassLynx version 4.0 software by allowing maximum charge state of up to +3 in the mass range of 500-4000 Da. All processed mass spectra were subjected to PMF analysis by database searching in the manner described in Section 5.2.6.

5.2.8 Monolith-Based HPLC-MS/MS

In order to sequence the tryptic peptide sequence of interest, HPLC-MS/MS experiments were performed using linear IT MS (LTQ, Thermo Finnigan, San Jose, CA). An identical platform for HPLC separation, described in the previous section, was utilized. The capillary transfer tube was set at 175°C, and electrospray needle was held at +3.5 kV. A sheath gas flow of 15 arbitrary units was used. The ion activation was achieved by utilizing helium at a normalized collision energy of 35%. All MS/MS

data obtained were analyzed by using the TurboSequest feature of Bioworks 3.1 SR1 (Thermo Finnigan). By allowing the maximum missed cleavage of two, peptide ions are automatically assigned with the X_{corr} values to consider >3.5 for +3 ions, >2.5 for +2 ions, and >1.5 for +1 ions.

5.3 Results and Discussion

The monolithic column was used in this work to analyze proteins separated from breast cancer cell lines. All proteins from the human breast cancer cells in this experiment were analyzed with the same monolithic-capillary column. The column exhibits excellent ruggedness, where the stability of this type of column has been described elsewhere [24]. The salts and other species that might otherwise clog the column in CE-MS were cleaned up by running a 100% aqueous solvent prior to RP separation to wash off the impurities. In addition, only a few femtomoles of injection was possible with CE-MS in previous work [15], making it difficult to analyze proteins of low abundance, whereas monolith-based LC offers much higher loadability to improve detection.

5.3.1 High Protein Sequence Coverage with Monolithic LC-MS

A CA1a cell lysate was separated by 2-D liquid separations where NPS-RP-HPLC separation of proteins prefractionated at the pH of 7.55 is shown in Figure 5.2. A total of ten proteins contained in this fraction were analyzed for identification through PMF analysis. The protein digests were analyzed by capillary monolithic LC-MS where a typical high-resolution separation profile, completed in less than 10 min,

is shown in Figure 5.3 for one of the proteins (peak no. 6 in Figure 5.2). Typical full peak widths at half height are in the range of 3-5 s. The speed of separation is important since in analysis of complex cell lysates, there may be large numbers of proteins that need to be analyzed. Considering that recovery of proteins separated by NPS silica column is approximately 80% [25], it is estimated by performing peak quantitation with manual baseline by Origin software (version 6.0, Microcal Software, Northampton, MA, USA) that approximately 40 ng of protein digest was injected into the monolithic-capillary column. The mass analysis and database search identify this protein as fructose bisphosphate aldolase A, a 40 kDa protein so that approximately 1 pmol was consumed for analysis.

Table 5.1 summarizes the comparison between theoretical versus experimental intact M_r, pI, and the sequence coverage by monolith-based LC/MS and MALDI-TOF MS for each protein denoted in Figure 5.2. Also shown is the overall sequence coverage obtained by combining peptide masses from each method into database searching. In many cases, high sequence coverage was readily obtained solely by the monolithic-LC-MS method. Figure 5.4 A illustrates the protein coverage map by each MS method for annexin II, which is found to behighly expressed in the malignant cell of human breast cancer and suggested as a potential biomarker based on the previous differential expression study [9]. The sequence coverage of 90% for this protein of ca. 40 kDa was analyzed by monolithic LC-MS alone. As shown in the coverage map, monolithic LC-MS missed only three peptides, including 1-10, 79-88, and 207-212, excluding three tryptic peptides composed of one or two amino acids that cannot be detected in the mass range set for 500-4000 Da for both of the mass spectrometric

methods. This implies that the protein coverage by monolithic LC-MS was almost complete. A similar observation can be made with superoxide dismutase, for which the coverage map is shown in Figure 5.4 B. For this protein, a single tryptic peptide sequence composed of 38 amino acids (157-194; 4236.16 Da) without any cleavage site and a single amino acid residue (222) were not detected, of which detection is limited by the mass range examined in this experiment.

Although significantly higher protein sequence coverage was consistently obtained with monolith-based LC/MS as compared to MALDI-TOF MS, it is observed that the addition of peptide mass data from MALDI-TOF MS contributed to an increase of >10% of the overall protein sequence coverage for many of the proteins analyzed. It shows that this unique approach of combining different MS methods typically covers from 80% to over 90% of the overall sequence of the proteins ranging in size from a few thousands to over 70 kDa. Table 5.2 shows that proteins contained in other pH fractions of CA1a and a different cell line, CA1d, were also successfully analyzed.

Occasionally, the tryptic peptide sequences that were detected by neither monolith-based LC-MS or MALDI-TOF MS were detected and successfully sequenced by linear IT MS. Two of the tryptic peptides of heterogeneous nuclear ribonucleoproteins A2/B1, LFVGGIK, shown in Fig. 5.5 with the MS/MS spectrum, and EESGKPGAHVTVK (110-112) were not detected by either method for PMF analysis, yielding the overall coverage of only 74%. However, the tandem mass spectrometric analysis identified these peptides, therefore increasing the overall sequence coverage to 80%.

5.3.2 Enhanced Peptide Detection by Monolithic LC-MS

In previous work, lamin, which helps maintain and establish the shape and strength of the interphase nucleus [26,27], was analyzed with CE-MS with sequence coverage ranging from ca. 68-76% for a few isoforms (Table 5.1)[15]. The same protein was analyzed by monolithic LC-MS to reveal sequence coverage of higher than 80-90%, which increased up to 96% with the addition of the results from MALDI-TOF MS. Figure 5.6 A shows the monolithic LC-MS spectrum scanned over the actual separation time of a digest of lamin identifying five peptide sequences, mostly within mass accuracy of less than 100 ppm, that were not detected by the CE-MS method. Given that both methods utilized the same ionization method, it is strongly believed that the separation characteristics of monolithic LC, including high recovery and loadability, may have contributed to revealing peptides that went undetected by CE. This is an important observation because peptides bearing certain PTMs are difficult to detect due to their presence in relatively low concentrations and the monolithic LC may be a suitable means to overcome this problem. Two additional tryptic peptide sequences in lamin, key to identifying sequence variations involving isoforms, were detected by monolithic LC-MS analysis and are discussed in the following section. Figure 5.6 B shows the direct comparison of sequence coverage for lamin isoforms obtained by monolithic LC-MS and CE-MS, where it is clearly visualized that enhanced sequence coverage was observed by monolithic LC-MS.

5.3.3 Analysis of Isoforms and PTMs Using Monolithic LC-MS and NPS-RP-LC-MS

In addition to very high protein sequence coverage by PMF analysis, an excellent match between experimental M_r obtained by NPS-RP-HPLC/ESI-TOF MS and the

theoretical M_r of many of the proteins analyzed in this experiment serves to further suggest that the protein identification procedures are highly reliable, as several proteins in Tables 5.1 and 5.2 exhibited a very close agreement. Ubiquitin, for example, contains no suggested modifications in its sequence, and therefore experimental and theoretical molecular weights are identical.

In our previous study [15], the usefulness of intact M_r in determining sequence variations of proteins was demonstrated where unique peptides were detected by MALDI-TOF MS. The use of M_r measurements allowed the identification of the presence of isoforms of lamin. A single gene of lamin A/C encodes four isoforms, the most widely studied being lamins A and C, as well as ADelta10, resulting from alternative splicing [26,28]. The lamin C is identical to that of lamin A for the sequence (1-566), while a slight difference is observed in (567-572) where GSHCSS in lamin A is replaced by VSGSRR in lamin C. The rest of the sequence in lamin A (573-664) is absent in lamin C. In lamin ADelta10, a sequence (537-566) is missing from lamin A, the rest being identical [27]. In the present study, we successfully identified unique peptides with no missed cleavages contained in two isoforms, lamins A and C, with both monolith LC-MS and MALDI-TOF MS, as summarized in Table 5.3. Figures 5.7 A and B show the detection of the peptide sequences unique to lamins A and C, respectively, by monolithic LC-MS. Figure 5.8 shows that one of these unique peptides, TALINSTGEEVAMR (528-541), was successfully sequenced, further confirming its detection by utilizing monolith-based LC separation. The unique peptide of lamin ADelta10 was not detected by the present study. It is interesting to note that none of these unique peptides were observed by CE-MS analysis in our previous

study, while their presence was observed using monolithic LC-MS, further suggesting its usefulness in PMF analysis.

Another interesting observation of isoforms was made with cytoplasmic actin. There are three main groups of actin isoforms in vertebrates, including α, β, and γ forms [29,30]. β-Actin and γ-actin are known to coexist in most cell types as components of the cytoskeleton and as mediators of internal cell motility. This protein has been studied earlier based on intact M_r from NPS-RP-HPLC [31], although unique peptides were not found by MALDI-TOF MS analysis due to their very high degree of homology where 98.9% (371/375) of the sequence is identical. In this work, both β-actin and γ-actin were identified with overall coverage of 83 and 88%, respectively, with closely matching intact M_r values, when peptide maps from different MS methods were combined (Table 5.2). The unique peptides of these two isoforms are found in four of the first ten amino acids in their sequences, MEEEIAALVI for γ-actin and MDDDIAALVV for β-actin, while the remainder of the 371 amino acids in the sequence of these proteins are identical, making it a difficult task to distinguish them.

In their mature forms, the first amino acid residue of the sequence, methionine, of both forms of actins is missing and the second amino acid in each of the sequence, glutamic acid for γ-actin and aspartic acid for β-actin, are acetylated at the N-terminus [32-34]. Many studies report that the histidine residue at position 73 is methylated in actins [35,36]. Considering these modifications and alterations in sequence, theoretically, γ-actin should have MW of 41,718 Da, which closely matches our experimentally determined intact MW of 41,714 Da, or less than 100 ppm of mass

accuracy. The same consideration can be made with β-actin which should have MW of 41,662 Da, in comparison to the MW of 41,670 Da determined experimentally. In Figure 5.9, the observation of the presence of the unique peptide sequence of γ-actin with M_r including acetylation at a glutamic acid residue in its mature form by both monolithic LC-MS and MALDI-TOF MS is illustrated. Neither MS method revealed the unique peptide sequence contained in β-actin.

A closer examination of the mass spectrum obtained for actin allowed for the detection of the presence of another PTM. Shown in Figure 5.10 is a doubly and triply charged peptide sequence containing methylated histidine, YPIEHGIVTNWDDMEK (69-84, 1960.911 Da), by monolithic LC-MS. This modified sequence was also found by MALDI-TOF MS analysis. The modification at this particular residue of this peptide sequence, common to both β- and γ-actins, was suggested by a recent study where it appears to play a role in polymerization of actin and ATP hydrolysis [37]. Although detailed structural study by MS/MS is necessary for characterization, the results presented here strongly indicate that monolith-based LC-MS followed by PMF analysis is capable of identifying the presence of PTMs.

5.4 Conclusions

A method for obtaining consistently high sequence coverage of proteins separated from lysates of human breast cancer cell lines has been successfully demonstrated by utilizing a polymer-based monolithic-capillary column for LC-MS to obtain a rapid and high-resolution separation. Very high sequence coverage of proteins readily obtained by combining PMF results from monolithic LC/MS and MALDI-TOF MS

enhances the reliability of the protein identification procedures. Although CE-MS could also be used to analyze these samples, an important advantage of the monolith HPLC-ESI-TOF MS is the ruggedness of the technique. In CE-MS using a sheathless interface the electrical connection has been found to be the limiting factor where the connection needs to be redone after a couple of runs. Although CE can provide much improved resolution [38] compared to the monolithic columns in the 10 min separation interval, with the use of MS analysis sufficient resolution is obtained for distinguishing each peak. In addition, the monolithic columns provide much improved loadability compared to CE and much shorter separation times compared to packed capillary chromatography.

The monolithic column separations coupled to MS was also used to elucidate the presence of sequence variations, such as isoforms, and PTMs of proteins, aided by intact M_r information and MS/MS using linear IT MS. It is important to emphasize the uniqueness of this study due to the nature of samples analyzed, which originate from highly complex biological mixtures, suggesting the versatility of this approach for many other applications. In addition, this method requires no sample preparation or purification upon completion of enzymatic digestion of proteins prior to MS analysis.

Several proteins analyzed in this work, including annexin II, fructose bisphosphate aldolase A, phosphoglycerate kinase 1, and peptidyl-prolyl cis-trans isomerase A, have been reported to be highly expressed in tumor cells and described as potential biomarkers based on differential expression studies [9,39]. Although further work will be necessary to fully characterize these proteins, it is clearly demonstrated in this

study that the method of monolith-based LC on-line hyphenated with MS has great potential to become a high-throughput methodology. It may be used to characterize large numbers of potential biomarkers in various types of tumor and other diseased cells and to study the presence of a wide variety of modifications and structural changes in protein expression.

Table 5.1: Comparison between theoretical and experimental intact M_r and pI and sequence coverage between different MS methods for proteins in pH fraction of 7.55 of CA1a cell line (see Figure 5.2 for peak number.)

Peak number	Protein name	Accession number	Theoretical		Experimental		Protein sequence coverage (number of amino acids)		
			M_r	pI	M_r	pI	MALDI-TOF	Monolith LC/MS	Overall
1	Ubiquitin	P68198 M	8 565	6.6	8 565	7.55	61% (46/76)	95% (72/76) 100% (76/76)[b]	95% (72/76)
2	Peptidyl-prolyl *cis-trans* isomerase A	P62937	18 013	7.7	18 052	7.55	46% (76/165)	78% (128/165)	90% (149/165)
3	Nucleophosmin	P06748	32 575	4.6	33 104	7.55	59% (173/294)	69% (203/294)	84% (248/294)
4	Lamin A	P02545[a]	74 140	6.6	74 601	7.55	44% (290/664)	82% (545/664) 69% (457/664)[b]	86% (573/664)
	Lamin C		65 135	6.4	66 056	7.55	55% (316/572)	91% (520/572) 77% (438/572)[b]	96% (548/572)
	Lamin ADelta10		70 660	8.6	70 114	7.55	44% (272/634)	83% (527/634) 71% (451/634)[b]	88% (555/634)
5	Superoxide dismutase	P04179	24 722	8.3	24 606	7.55	60% (134/222)	78% (174/222)	82% (183/222)
6	Fructose bisphosphate aldolase A	P04075	39 420	8.3	39 282	7.55	40% (147/364)	71% (257/364)	80% (293/364)
7	α-Enolase	P06733	47 169	7.0	47 073	7.55	48% (210/434)	91% (397/434)	91% (397/434)
8	Annexin II	P07355	38 604	7.6	38 614	7.55	47% (158/339)	90% (304/339)	95% (321/339)
9	ATP synthase α chain	P25705	59 751	9.2	59 750	7.55	47% (259/553)	76% (423/553)	80% (445/553)
10	Phosphoglycerate kinase 1	P00558	44 615	8.3	44 528	7.55	55% (231/417)	88% (369/417)	88% (369/417)

a) Isoforms found. Database searching performed with both Swiss-Prot and NCBI databases.
b) These percentage coverages are the previous results from CE-MS, adapted from [15] with permission.

Table 5.2: Comparison between theoretical and experimental intact M_r and pI and sequence coverage between different MS methods for proteins in other pH fractions of CA1a and CA1d cell lines

Cell line	Protein name	Accession number	Theoretical		Experimental		Protein sequence coverage (number of amino acids)		
			M_r	pI	M_r	pI	MALDI-TOF	Monolith LC-MS	Overall
CA1a	Splicing factor, arginine/serine-rich 1	Q07955	27 745	10.4	28 163	7.00	58% (145/248)	73% (182/248)	88% (218/248)
CA1a	Actin[a]	P63260 (γ-actin)	41 793	5.3	41 714	6.21	53% (198/375)	81% (305/375)	88% (331/375)
		P60710 (β-actin)	41 737	5.3	41 670	6.21	48% (181/375)	77% (287/375)	83% (313/375)
CA1d	Keratin, type I cytoskeletal 17 (CK17)	Q04695	48 106	5.0	48 077	5.71	56% (241/432)	72% (309/432)	84% (365/432)
CA1d	Heterogeneous nuclear ribonucleoproteins A2/B1[b]	P22626	37 430	9.0	37 411	6.91	43% (153/353)	74% (261/353)	80% (281/353)

a) Isoforms found.
b) Overall protein sequence coverage includes two tryptic peptides (110–112 and 114–120) analyzed by monolithic LC/MS/MS.

Table 5.3: Unique tryptic peptides detected by different MS methods to distinguish isoforms

Protein name	Isoform(s)	Unique tryptic peptide sequence (amino acids)	$[M + H]^+$	MS method to detect		
				MALDI-TOF MS	Monolith LC-MS	CE-MS
Lamin A/C	Lamins A and C	TALINSTGEEVAMR (528–541)	1491.747	Yes	Yes	No
	Lamin C	SVTVVEDDEDEDGDDLLHHHHVSGSR (546–571)	2899.272	Yes	Yes	No
Actin[a)]	β-Actin	DDDIAALVVDNGSGMCK (2–18)	1764.778	No	No	NA
	γ-Actin	EEEIAALVIDNGSGMCK (2–18)	1820.841	Yes	Yes	NA

a) $[M + H]^+$ values include *N*-terminal acetylation for each of the peptide.

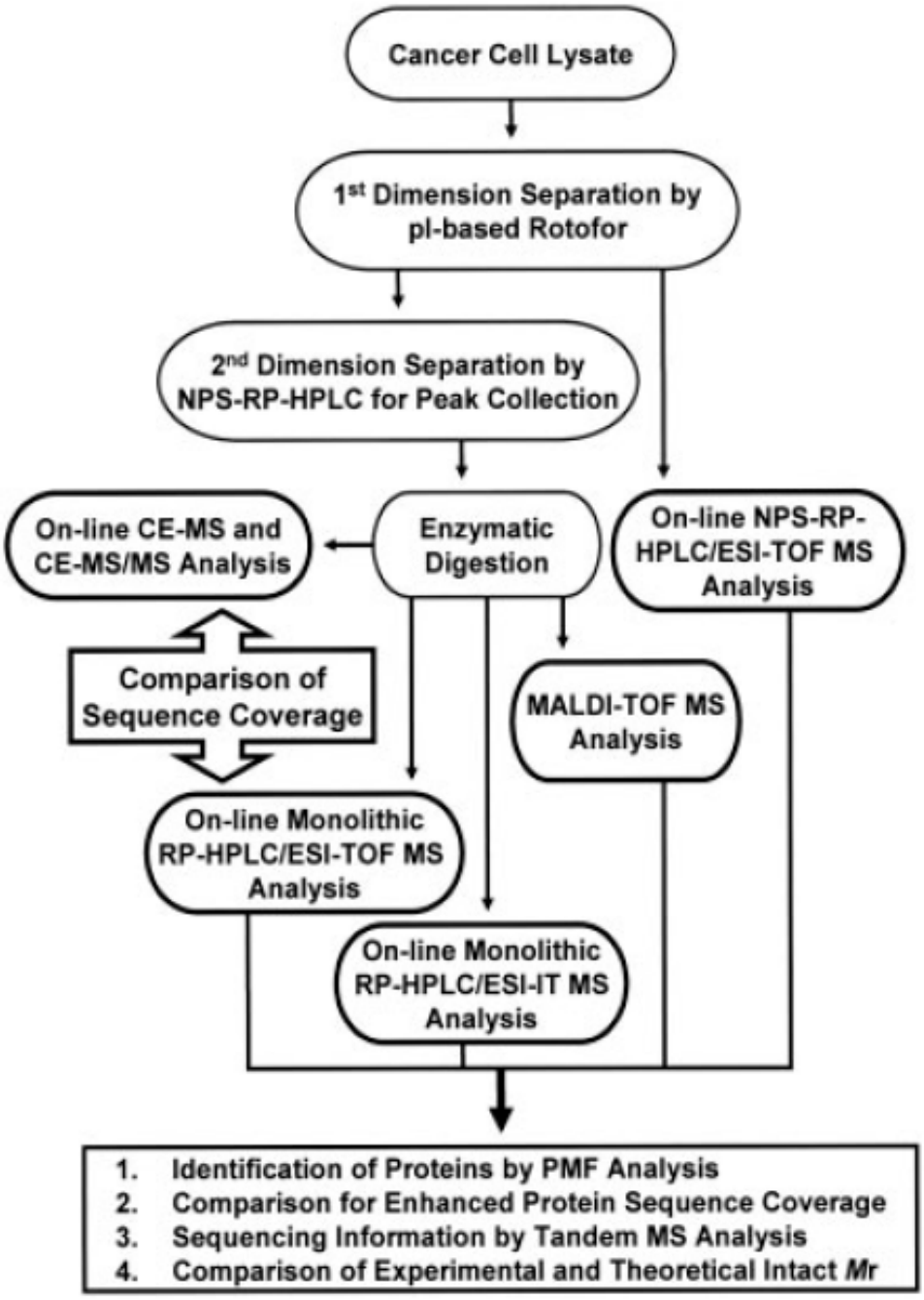

Figure 5.1: Experimental scheme of the 2-D liquid-phase separation techniques followed by PMF analysis and sequencing from different MS methods for identification of proteins in human breast cancer cell lines with high overall sequence coverage.

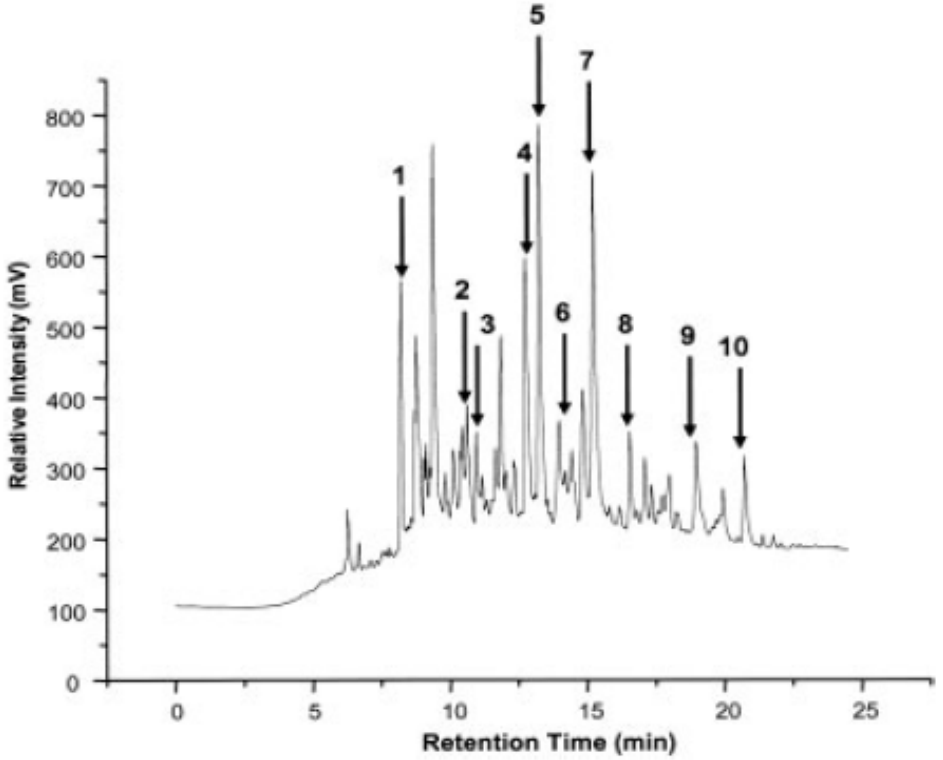

Figure 5.2: NPS-RP-HPLC chromatogram of proteins in pH fraction of 7.55 from CA1a cell line with peaks annotated for protein identification results shown in Table 5.1.

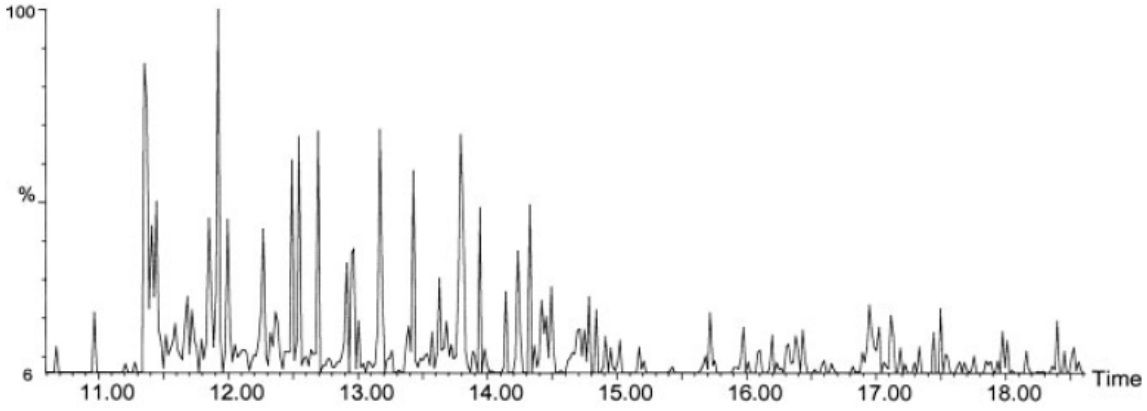

Figure 5.3: Monolith-based RP-HPLC-MS chromatogram of tryptic digest of fructose bisphosphate aldolase A.

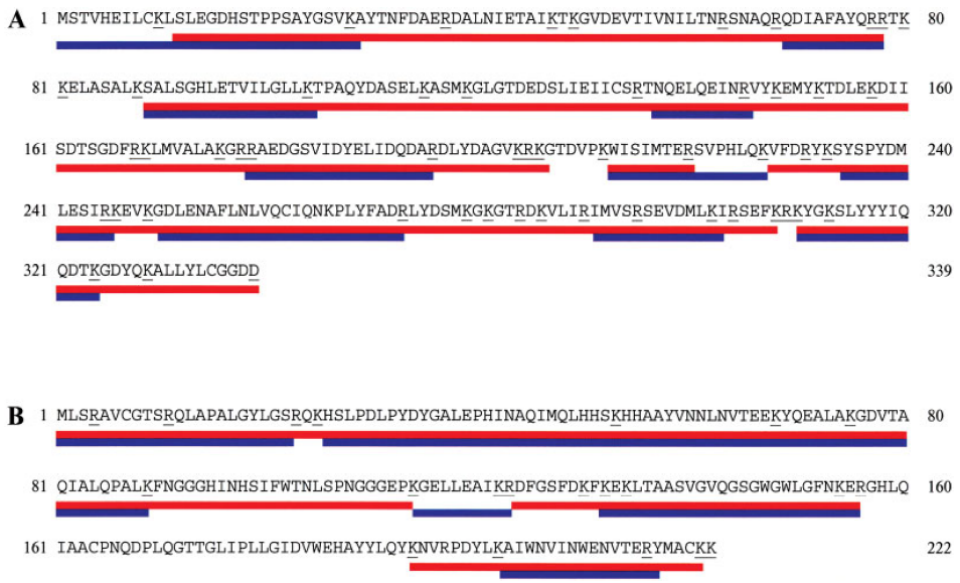

Figure 5.4: Illustration of protein sequence coverage maps (red bar by monolith-based HPLC/ESI-TOF MS; blue bar by MALDI-TOF MS) of (A) annexin II and (B) superoxide dismutase.

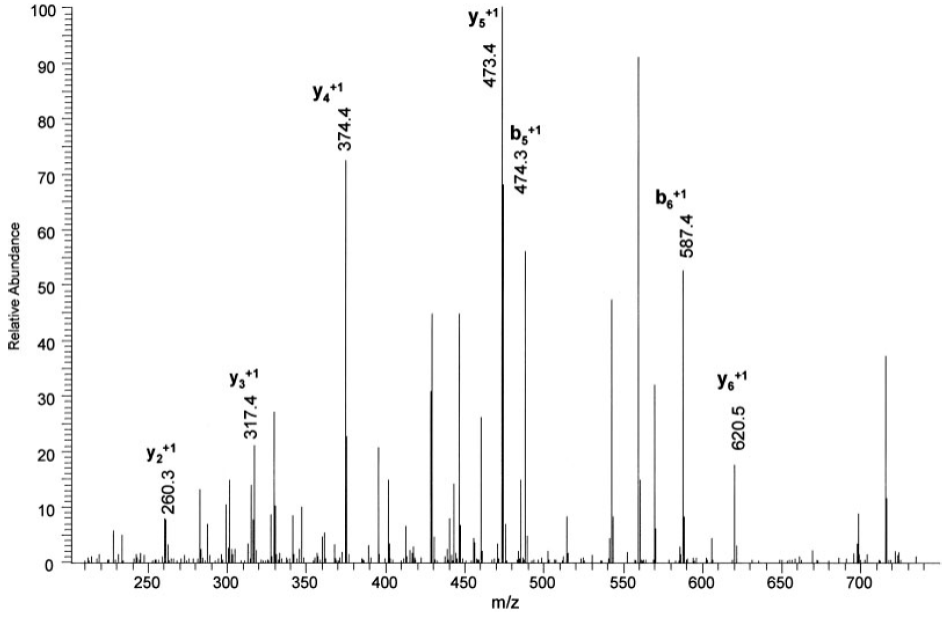

Figure 5.5: Monolith-based LC-MS/MS spectrum of the tryptic peptide, LFVGGIK (114-120), of heterogeneous nuclear ribonucleoprotein A2/B1.

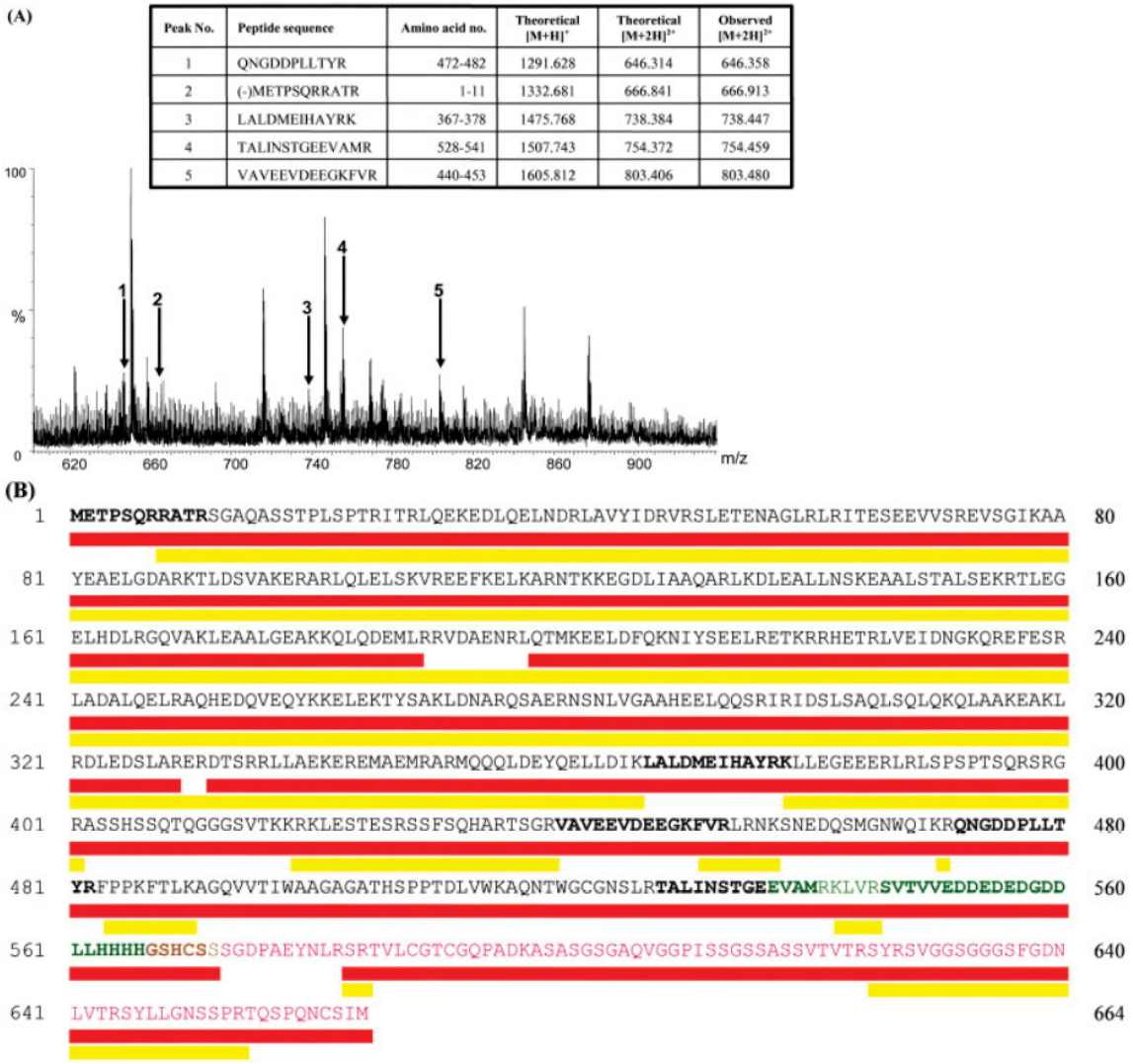

Figure 5.6: (A) Illustration of several tryptic peptides (sequence information in table inset) from lamin detected by monolithbased HPLC/ESI-TOF MS that were not detected by CE-ESI-TOF MS in previous study. (B) Comparison of protein sequence coverage for lamin A/C (red bar monolith-based LC-MS; yellow bar CE-MS, adapted from [15] with permission). All tryptic peptides only detected by monolithic LC-MS are in bold. Note: Sequence in green (537-566) is missing in lamin Adelta10. Sequence in pink (573-664) is missing in lamin C. Sequence in brown (567-572) is replaced with VSGSRR in lamin C.

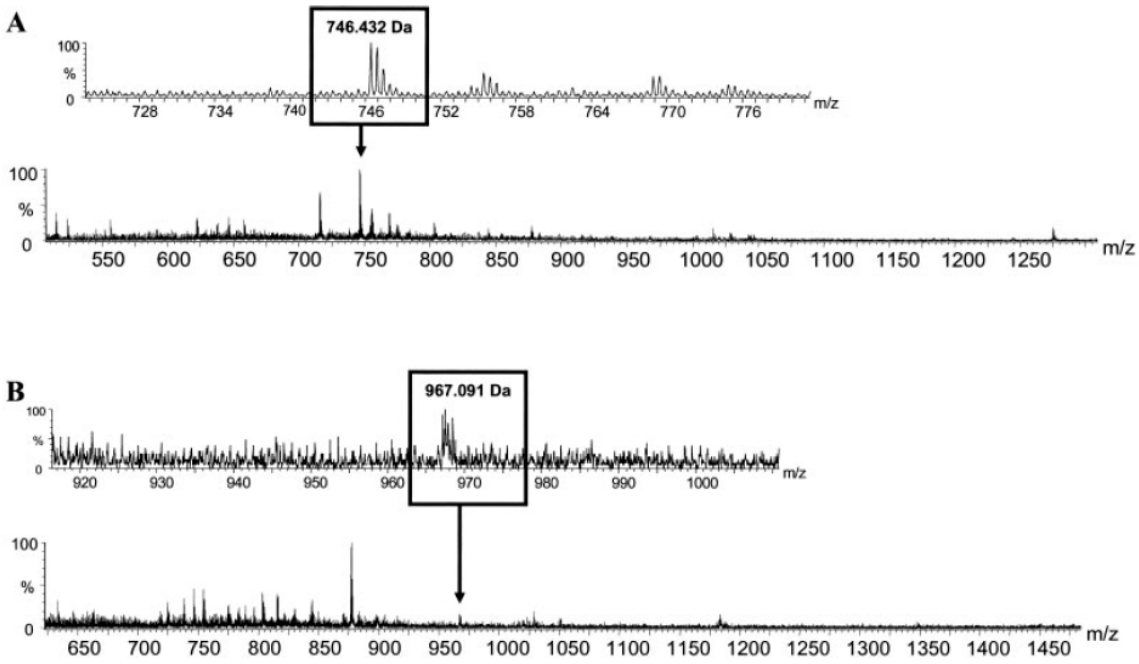

Figure 5.7: Identification of unique peptides (multiply charged; see Table 5.3 for peptide sequence information) in the isoforms of lamin, (A) lamins A and C and (B) lamin C, detected by monolith-based HPLC/ESI-TOF MS with zoomed-in view.

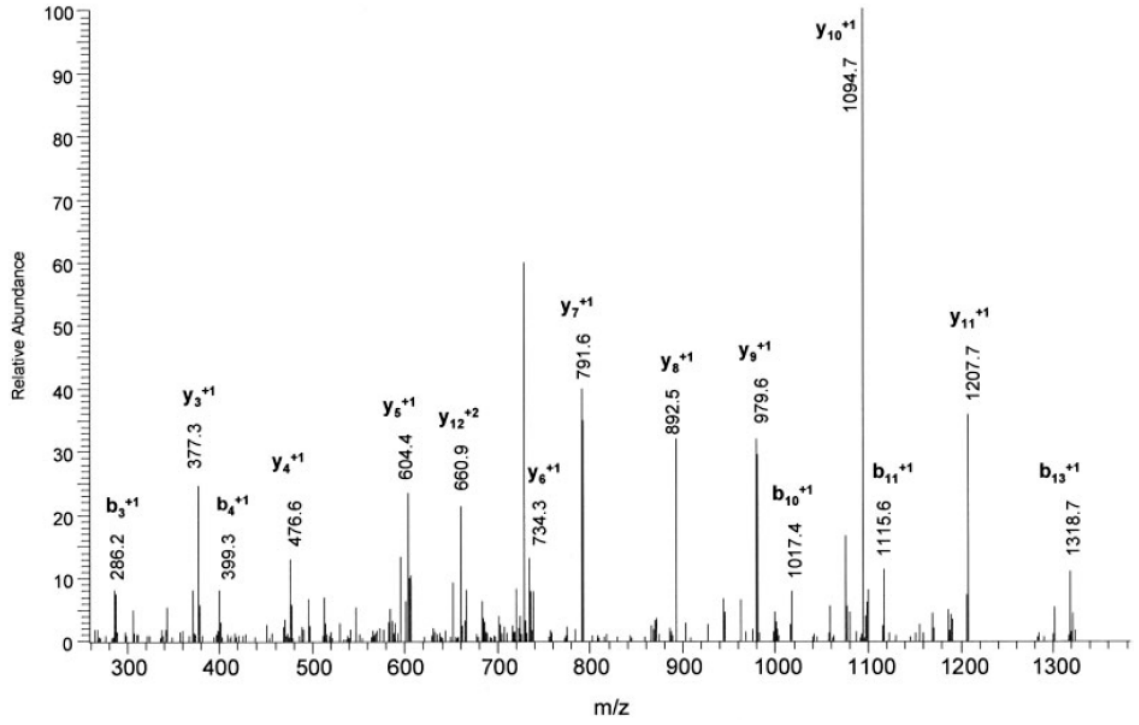

Figure 5.8: Monolith-based LC-MS/MS spectrum of the tryptic peptide, TALINSTGEEVAMR (528-541), of lamins A and C.

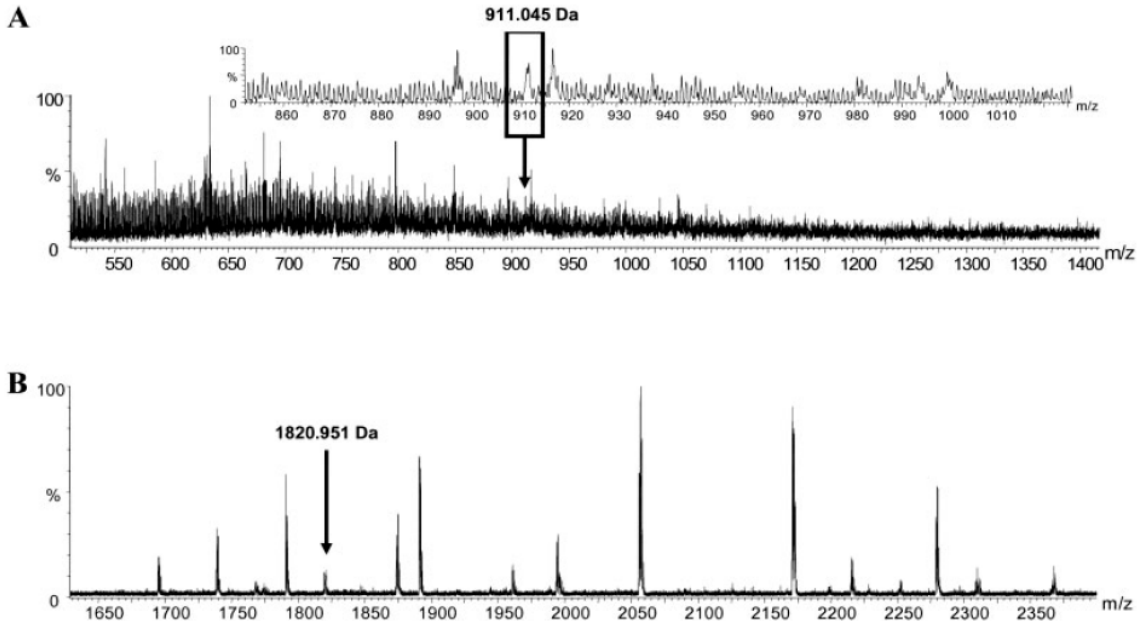

Figure 5.9: Identification of unique peptide (see Table 3 for peptide sequence information) in g-actin by (A) monolith-based HPLC/ESI-TOF MS with zoomed-in view and (B) MALDI-TOF MS.

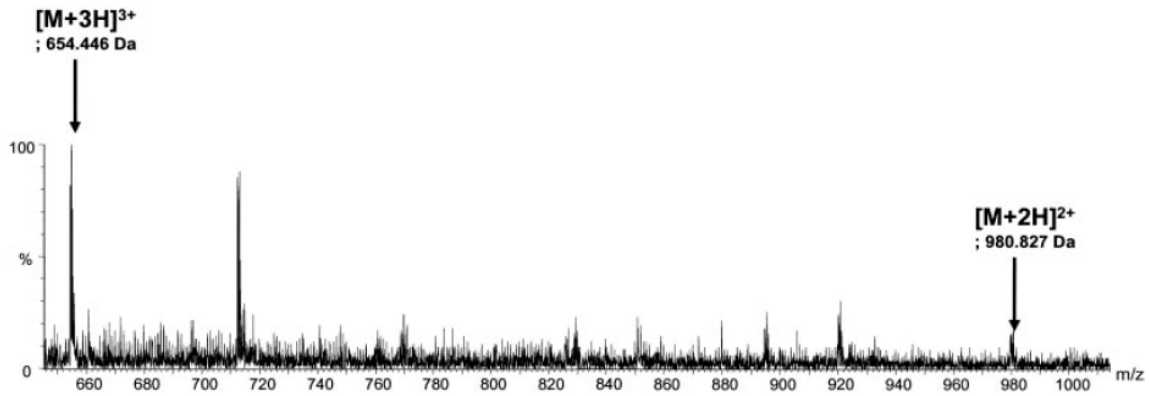

Figure 5.10: Identification of multiply charged tryptic peptide containing methylated histidine residue in actin by monolithbased HPLC/ESI-TOF MS.

References

[1] Vlahou A.; Fountoulakis M. *J. Chromatogr. B*, 814:11-19, 2005.

[2] O'Donovan C.; Apweiler R.; Bairoch A. *Trends Biotechnol.*, 19:178-181, 2001.

[3] Mann M.; Jensen O. N. *Nat. Biotechnol.*, 21:255-261, 2003.

[4] Han K. K.; Martinage A. *Int. J. Biochem.*, 24:19-28, 1992.

[5] Wu W.; Hu W.; Kavanagh J. J. *Int. J. Gynecol. Cancer*, 12:409-423, 2002.

[6] O'Farrell P. H. *J. Biol. Chem.*, 250:4007-4021, 1975.

[7] Bernova-Giorginni S. *Trends Anal. Chem.*, 22:273-281, 2003.

[8] Goodlett D. R.; Yi E. C. *Funct. Integr. Genomics*, 2:138-153, 2002.

[9] Hamler R. L.; Zhu K.; Miller F. R.; Lubman D. M. et al. *Proteomics*, 4:562-577, 2004.

[10] Kreunin P.; Urquidi V.; Lubman D. M.; Goodison S. *Proteomics*, 4:2754-2765, 2004.

[11] Wang H.; Kachman M. T.; Schwartz D. R.; Cho K. R.; Lubman D. M. *Proteomics*, 4:2476-2495, 2004.

[12] Padliya N. D.; Wood T. D. *Proteomics*, 4:466-473, 2004.

[13] Krause E.; Wenschuh H.; Jungblut P. R. *Anal. Chem.*, 71:4160-4165, 1999.

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

[15] Zhu K.; Kim J.; Yoo C.; Miller F. R.; Lubman D. M. *Anal. Chem.*, 75:62096217, 2003.

[16] Moore R. E.; Licklider L.; Schumann D.; Lee T. D. *Anal. Chem.*, 70:4879-4884, 1998.

[17] Walcher W.; Oberacher H.; Troiani S.; Huber C. G. et al. *J. Chromatogr. B*, 782:111-125, 2002.

[18] Holzl G.; Oberacher H.; Pitsch S.; Stutz A.; Huber C. G. *Anal. Chem.*, 77:673-680, 2005.

[19] Premstaller A.; Oberacher H.; Huber C. G. *Anal. Chem.*, 72:4386-4393, 2000.

[20] Premstaller A.; Oberacher H.; Walcher W.; Timperio A. M. et al. *Anal. Chem.*, 73:2390-2396, 2001.

[21] Lee D.; Svec F.; Frechet J. M. J. *J. Chromatogr. A*, 1051:53-60, 2004.

[22] Kimura H.; Tanigawa T.; Morisaka H.; Ikegami T. et al. *J. Sep. Sci.*, 27:897-904, 2004.

[23] Santner S. J.; Dawson P. J.; Tait L.; Soule H. D. et al. *Breast Cancer Res. Treat.*, 65:101-110, 2001.

[24] Toll H.; Oberacher H.; Swart R.; Huber C. G. *J. Chromatogr. A*, 1079:274-286, 2005.

[25] Wall D. B.; Lubman D. M.; Flynn S. J. *Anal. Chem.*, 71:3894-3900, 1999.

[26] Broers J. L.; Hutchinson C. J.; Ramaekers F. C. *J. Pathol.*, 204:478-488, 2004.

[27] Moir R. D.; Spann T. P. *Cell. Mol. Life Sci.*, 58:1748-1757, 2001.

[28] Wilson K. L.; Zastrow M. S.; Lee K. K. *Cell*, 104:647-650, 2001.

[29] Vandekerckhove J.; Weber K. *J. Mol. Biol.*, 126:783-802, 1978.

[30] Garrels J. I.; Gibson W. *Cell*, 9:793-805, 1976.

[31] Wall D. B.; Kachman M. T.; Gong S. S.; Lubman D. M. et al. *Rapid Commun. Mass Spectrom.*, 15:1649-1661, 2001.

[32] Rubenstein P. A.; Martin D. J. *J. Biol. Chem.*, 258:3961-3966, 1983.

[33] Vandekerckhove J.; Weber K. *Eur. J. Biochem.*, 90:451-462, 1978.

[34] Abe A.; Saeki K.; Yasunaga T.; Wakabayashi T. *Biochem. Biophys. Res. Commun.*, 268:14-19, 2000.

[35] Asatoor, A. M.; Armstrong M. D. *Biochem. Biophys. Res. Commun.*, 26:168-174, 1967.

[36] Raghavan M.; Smith C. K.; Schutt C. E. *Anal. Biochem.*, 178:194-197, 1989.

[37] Nyman T.; Schuler H.; Korenbaum E.; Lindberg U. et al., J. *Mol. Biol.*, 317:577-589, 2002.

[38] Pelzing M.; Neusss C. *Electrophoresis*, 26:2717-2728, 2005.

[39] Bini L.; Magi B.; Marzocci B.; Tosi P. et al. *Electrophoresis*, 18:2832-2841, 1997.

CHAPTER VI

Automated integration of monolith-based protein separation with on-plate digestion for mass spectrometric analysis of esophageal adenocarcinoma human epithelial samples

6.1 Introduction

The immense complexity of the human proteome [1] presents a great challenge that requires the analysis of large numbers of proteins [2]. Therefore, it is important that proteomic methodologies should be simple, automated, and versatile for rapid and reproducible analysis [3]. 2D-GE [4] is still the most widely utilized method for large scale proteomics applications, where in-gel digestion of each protein spot is analyzed by subsequent ESI- or MALDI-based MS analysis. Despite development of robotic systems for gel-spot picking and excision [5-7], 2D-GE coupled with MS remains difficult and time-consuming due to extensive sample cleanup and digestion of a few hundred to several thousands of spots that may be present in a sample representing the human proteome.

Recently, the integration of HPLC with offline fraction collection for MALDI-MS analysis has received attention [8,9] and has been applied to several systems [10-13]. LC/MALDI has also proven useful for unique applications and has been demon-

strated to be useful in quantitative studies [14] and in the characterization of post-translational modifications [15] of proteins. In experiments utilizing LC/MALDI and applied to complex biological mixtures, proteins have been exclusively analyzed by enzymatic digestion of whole cell lysates [16-20] using sequential coupling of ion exchange and RP-HPLC separations. Although this shotgun proteomics approach is useful for a comprehensive analysis on a global scale [21], the results can be misleading, where the complexity of samples may result in false positive identifications when only a small number of peptides are matched [22]. It was previously shown that proteome analysis performed at the level of intact proteins [23] through 2-D liquid phase separations can help avoid such problems, where each of the protein fractions was further examined in detail and compared with intact molecular weight (MW) analysis [24-27]. This concept was used to develop a novel method that integrates monolithic HPLC separation of intact proteins with on-MALDI plate tryptic digestion [28] for rapid identification of proteins. It has been recently applied to a human breast cancer cell line [29], where lengthy experimental procedures required by in-solution digestion were greatly minimized. Also, unique characteristics of the monolith [30-32] as a separation medium, including high recovery and rapid separation speed [33,34], were found to be ideally suited for the purpose of LC/MALDI analysis.

Automation of the sample-handling process is an important issue to be considered in proteomics in order to develop a truly high-throughput methodology facilitating the analysis of large numbers of proteins [35]. An automated integration of HPLC with offline fraction collection for MALDI-MS analysis has been attempted previously in

several studies for the analysis of protein digests [10,36]. In the present studies, we focus on the advancement of this method by automation of all liquid-handling procedures. The proteins from human esophageal adenocarcinoma [37,38] a cancer type that is demonstrating an alarming increase in incidence, were analyzed to demonstrate the versatility and applicability of this method using whole tissue proteins. A selected set of pH-fractionated Barrett's cancer tissue samples was analyzed for protein identification. HPLC/ESI-TOF MS was performed to compare intact protein MW for further confirmation.

6.2 Materials and Methods

Proteins from esophageal tissue samples were separated by chromatofocusing (CF) for pH fractionation. One of the fractions was selected for further separation by monolithic capillary RP-HPLC connected to the modified nano-plotter (GeSim), equipped with xyz-robotic unit for automatic offline peak collection. The fractions at 30 s intervals were collected directly on the MALDI plate precoated with trypsin for on-plate digestion and subsequent MS analysis. A simplified LC/MALDI configuration is described in Figure 6.1, where all liquid-handling procedures were automated in this experiment. The same pH fraction was also analyzed by online nonporous (NPS) RP-HPLC/ESI-TOF MS to obtain accurate intact protein MW values for comparison with theoretical MW values.

6.2.1 Sample Preparation

Patient consent was received according to guidelines set forth by the institutional review board (IRB) of the University of Michigan. Tissues were collected and analyzed from patients undergoing esophagectomy at the University of Michigan Health System between 1991 and 2001. Patients received no preoperative radiation or chemotherapy. Each esophageal adenocarcinoma tumor specimen was immediately frozen in liquid nitrogen and stored at -80°C. Cryostat sectioning of all tumors were performed and only portions of tumors containing >80% tumor cellularity were utilized for subsequent protein isolation.

6.2.2 Cell Lysis and Buffer Exchange

200 μg of tissue sample was lysed with 2 mL of lysis buffer which consisted of 7.5 M urea, 2.5 M thiourea, 4% OG, 10 mM TCEP, 10% glycerol, 50 mM Tris (all from Sigma, St. Louis, MO), and 40 μL protease inhibitor solution (one tablet in 1 mL PBS buffer, Roche, Indianapolis, IN). Samples were then homogenized, vortexed frequently for 1 hr at room temperature, and centrifuged at 30,000 rpm for 70 min at 4°C. The collected supernatant was subjected to buffer exchange against CF start buffer using a PD-10 G-25 column (Amersham Biosciences, Piscataway, NJ). Bradford protein assay kit (BioRad, Hercules, CA) was used to quantify the amount of proteins in mixtures. The buffer-exchanged protein mixtures were stored at -80°C until further use.

6.2.3 Chromatofocusing

CF separation was performed on an HPCF-1D column (250×2.1 mm, Beckman Coulter, Fullerton, CA) using a Beckman System Gold HPLC. Prior to sample loading, the column was equilibrated with a start buffer containing 25 mM BisTris propane (Sigma), 6 M urea, and 1% OG, where its pH was adjusted to pH 7.4 with saturated IDA (Sigma). 4.5 mg protein was loaded to the elution buffer at pH 4.0 containing 10% Polybuffer 74 (Amersham Pharmacia), 6 M urea, and 1% OG at a flow rate of 0.2 mL/min. A linear pH gradient was generated so that proteins eluted off according to their pI for detection at 280 nm. Accurate pH was measured online by a postdetector pH electrode (Lazar Research Laboratories, Los Angeles, CA) for fraction collection at every 0.2 pH unit in the range of pH 7.0-4.0.

6.2.4 Online NPS-RP-HPLC/ESI-TOFMS for intact protein molecular weight determination

Fractions obtained from CF were subjected to NPS-RP HPLC separation using an ODSIII-E column (33×4.6 mm, Eprogen, Darien, IL) packed with 1.5 μm NPS silica. The column was maintained at 60°C by a column heater (model 7971, Jones Chromatography, Resolution Systems, Holland, MI) to perform separation at a flow rate of 0.5 mL/min, where 40% was split into an ESI-TOF MS (LCT Premier, Waters/Micromass, Milford, MA). Formic acid (0.5%, Sigma) was added post splitter using a syringe pump (Harvard Apparatus, Holliston, MA). The remainder of the flow was monitored at 214 nm (detector model 166, Beckman Coulter) for offline peak collection using a fraction collector (model SC100, Beckman Coulter) controlled by an in-house acquisition program. A mobile phase system of A and B was comprised

of 0.1% TFA (J. T. Baker, Phillipsburg, NJ) in purified DI water (Millipore RG system, Bedford, MA) and ACN (Sigma), respectively. The gradient profile was as follows: 5-15% B in 1 min, 15-25% B in 2 min, 25-31% B in 3 min, 31-41% B in 10 min, 41-47% B in 3 min, 47-67% B in 4 min, and 67-100% B in 1 min. The capillary voltage for electrospray was set at 3200 V, sample cone at 35 V, extraction cone at 3 V, and reflectron at 750 V. Desolvation temperature was maintained at 330°C and source at 130°C with desolvation gas flow of 650-800 L/h. Intact protein MW was obtained by automatic deconvolution using Protein Trawler software.

6.2.5 Monolithic Capillary HPLC for Protein Separation

The preparation of copolymerized styrene and divinylbenzene monolithic capillary columns (0.2×66 mm) was performed according to procedures described elsewhere [33]. A capillary pump (Ultra-Plus II MD, Micro-Tech Scientific, Vista, CA) was used for separation. The capillary column was directly mounted on a microinjector with a 500 nL internal sample loop (Valco Instruments, Houston, TX). The capillary HPLC separation was controlled at 60°C with an in-house column heater. The flow from the solvent delivery pump was split precolumn to generate a flow rate of ca. 2.5 μL/min through the monolithic columns. A mobile phase system of two solvents was used, wherein solvents A and B were composed of 0.05% formic acid in HPLC-grade water (Fisher Scientific, Hanover Park, IL) and ACN, respectively. A linear gradient of 0-100% B in 18 min was applied and held at 100% B for 3 min.

6.2.6 Integration of Protein Separation to Automated on-MALDI Plate Enzymatic Digestion

A syringe pump was interfaced with an xyz-module of Nano-Plotter 2.0 piezoelectric pipetting system (GeSiM, Germany) by a capillary tubing in order to precoat the MALDI plate with 0.5 μL of TPCK-modified trypsin solution of 0.15 μg/mL (Promega, Madison, WI) into each spot. Prefractionated esophageal tumor proteins at pH 5.6-5.8 were loaded onto a monolithic capillary column for separation, where the connecting capillary tubing off the column was interfaced directly to the print head of the nano-plotter. An instrument controller program written in nano-plotter language (NPL) was used to control the print head for precise movements along the x, y, and z axes for automatic control over the designated fraction deposition time of 30 s intervals onto the MALDI plate. Fraction collection was performed in a real-time mode with proper calibration. Following the protein collection onto each trypsinized spot, 0.5 μL of 50 mM NH_4HCO_3 (Sigma) was automatically added to the top layer of each spot using a syringe pump and the plate was maintained at room temperature for digestion for around 10 min. Then, 0.5 μL of 0.1% TFA was added to each spot to halt digestion, followed by addition of 0.5 μL of CHCA matrix solution prepared by diluting saturated CHCA with 60% ACN/0.1% TFA at a 1:4 ratio. The internal standards of angiotensin I, adrenocorticotropic hormone (ACTH) fragment 1-17, and 18-39 (all from Sigma) were included to have a final concentration of 50 fmol for every spot.

6.2.7 MALDI-TOF MS Analysis and Database Searching

The MALDI-TOF MS analysis was performed on a Tof-Spec2E (Waters/Micromass) equipped with delayed extraction in reflectron mode using a 337 nm Nd:YAG laser as the ionization source. The delay time was set at 520 ns, the extraction voltage at 1:1 to the source voltage at 20 kV, and the pulse voltage at 2300 V. Each spectrum was internally calibrated and monoisotopic peptide masses were obtained using MassLynx software version 4.0 (Waters-Micromass) for submission to the MS-Fit search engine using Swiss-Prot database for protein identification. The search was carried out under the species of *Homo sapiens* at <100 ppm of mass tolerance with no limitation set for MW and pI. One missed cleavage was allowed and the possible modifications included N-terminal Gln to pyroGlu, oxidation of Met, N-terminal acetylation, and phosphorylation at S, T, and Y. Protein identification was accepted as a match by filtering according to the following parameters: MOWSE score of $>10^3$ and sequence coverage of >20% and comparison with intact MW values.

6.2.8 MALDI-TOF/TOF MS Analysis and Database Searching

The MALDI-TOF/TOF MS analysis was performed using a 4800 MALDI TOF/TOF analyzer (Applied Biosystems, Foster City, CA) with a 384-well plate. A CID voltage of 2 kV was used throughout the runs. S/N threshold of 50 and 30 were used for selecting the precursor and fragment peaks, respectively. The MASCOT available online at `www.matrixscience.com` was used for database searching under Swiss-Prot for Homo sapiens. Mass tolerances of 50 ppm and 0.3 Da were used for precursor and fragment ions, respectively, by allowing one missed cleavage. The same variable modifications were allowed as in PMF analysis.

6.3 Results and Discussion

6.3.1 Design of Automated Platform for Integration of Monolithic LC-based Protein Separation and on-plate Trypsin Digestion

A number of different approaches have been developed to interface HPLC and MALDI-MS. It has recently been shown that the use of sheath gas flow around the emerging droplet of LC effluent assisted solvent evaporation, but prevented the oxidation of analytes [18]. In another study, a pull-down deposition method, applying an electric field to the analyte-collection device, has been demonstrated to effectively deposit a droplet from the HPLC [36]. This method can be particularly useful where very small droplets are generated through nanoscale LC to reduce droplet adhesion to the connecting tubing. However, it may complicate the experimental platform with the use of high voltage applied across the entire analyte-collection device, where electrochemical reactions may adversely affect MS spectra. Others reported a pressure pulse-driven dispenser device [39] and a heated droplet interface [40].

Figure 6.1 describes the schematic of the automated configuration to integrate a monolithic capillary HPLC with onplate digestion of intact proteins for subsequent MALDI-TOF MS analysis. This experimental configuration involves a continuous liquid deposition through contact to the MALDI plate for simplicity, which also proved to provide sufficiently high precision for the purpose of this study. Given the flow rate used for HPLC separation in our experiment, around 1.2 μL from the HPLC is deposited on each spot. Unlike nano-LC applications, this effluent volume is considered to be sufficiently large so that droplet adhesion to the connecting cap-

illary is less of a concern, eliminating the need for a sophisticated platform for liquid collection.

Figure 6.2 shows a very close proximity between the interface of LC effluent and the MALDI plate, which measures around 0.1 mm. Compared to the protein collection and liquid deposition performed manually in our previous study [29], this automated device can ensure highly precise positioning of liquid deposition into a very small area of the spot surface. This provides a method enhancing protein concentration for highly efficient digestion, therefore, detection for improved sensitivity for MS analysis. This added advantage can be particularly useful for analyzing proteins present in low abundance or limited amounts of sample.

The optimal time interval for analyte collection will be dependent on the complexity of the samples [17]. Although a 30 s interval was used for protein collection time in this study, our automation device can be programed for shorter analyte deposition time down to several seconds or less that may be necessary for the analysis of very complex proteomes. The flexibility of the automated system allows multiple numbers of the MALDI plates of varying sizes to be accommodated, which can further enhance throughput of this method.

6.3.2 Identification of Human Esophageal Cancer Tissue Protein

In our previous study, human breast cancer cell line proteins prefractionated by a Rotofor device of preparative scale were chosen as a model to show the applicability of LC/MALDI to biological mixtures [29]. However, the use of the Rotofor device

is less appropriate when the samples to be analyzed are limited in quantity, as it is often the case for human tissue samples. Recently, the usefulness of CF [41] as a fractionation method has been successfully demonstrated for the analysis of various human cancer proteomes [24,25]. In addition to reducing the complexity of samples, it is important to note that experimental pI values of proteins can be monitored online by CF so that the presence of possible PTM can be assessed by comparing to theoretical pI values [42]. In this experiment, we extended the application of the LC/MALDI scheme to human cancer tissue samples, where prefractionation was performed by CF.

Proteins fractionated in the pH range of 5.6-5.8 from an esophageal adenocarcinoma tissue sample was subjected to monolithic HPLC separation interfaced with onplate digestion for subsequent MALDI-TOF MS and MALDI-TOF/TOF MS analyses by automating all liquid handling procedures. In these experiments, a limited amount of CF-fractionated sample was required due to the low loading capacity of the monolithic capillary column [43]. Table 6.1 shows a list of proteins identified through both PMF and intact protein MW analyses in each of the collection times, where the LC system used a delay time of around 6.5 min. A monolith-based protein separation was performed for the collection time of 20 min. As discussed in a recent work [29], a monolithic capillary column of the small id used in the experiment may not be the most suitable means to separate many proteins present in complex mixtures with high resolution and high efficiency. However, a somewhat lower resolution of protein peaks obtained from HPLC separation does not pose a serious problem for reliable protein identification, since subsequent MS analysis can resolve the proteins.

Figure 6.3 illustrates a MALDI-TOF MS spectrum obtained for the splicing factor, proline- and glutamine-rich, a large protein of >70 kDa, that was identified with 43% of sequence coverage, where its identified tryptic peptide sequences are summarized in Table 6.2. As shown in Table 6.1, relatively high sequence coverage was obtained for most of the proteins and this is considered to be primarily due to minimal sample transfers associated with direct LC/MALDI interface to prevent sample loss. Also, compared to traditional in-solution digestion, it appears that protein enrichment effects obtained through a small spot size due to precise control of liquid-handling combined with a large surface area for enzymatic reactions provided by precoated trypsin perhaps resulted in more efficient digestion. Also, reduced trypsin autolysis relative to in-solution digestion may be expected due to trypsin immobilization [44] on the plate to generate less complicated mass spectra for unambiguous interpretation. In order to further confirm the protein identification procedures, LC/MALDI based sequencing analysis was performed using MALDI-TOF/TOF MS. Figure 6.4 shows an example for one of the proteins in Table 6.1, calgranulin B, where one of its tryptic peptides, NIETIINTFHQYSVK (11-25) has been successfully sequenced. This protein was identified with >50% of sequence coverage through sequencing analysis.

Also, since proteins are sufficiently separated from each other for direct deposition onto the MALDI target plate to be confined into a very small spot size with minimal sample loss expected, several less abundant proteins were identified in this work, such as cyclin H [45] in Table 6.1. Another protein, MAPK-interacting serine/threonine

kinase 1 [46], activated by phosphorylation, is also known to be expressed at low levels. It would be difficult to detect these proteins using a traditional proteomics approach, where sample loss due to sample-handling procedures is inevitable.

In addition to helping to confirm protein identifications determined by PMF and sequencing analyses, intact protein MW values were obtained separately by NPS-RP HPLC/ESI-TOF MS for comparison to theoretical MW values with excellent agreement, as shown in Table 6.1. Although the LC/MALDI method alone can provide highly confident protein identification due to relatively high sequence coverage, the usefulness of intact MW values can be observed from a close examination of several proteins. For instance, the splicing factor, proline- and glutamine-rich, is annotated with two distinct isoforms A and F (Acc. nos. P23246-1 and P23246-2, respectively) that slightly differ in amino acid compositions in the region of 663-707 to result in the isoform A being larger than the isoform F by around 4000 Da (`www.expasy.org`).

Protein identification by PMF or tandem MS-based sequencing analysis is solely based on partial sequence coverage of proteins, where sequence variations, such as truncation and isoforms, are often undetected. In our experiment, however, a comparison between theoretical and experimental intact protein MW allows us to conclude that it is likely that the isoform A of splicing factor, proline and glutamine-rich, is present in the esophageal tumor sample. Also, significant deviation observed between the experimental and theoretical protein MW values can suggest the presence of possible modifications in certain proteins. An example can be found from ATP synthase beta chain that becomes truncated by the loss of transit peptide sequence

upon entry into mitochondria, where this can also be explained by an excellent agreement between its theoretical intact MW value and experimental MW value, as shown in Table 6.1. A close match between theoretical and experimental MW values of all proteins in Table 6.1 further confirms the protein identification obtained by PMF analysis, where slight deviations can suggest the presence of potential PTMs [24].

6.4 Conclusiuons

We have developed a novel method that automates the integration of monolithic capillary HPLC separation of intact proteins and on-plate enzymatic digestion for MALDI-based MS analysis. The method is shown to be highly versatile and robust and is suitable for the analysis of complex proteomes, as successfully demonstrated for the analysis of esophageal tissue proteins. It is important to emphasize that the current study analyzed the proteome at the protein level, unlike the majority of previous LC/MALDI work performed by shotgun proteomics that may result in ambiguous identifications due to the complexity of mixtures upon digestion of a large number of proteins. Based on the measured intact protein MW values and substantially higher protein sequence coverage from the protein-based LC/MALDI, compared to shotgun-derived LC/MALDI, it is likely that the developed method can help analyze PTM or other sequence variations. Although the analysis of complex proteomes is often performed by ESI-based MS, integration of protein separation directly to MALDI is an effective means to solve compatibility issues with various LC solvent compositions and contaminants to which the ESI process is substantially less tolerant.

Automated coupling of LC-based protein separation for direct on-plate enzymatic digestion is an attractive means to realize high-throughput proteomics, where the process can be fully automated and enhance reproducibility. Another distinct advantage includes accurate control of the spot surface area upon deposition that allows for the analysis of low-abundance proteins by enrichment of analytes [47]. This is essential in proteomics studies, where a wide dynamic range of protein concentration presents a great challenge. Considering that the droplet size is dependent on the surface tension and viscosity of the solution [48], further development is ongoing to diminish the spot diameter. The method has potential to become even more powerful when protein identification through automated software analysis can be combined.

Table 6.1: List of proteins identified by automated interfacing of monolithic HPLC with MALDI-TOF MS through on-plate digestion for Barrett's esophageal adenocarcinoma samples prefractionated by CF at pH 5.6-5.8

Collection time (min)	Protein name	Accession no.	Theoretical		Experimental MW	Sequence coverage (%)	MOWSE score
			MW	p*I*			
6.5–7	ATP synthase beta chain, mitochondrial precursor	P06576	56 560 (51 736)[a)]	5.3	51 636	36	1.397e + 04
9–9.5	Neuron-specific calcium-binding protein hippocalcin	P84074	22 428	4.9	22 574	48	5041
10–10.5	NADH-ubiquinone oxidoreductase B16.6 subunit	QP0J0	16 698	8.0	16 682	45	1.268e + 04
11–11.5	Propionyl-CoA carboxylase beta chain	P05166	58 206	7.6	57 695	30	3182
12.5–13	Paraplegin	Q9UQ90	88 136	8.7	88 634	25	6.390e + 05
13.5–14	Triosephosphate isomerase	P60174	26 670	6.4	26 652	40	1.456e + 04
	Heterogeneous nuclear ribonucleoprotein A1	P09651	38 846	9.3	39 461	33	3.354e + 05
14.5–15	Splicing factor, proline- and glutamine-rich	P23246	76 150	9.5	75 490	43	1.113e+12
15–15.5	Aldo-keto reductase family 1 member C3	P42330	36 845	8.1	36 886	52	2.008e + 07
16–16.5	Calgranulin B	P06702	13 242	5.7	13 816	66	8.249e + 04
17–17.5	Lymphoid-enhancer-binding factor 1	Q9UJU2	44 201	6.9	44 549	20	1.104e + 04
18–18.5	HLA class I histocompatibility antigen	Q29718	40 421	5.7	40 195	35	1.815e + 04
18.5–19	Aldehyde dehydrogenase 1A1	P00352	54 862	6.3	54 862	26	6.096e + 04
19–19.5	39S Ribosomal protein L45	Q9BRJ2	35 351	9.1	35 214	26	4949
19.5–20	Tumor necrosis factor receptor superfamily member 11B precursor	O00300	46 041	8.7	45 799	32	1.943e + 04
	Cytokine-inducible SH2-containing protein	Q9NSE2	28 663	6.5	28 635	47	1.415e + 04
21–21.5	MAP kinase-interacting serine/threonine kinase 1	Q9BUB5	51 343	6.3	50 686	38	2.607e + 04
22.5–23	Spindlin (ovarian cancer-related protein)	Q9GZT2	27 077	5.8	27 246	30	1.059e + 05
23.5–24	Transgelin-3	Q9UI15	22 473	6.8	22 878	38	6968
24.5–25	Alcohol dehydrogenase	P14550	36 573	6.3	36 531	25	1.415e + 04
25–25.5	Cyclin H	P51946	37 644	6.7	37 533	43	2496
26.5–27	Peptidyl-prolyl *cis–trans* isomerase	Q13451	51 213	5.7	50 686	31	4.837e + 04

a) The number in parentheses indicates theoretical MW of the truncated form.

Table 6.2: A list of sequenced tryptic peptides identified from splicing factor, proline- and glutamine-rich, where peptides with * are shown in Figure 6.3

Exp. $[M+H]^+$	Theor. $[M+H]^+$	Possible modifications	Start	End	Missed cleavages	Database sequence
770.3267	770.3321		594	599	0	EESYSR
830.4011	830.4484		543	548	1	QEELRR
830.4011	830.4484		542	547	1	RQEELR
*869.3300	869.3650		600	606	0	MGYMDPR
886.4747	886.4998		400	407	0	AVVIVDDR
1120.4904	1120.5210		682	693	0	GMGPGTPAGYGR
*1143.5992	1143.6275		366	376	0	FATHAAALSVR
*1170.5725	1170.5036	1Met-ox	600	608	1	MGYMDPRER
1235.5787	1235.5149	PyroGlu 1Met-ox	582	590	1	QREMEEQMR
1235.5787	1235.6497		287	296	1	RPGEKTYTQR
1245.6538	1245.6955		414	425	0	GIVEFASKPAAR
*1252.5765	1252.5414	1Met-ox	582	590	1	QREMEEQMR
*1252.5765	1252.6214		320	330	0	YGEPGEVFINK
*1267.5986	1267.6217		33	44	0	SPPPGMGLNQNR
*1341.6605	1341.6664		667	681	0	FGQGGAGPVGGQGPR
*1572.7620	1572.7263		549	560	1	MEELHNQEMQKR
*1572.7620	1572.7263		548	559	1	RMEELHNQEMQK
*1743.8246	1743.8951		343	358	1	ALAEIAKAELDDTPMR
*1762.7778	1762.7825		480	493	0	FAQHGTFEYEYSQR
*1807.8894	1807.8241	1PO4	664	681	1	TERFGQGGAGPVGGQGPR
*1807.8894	1807.9118		299	314	0	LFVGNLPADITEDEFK
*1963.9795	1964.0129		299	315	1	LFVGNLPADITEDEFKR
*2264.9968	2264.9647	1PO4	519	536	0	LESEMEDAYHEHQANLL
*2370.1633	2370.2253		212	236	0	MPGGPKPGGGPGLSTPGGHPKPPHR
*2403.0925	2403.1356		246	267	0	QHHPPYHQQHHQGPPPGGPGGR
*2428.0911	2428.1203		517	536	1	DKLESEMEDAYHEHQANLLR
*2639.1956	2639.2993		377	399	0	NLSPYVSNELLEEAFSQFGPIER
*3195.3171	3195.4372		631	663	0	FPPLGGGGGIGYEANPGVPPATMSGSMMGSDMR
*3467.1528	3467.3260	2Met-ox 3PO4	631	663	0	FPPLGGGGGIGYEANPGVPPATMSGSMMGSDMR

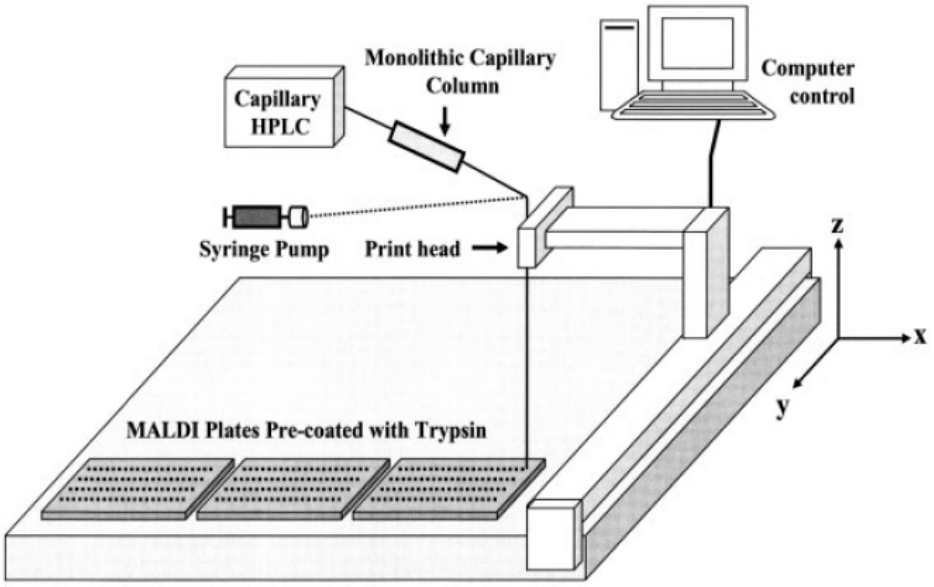

Figure 6.1: A simplified diagram of automated LC/MALDI configuration constructed by modifying nano-plotter (not to scale) for all liquid-handling procedures.

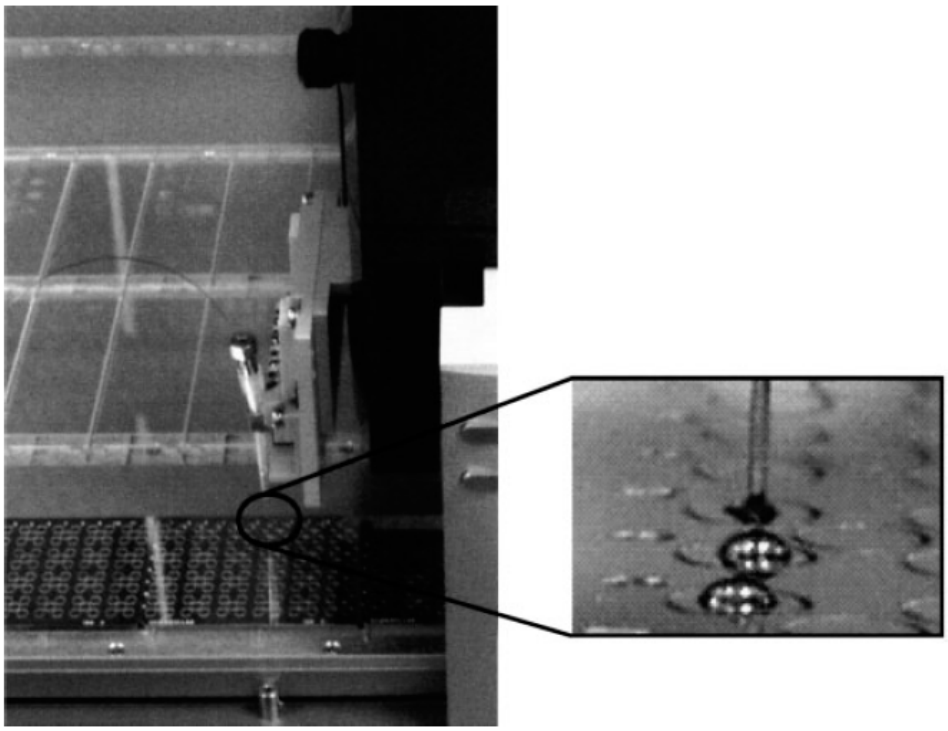

Figure 6.2: Automated LC/MALDI interface where proteins separated by capillary monolithic HPLC are deposited directly onto the MALDI target plate precoated with trypsin.

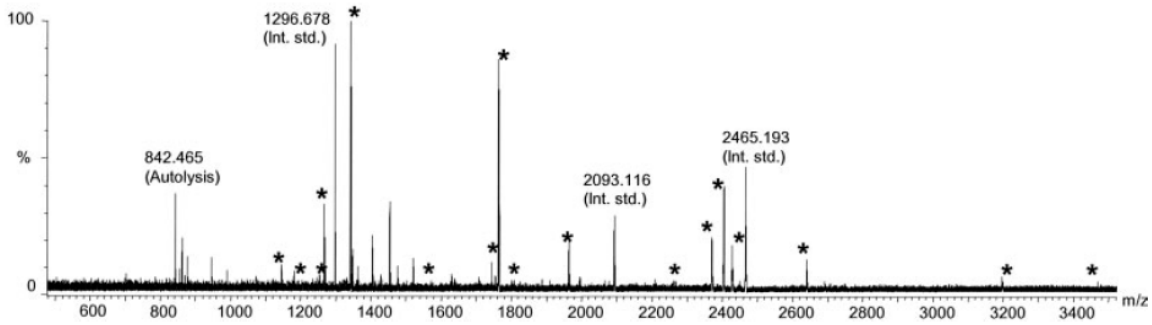

Figure 6.3: MALDI-TOF MS spectrum (unprocessed) obtained for the splicing factor, proline- and glutamine-rich, from an esophageal tissue sample. * Indicates peptides identified by monolithic LC/MALDI scheme through on-plate digestion. Refer to Table 6.2 for information regarding each peptide.

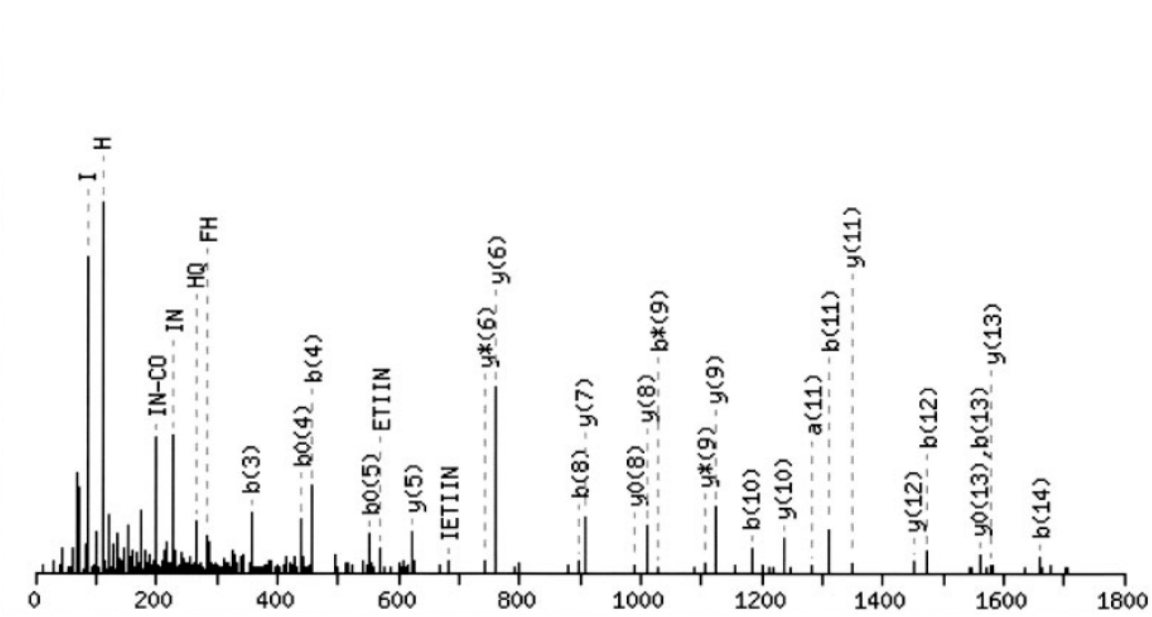

Figure 6.4: MALDI-TOF/TOF MS spectrum for a tryptic peptide NIETIINTFHQYSVK (11-25) of calgranulin B.

References

[1] Han K. K.; Martinage A. *Int. J. Biochem.*, 24, 19–28, 1992.

[2] Cagney G.; Amiri S.; Premawaradena T.; Lindo M.; Emili A. *Proteome Sci.*, 1:5, 2003.

[3] Quadroni M.; James P. *Electrophoresis*, 20:664–677, 1999.

[4] O'Farrell P. H. *J. Biol. Chem.*, 250:4007–4021, 1975.

[5] Weeks M. E.; Sinclair J.; Jacob R. J.; Saxton M. J. et al. *Proteomics*, 5:1669–1685, 2005.

[6] Malmstrom J.; Larsen K.; Malmstrom L.; Tufvesson E. et al. *Electrophoresis*, 24:3806–3814, 2003.

[7] Klene T. G.; Andreasen c. M.; Kjeldal H. O.; Leonardsen L. R. et al. *Anal. Chem.*, 76:3576–3583, 2004.

[8] Foret F.; Preisler J. *Proteomics*, 2:360–372, 2002.

[9] Zhong H.; Marcus S. L.; Li L. *J. Am. Soc. Mass Spectrom.*, 16:471–481, 2005.

[10] Brombacher S.; Owen S. J.; Volmer D. A. *Anal. Bioanal. Chem.*, 376:773–779, 2003.

[11] Keil O.; LeRiche T.; Deppe H.; Volmer D. A.; *Rapid Commun. Mass Spectrom.*, 16:814–820, 2002.

[12] Tegeler T. J.; Merchref Y.; Boraas K.; Reilly J. P.; Novotny M. V.; *Anal. Chem.*, 76:6698–6706, 2004.

[13] Zhen Y.; Xu N.; Richardson B.; Becklin R. et al. *J. Am. Soc. Mass Spectrom.*, 15:803–822, 2004.

[14] Griffin T. J.; Gygi S. G.; Rist B.; Aebersold R. et al. *Anal. Chem.*, 73:978–986, 2001.

[15] Lochnit G.; Geyer R. *Biomed. Chromatogr.*, 18:841–888, 2004.

[16] Canelle L.; Pionneau C.; Marie A.; Bousquet J. et al. *Rapid Commun. Mass Spectrom.*, 18:2785–2794, 2004.

[17] Young J. B.; Li L. *J. Am. Soc. Mass Spectrom.*, 17:325–334, 2006.

[18] Mirgorodoskaya E.; Braeuer C.; Fucini P.; Lehrach H.; Gobom J. *Proteomics*, 5:399–408, 2005.

[19] Hattan S. J.; Marchese J.; Khainovski N.; Martin S.; Juhasz P. *J. Proteome Res.*, 4:1931–1941, 2005.

[20] Chen V. C.; Cheng K.; Ens W.; Standing K. G. et al. *Anal. Chem.*, 76:1189–1196, 2004.

[21] Wang Y.; Rudnick P. A.; Evans E. L.; Li J. et al. *Anal. Chem.*, 77:6549–6556, 2005.

[22] Veenstra T. D.; Conrads T. P.; Issaq H. *Electrophoresis*, 25:1278–1279, 2004.

[23] Slysz G. W.; Schriemer D. C. *Anal. Chem.*, 77:1572–1579, 2005.

[24] Kreunin P.; Urquidi V.; Lubman D. M.; Goodison S. *Proteomics*, 4:2754–2765, 2004.

[25] Zhu K.; Miller F. R.; Barder T. J.; Lubman D. M. *J. Mass Spectrom.*, 39:770–780, 2004.

[26] Zheng S.; ONeil K. A.; Barder T. J.; Lubman D. M. *Bio-Techniques*, 35:1202–1212, 2003.

[27] Zhou F.; Johnston M. V. *Electrophoresis*, 26:1383–1388, 2005.

[28] Harris W. A.; Reilly J. P. *Anal. Chem.*, 74:4410–4416, 2002.

[29] Zheng S.; Yoo C.; Miller F. R.; Huber C. G.; Lubman D. M. *Anal. Chem.*, 78:5198–5204, 2006.

[30] Kato M.; Sakai-Kato K.; Jin H.; Kubota K. et al. *Anal. Chem.*, 76:1896–1902, 2004.

[31] Zou H.; Huang X.; Ye M.; Luo Q. *J. Chromatogr. A*, 954:5–32, 2002.

[32] Hemstrom P.; Nordborg A.; Irgum K.; Svec F.; Frchet J. M. J. *J. Sep. Sci.*, 29:25–32, 2006.

[33] Premstaller A.; Oberacher H.; Walcher W.; Timperio A. M. et al. *Anal. Chem.*, 73:2390–2396, 2001.

[34] Walcher W.; Toll H.; Ingendoh A.; Huber C. G. *J. Chromatogr. A*, 1053:107–117, 2004.

[35] Lopez M. F. *Electrophoresis*, 21:1082–1093, 2000.

[36] Ericson C.; Phung Q. T.; Horn D. M.; Peters E. C. et al. *Anal. Chem.*, 75:2309–2315, 2003.

[37] Shaheen N. J. *Gastroenterology*, 128:1554–1566, 2005.

[38] Devesa S. S.; Blot W. J.; Fraumeni J. F. *Am. Cancer Soc.*, 83:2049–2053, 1998.

[39] Onnerfjord P.; Nilsson J.; Wallman L.; Laurell T.; Marko-Varga G. *Anal. Chem.*, 70:4755–4760, 1998.

[40] Zhang B.; McDonald C.; Li L. *Anal. Chem.*, 76:992–1001, 2004.

[41] Sluyterman L. A. A.; Elgersma O. *J. Chromatogr.*, 150:17–30, 1978.

[42] Zhu K.; Zhao J.; Lubman D. M.; Miller F. R.; Barder T. J. *Anal. Chem.*, 77:2745–2755, 2006.

[43] Oberacher H.; Premstaller A.; Huber C. G. *J. Chromatogr. A*, 1030:201–208, 2004.

[44] Gabel D.; Kasche V. *Biochem. Biophys. Res. Commun.*, 48:1011–1018, 1972.

[45] Lolli G.; Johnson L. N. *Cell Cycle*, 4:572–577, 2005.

[46] Parra J. L.; Buxade M.; Proud C. G. *J. Biol. Chem.*, 280:37623–37633, 2005.

[47] Miliotis T.; Kjellstrom S.; Nillson J.; Laurell T. et al. *J. Mass Spectrom.*, 35:369–377, 2000.

[48] Daniel J. M.; Laiko V. V.; Doroshenko V. M.; Zenobi R. *Anal. Bioanal. Chem.*, 383:895–902, 2005.

CHAPTER VII

Conclusions

Studies in proteomics present us with a direct approach for investigating disease at the individual patient level. Since most therapeutic targets are proteins, proteomics is vital for developing methods for cure. The human proteome is so more complex than previously assumed, that it is almost impossible to depend on one single method for its reliable in-depth characterization. The ultimate goal of application of proteomics in personalized clinical diagnostics requires methods with higher sensitivity, dynamic range, throughput and multiplexing capability than the traditional methods in use today. The multitude of methods presented here utilizing the power of mass spectrometry, multidimensional separations and microarray based discovery, detection, identification and characterization techniques attempts to develop a holistic approach to realize that goal.

Over the years, mass spectrometry has demonstrated its critical role in biomolecule analysis and is used almost exclusively in proteomics today. But mass spectrometry has very little to no capability in analyzing complex mixtures such as the proteome, unless they have been fractionated and purified of salts and other interfering agents. In this respect, many factors including ease in interfacing makes multidimensional liq-

uid phase separations the method of choice for mass spectrometry based proteomics. The work presented in this thesis applies multidimensional separations for generating reverse phase protein microarrays and applies mass spectrometry to identify proteins of interest.

The majority of studies in biomarker discovery has been limited to individual proteins, but it is gradually becoming clear that information on entire networks may be necessary for robust diagnostics. The use of microarrays provides an excellent platform for such large scale parallel analysis. New types of protein arrays combined with advanced bioinformatics for data processing have been presented here to identify molecular signatures of individual tumors based on protein pathways and signaling cascades. Global protein phosphorylation detection and discovery methods have been realized using a small molecule phospho-sensor dye. This provides an improved alternative than antibody based approaches but despite the speed and sensitivity of dye-array techniques must be enhanced significantly for detection of the biologically relevant proteins. Using nitrocellulose slides with high adsorption capacity, autoantibody response methods have been used to obtain biomarker pools for improved sample discrimination. Correlations between classification based on biomarker pools and prostate cancer gleason grades presented in Chapter IV demonstrate the power of such techniques. Multiplexed analysis from microarrays where complementary molecular concept modeling approaches have helped discover signaling and metabolic pathways have been verified through experimental means to be a practical approach for diagnosis. The use of piezo-electric pins for non-contact printing have significantly reduced spot variability as is evident from comparisons

between array images presented in Chapter II and IV where quill-type pins had been used for contact printing in the earlier work. However, subtle differences in protein expression levels may still be difficult to detect owing to factors mostly related to solution phase chemistry where uneven background staining from local variations in concentration or solvent hydrodynamics can degrade the quality of microarray data. Great care had been taken for the experiments described here and all comparative experiments were conducted the same day whenever possible to eliminate any day-to-day variations. Subtle non-critical procedural improvements were readily applied whenever possible and have often provided huge improvements in the quality of data obtained.

More than half of the serum and a significant portion of the cellular or tissue proteome contain proteins which have no role in signaling or are not useful in diagnostics. Multidimensional liquid separation techniques have been able to fractionate such samples so that the low abundant proteins are available for analysis. The work presented in this thesis describes a monolithic capillary LC based technique for analysis of low concentrations of proteins and may be useful in the study of the low abundant proteins in the proteome. The above method was applied for on-line interfacing with both ESI- and MALDI-MS based techniques. Both techniques provide highly efficient separations of protein digests within several minutes and peak widths of only a few seconds thereby providing high peak capacities suitable for fast high-throughput analysis. LC separations using the polymer based monolithic capillary was described as an alternative to CE in obtaining nearly the same peptide coverages. More rugged than CE-MS, monolithic-LC MS provided peak capacities of nearly 100 within a 10

min window. With extremely high sequence coverage obtained from this method, several sequence variations could be observed among proteins. Monolithic-LC/MALDI method on the other hand demonstrates the ability of such columns for intact protein separations and robot assisted MALDI spotting was used to increase the throughput. This method also provides higher sensitivity because there is very little sample loss between the separation and mass spectrometry based identification steps. This method also allows for scaling down proteomic analysis so that 2-D virtual maps can be obtained from as little as 100 μg of sample.

The multiplexed high-throughput proteomics platforms and micro-scale liquid separation methods implemented in this research may help in the detection of low abundant proteins and their associated characterization using mass spectrometry. The ability to analyze lower amounts of sample will be helpful in analyzing alternative viable sources of biomarkers, eg. buccal cells so that strategies to amplify and harvest biomarkers will greatly enhance the capabilities of current proteomics modalities. Additional methods of validation of proteins identified through mass spectrometry will increase the reliability of such techniques.