Winter 2006

The Use of Proteomic Technologies to Identify Serum Glycoproteins for the Early Detection of Liver and Prostate Cancers

Elizabeth Ellen Schwegler
Old Dominion University

THE USE OF PROTEOMIC TECHNOLOGIES TO IDENTIFY SERUM GLYCOPROTEINS FOR THE EARLY DETECTION OF LIVER AND PROSTATE CANCERS

by

Elizabeth Ellen Schwegler
B.S. May 2002, Baylor University

A Dissertation Submitted to the Faculty of
Eastern Virginia Medical School and Old Dominion
University in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

MOLECULAR AND INTEGRATIVE BIOMEDICAL SCIENCES

EASTERN VIRGINIA MEDICAL SCHOOL
AND OLD DOMINION UNIVERSITY
DECEMBER 2006

Approved by:

Richard Drake (Director)

Richard Ciavarra (Member)

Julia A. Kerry (Member)

O. John Semmes (Member)

ABSTRACT

THE USE OF PROTEOMIC TECHNOLOGIES TO IDENTIFY SERUM GLYCOPROTEINS FOR THE EARLY DETECTION OF LIVER AND PROSTATE CANCERS

Elizabeth Ellen Schwegler
Eastern Virginia Medical School and Old Dominion University, 2006
Director: Dr. Richard R. Drake

The application of proteomic technologies to identify serum glycoproteins is an emerging technique to identify new biomarkers indicative of disease severity. Many of these newly evolving protein-profiling methodologies have evolved from previous global protein expression profiling studies such as those involving SELDI-TOF-MS technologies. Though the SELDI approach could distinguish disease from normal by utilizing protein patterns as shown herein with the HCC study of chapter II, it was unable to offer sequence information on the selected peaks, and did not have the ability to analyze the entire dynamic range of the serum/plasma proteome. To address these deficiencies, new strategies that incorporate the use of differential lectin-based glycoprotein capture and targeted immuno-based assays have been developed. The carbohydrate binding specificities of different lectins offers a biological affinity approach that both complements existing mass spectrometer capabilities and retains automated throughput options. A prostate cancer study using disease stratified samples is utilized herein to determine whether lectin capture can identify glycoproteins, which are indicative of different stages of prostate disease. By utilizing upfront lectin fractionation we show here evidence of glycoproteins and glycoprotein isoforms, which are specific to cancer progression. In addition, the incorporation of lectin fractionation followed by albumin depletion allows for a more in depth analysis of the entire dynamic range of the human serum and plasma proteome. Taken together we believe this approach is an attractive strategy for the discovery of proteins indicative of the early detection of liver and prostate cancers.

This dissertation is dedicated to my husband, Lawrence, and everyone who helped me during this process.

ACKNOWLEDGMENTS

I would like to thank my committee for their continuous support and guidance through this entire process. In addition I would like to specially thank my thesis advisor Dr. Richard Drake for always providing such insight and care concerning my future and research project.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

1.1 Introduction and Goals of Dissertation

The successful completion of the Human Genome Project in 2002 left an immense impact on the scientific community since, for the first time a comprehensive list of all genes within the human genome was readily accessible (1-3). Also forthcoming from this work is an interesting discovery regarding protein expression, in that less than 30,000 genes within the genome have encoding regions, yet there are over 400,000 known proteins and likely many more thousands uncharacterized (1-3). The combination of a defined human genome with numerous uncharacterized proteins has sparked great interest in the identification and characterization of biomarker proteins associated with disease onset and/or progression (4). In proteomics, a biomarker is defined as a "protein or protein fragment" which is easily detectable in a patient's sample, usually blood or urine that is specific to a disease or disease stage (5). The goal of this thesis research project is to use current proteomic technologies to determine whether disease specific protein expression patterns exist in clinical fluids of prostate and liver cancer patients. Whether these protein patterns can be used to predict and monitor cancer occurrence and progression will also be determined.

The specific hypothesis of this thesis is that the tumorogenic progression of liver and prostate cancers produces distinct disease-associated biomarker proteins derived from the tumor or surrounding environment that are detectable in clinical fluids, in particular serum or plasma. These newly discovered tumor associated proteins could then be further assessed for the development of novel diagnostic cancer assays, such as a blood test for cancer detection or reoccurrence (6-8). In theory, disease specific proteins could be shed directly from a tumor cell, or from cells in the tumor microenvironment, and include a range of protein classes such as truncated protein products, proteins with altered post translational modifications, and up-regulated

---

This manuscript is modeled after the Journal of Cancer Research.

proteins working to enhance the growth and maintenance of tumor cells (4, 7, 9). Fundamental to this biomarker discovery is the belief that newly discovered biomarkers will not only be able to enhance current clinical diagnostic tests, but also serve as prognostic markers of disease progression providing clues to the molecular mechanisms leading to cancer progression.

Since current proteomic technologies are a key component of this research project, a brief history of the evolution of proteomic technologies and their role in biomarker discovery will be included. This section will provide an overview of the specific methodologies used within this project and other techniques being used for detection of protein biomarkers. This section will then be followed by an overview of the clinical properties, detection strategies and pathology of liver and prostate cancers in humans, the two cancers assessed within this project.

1.2 Proteomics: History and Instrumentation

The common use of the term "proteome" found its origins in the early 1990s as a word to refer to the "protein complement of the genome," although to this point it had been used to describe 1970's and 1980's biochemistry based techniques useful for protein separation and small molecule analysis (10). During this same period, protein sequencing was achieved using Edman degradation, a method which owes its successes to the early accomplishments of the Nobel prize winner Fredrick Sanger whom, following 10 years of intense research, produced information describing the complete chemical structure of insulin (2, 11). Sanger's early insulin studies were accomplished by use of a novel method, which labeled the end terminus of a protein or peptide to be sequenced. This technique is known as the DNP (dinitrophenyl group) labeling method. DNP labeling works by covalently modifying the amino terminus of an amino acid in a peptide. This modification can then be used as a chemical marker of the peptide, since it will continue to be attached to the amino group following hydrolysis (2, 11). Following hydrolysis the complete sequence of the peptide can be obtained by aligning all of the peptides and looking for overlaps in their sequences (2, 11, 12). Building on these end labeling sequencing methodologies of Sanger, Pehr Edman, a Swedish biochemist, went on to further develop his namesake the Edman Degradation, for protein sequencing reactions (2, 12).

The process of Edman degradation can be divided into three steps: coupling, cleavage and conversion (2). The coupling step uses phenylisothiocyanate (PITC) which reacts with the N-terminus of a protein in a chemical reaction, forming a cyclic intermediate known as PTC or (phenylthiocarbamyl). In the cleavage reaction, the PTC amino-terminal product is cleaved by anhydrous acid, which ensures that cleavage will occur at the amino terminal peptide bond. The cleavage reaction will allow the release of the amino terminal amino acid, which can be extracted from the remaining polar polypeptide chain due to its newly acquired hydrophobicity. The final step, conversion, occurs when the unstable free amino acid is converted into a more stable product (2, 12-15). The identity of the released amino acid can now be determined using analytical methods such as chromatography or electrophoresis. This "stepwise degradation" can continue for numerous cycles, removing one amino acid at a time for as many cycles as necessary (2, 12-15). Although Edman degradation is a reliable method for determining protein structure and sequence, at this time it required extensively purified, microgram amounts of sample and, furthermore was a slow process lacking the higher throughput capabilities currently associated with mass spectrometry based proteomic techniques (10).

Concurrently, two-dimensional gel electrophoresis (2DE) methodologies were being described as an alternative separation strategy for complex protein samples (2, 16, 17). Unlike one dimensional denaturing gel electrophoresis, which separates proteins by mass, 2DE offers further fractionation of proteins by separating proteins in two dimensions. The first dimension separates proteins based on their PI, or bioelectric point, while the second dimension separates proteins based on molecular weight (2, 16, 17). Though 2DE is an effective strategy for separating complex protein mixtures that might not otherwise be separated, this technique suffers various limitations (2, 10, 12, 16, 17).

For example, the original 2-D gels suffered from a variety of reproducibility issues largely resulting from the need to cast the first dimension isoelectric focusing gels in glass tubes. Especially for studies done on human clinical samples, it also became apparent that the dynamic complexity (e.g. post-translational modifications) and large disparate concentration range of the proteome contributed to the lack of reproducibility (2, 12). The introduction of commercially

available immobilized pH gradients (IPGs) on thin plastic strips did resolve many of the issues associated with first dimension IEF reproducibility. Also, further advancements allowed for 2DE methodologies to directly correlate protein expression differences between two disease subtypes, in a quantitative manner (18-20). Yet, nothing could account for the fact that when used alone 2DE allowed only a small part of the proteome to be examined, and these proteins analyzed were most likely only the most abundant proteins within the sample. This dynamic range issue continues to be a major hindrance to proteomic studies aimed at developing clinical fluid-based diagnostic tests. Furthermore, when 2DE analysis was coupled with Edman degradation for amino acid sequencing of peptides, it was clear that it would not be feasible for high throughput analysis, in part because Edman sequencing reactions can only sequence proteins with at most 60 amino acids, which was a large disadvantage (10).

Fortunately, around this time long-standing problems associated with the production of gas-phase ions of large molecules like proteins were being solved, offering new opportunities for these samples to be analyzed by mass spectrometry methods (2, 12). The first attempts to ionize proteins were electron and chemical ionization approaches, which included a vaporization step amenable only to small peptides with a maximum of 4 to 5 amino acids (12). Fast atom bombardment (FAB) was the first ionization method to allow analysis of proteins by mass spectrometry, developed in 1981 (12). The FAB experiments of this time were always run in tandem with Edman degradation reactions to ensure they could achieve the success of the Edman methods. In 1990, two new ionization methodologies capable of ionizing large molecules were made commercially available, electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). These technologies are commonly referred to as soft ionization methods since they efficiently ionize a protein for identification without causing fragmentation of the protein/peptides. Therefore, soft ionization methods do not offer structural information; making it necessary to couple these technologies with tandem mass spectrometry. Tandem mass spectrometry allows the molecular weight of various peptides to be assessed in an initial scan, and then a particular peptide of interest can be selected and re-fragmented using energetic dissociation to see its product ions, thus allowing structural characterization of a protein (12, 21,

22). Emergence of soft ionization technologies coupled with tandem mass spectrometry essentially revolutionized approaches to protein sequencing, allowing the mass spectrometer to emerge as a mainstream technology for high throughput protein sequencing experiments, thus quickly replacing the use of Edman degradation reactions (12, 21, 23). Electrospray and MALDI technologies are distinct approaches that achieve ionization in different ways. Electrospray ionization requires the peptides of interest to be dissolved in an acidic organic solution, which is then sprayed through a small diameter needle. A voltage applied to the needle produces droplets of the analyte, and protons from the acidic solution give the droplets a positive charge, causing them to move from the needle to the negatively charged instrument (2, 12). As the positive particles migrate towards the instrument, the droplets undergo a series of evaporation and condensation reactions. The evaporation process is aided by a flow of gas, typically nitrogen, and heat. This evaporation and condensation process repeats until the small size and charge of the droplets desorbs protonated peptides into the gas phase where they are directed into the mass spectrometer detector by different electric fields (2, 12). Electrospray is usually coupled to ion trap instruments, and liquid chromatography is also widely used with these technologies.

Matrix-assisted laser desorption/ionization, often referred to as 'MALDI,' achieves ionization by dissolving peptides into a solution of an energy-absorbing matrix, which is then placed on a target plate and inserted into the mass spectrometer. The idea is to saturate the solvent with an abundance of the matrix material, which co-crystallizes with the sample. This is highlighted in Figure 1 (12). Matrix choice is dependent upon the laser within the instrument, being either IR (infrared) or UV (ultraviolet), with the UV being more widely used. There are many different matrix compounds available, and their use is dependent upon the experimental design of the project. For example, α-cyano-4-hyroxycinnaminic acid (CHCA) is a good matrix for use with most samples allowing analysis of proteins and peptides of masses less then 10,000 Daltons. On the other hand if glycoproteins or glycolipids are being analyzed, 2,5-dihydroxybenzoic acids which are completely water soluble would be a more compatible choice (2, 12). In a MALDI instrument, the UV laser works by firing short pulses of laser light at the target, which vaporizes the matrix and peptide ions converting them to a gas phase. Thus,

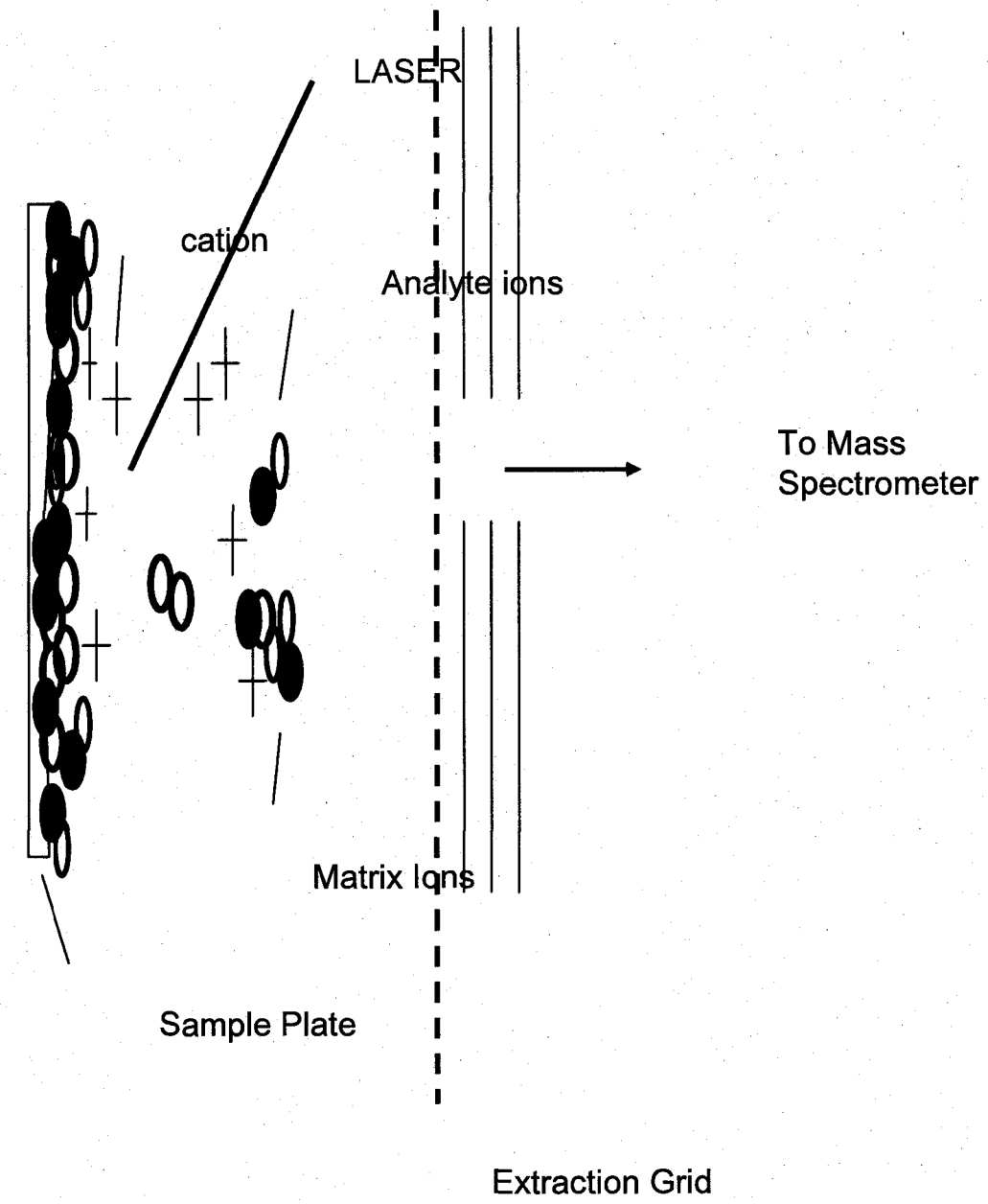


Figure 1. Saturation of Solvent in Matrix (MALDI-TOF/TOF). The matrix allows co-crystallization of solvent and sample so they can travel down the flight tube into the mass spectrometer.

ionization occurs by protonation of the solvent facilitated by the acidic matrix solution added to the sample (12).

MALDI is usually used with time of flight (TOF) analyzers. Time of flight instruments measure the mass to charge ratio (m/z) of the analyte by measuring the time it takes the ions to travel a distance down a flight tube and reach the detector, which should be proportional to their mass. Tandem TOF, or TOF/TOF, instruments contain two time of flight regions, separated by a collision source. This allows ions to be accelerated through the first TOF via linear mode to gather a scan of the peptides relative to their mass to charge values. If further fragmentation is warranted, a parent ion can be selected from the first scan by molecular mass, and only this parent ion will be introduced into the collision cell where it will be fragmented by an inert gas within a voltage cell (usually helium). The data from the fragmentation of the parent ion can now be used to search a genome database and determine the protein sequence (2, 10, 12). The MALDI-TOF/TOF can also operate in MS or reflectron mode which is utilized for peptide mass fingerprinting of protein samples previously digested with an appropriate enzyme, most often trypsin. In reflectron mode the ions will travel down the flight tube, but be reflected back by an ion mirror at the end of the tube. This lengthens the flight time of the ions, and compensates for similarly charged ions, which may have slightly different energies. This feature thereby produces higher mass accuracies and improved resolution. Whether using peptide mass fingerprinting in MS mode or fragmentation of parent ions in MS/MS mode, both approaches rely upon peptide database search engines and algorithms for the identification of unknown proteins (2). Peptides with the same m/z value will be broken down into two fragments forming N terminal ions or C terminal ions. This peptide fragmentation will generate a peak map, with the peaks corresponding to the ions, and the peak intensity representative of the frequency of ions with a particular m/z value (24). The peak masses will then be labeled and compiled into a peak list, which can then be used to search the appropriate database to find amino acid combinations that correspond to the indicated masses, and therefore indicate the possible identity of the protein (12, 24). There are numerous protein databases available for this analysis. Deciding on a database is reliant upon

personal preference, experimental design and the amount of redundancy within the database the investigator is willing to tolerate (12).

There are three basic types of public sequence databases currently available: primary nucleotide sequence databases, comprehensive nucleotide sequence databases and curated protein sequence databases (12). For example, the GenBank database is a nucleotide sequence database including over 5,691,170 total entries. These types of databases are notoriously complex since they have been constructed based upon sequence submissions from individual scientists. Although they have broad coverage, large databases suffer from high redundancies and error rates. Conversely, the Swiss-Prot database has 85,824 total entries and is a curated database, which means that every entry has been carefully evaluated by a group of scientists prior to inclusion. Curated databases will have lower redundancy and error rates, yet due to their size may also be unable to identify a protein if it is not included in the database (12). The GenBank and Swiss-Prot databases use amino acid sequences and peptide molecular masses to identify a protein; yet, SEQUEST is a more complex database, which makes use of uninterrupted product ion spectra for the search query. The SEQUEST program uses an algorithm to correlate the product ion spectra with a theoretical spectra generated in the database which it thinks matches. In this process a cross correlation value known as Xcorr will be generated which ranks the matches found within the database. The higher the number, the closer the match, and the better the fit between the two spectra. SEQUEST therefore is a more complex data sequence program, since it uses both the mass and intensity data (12).

Currently, the mass spectrometer is the core component to virtually all proteomic methodologies and high throughput biomarker discovery laboratories. The era of large-scale clinical proteomic analysis essentially began in 2000 with the use of SELDI-TOF/MS, a modification of MALDI approaches. SELDI (surface enhanced laser desorption ionization) made use of aluminum chips with eight spots containing different surface chemistries/affinities as a pre-fractionation step prior to the mass spectrometry analysis the entire process is illustrated in Figure 2 (25). The multiple affinity surfaces on these Protein Chips allowed for targeted capture of proteins within complex mixtures (25). With this technology, serum samples could be assessed

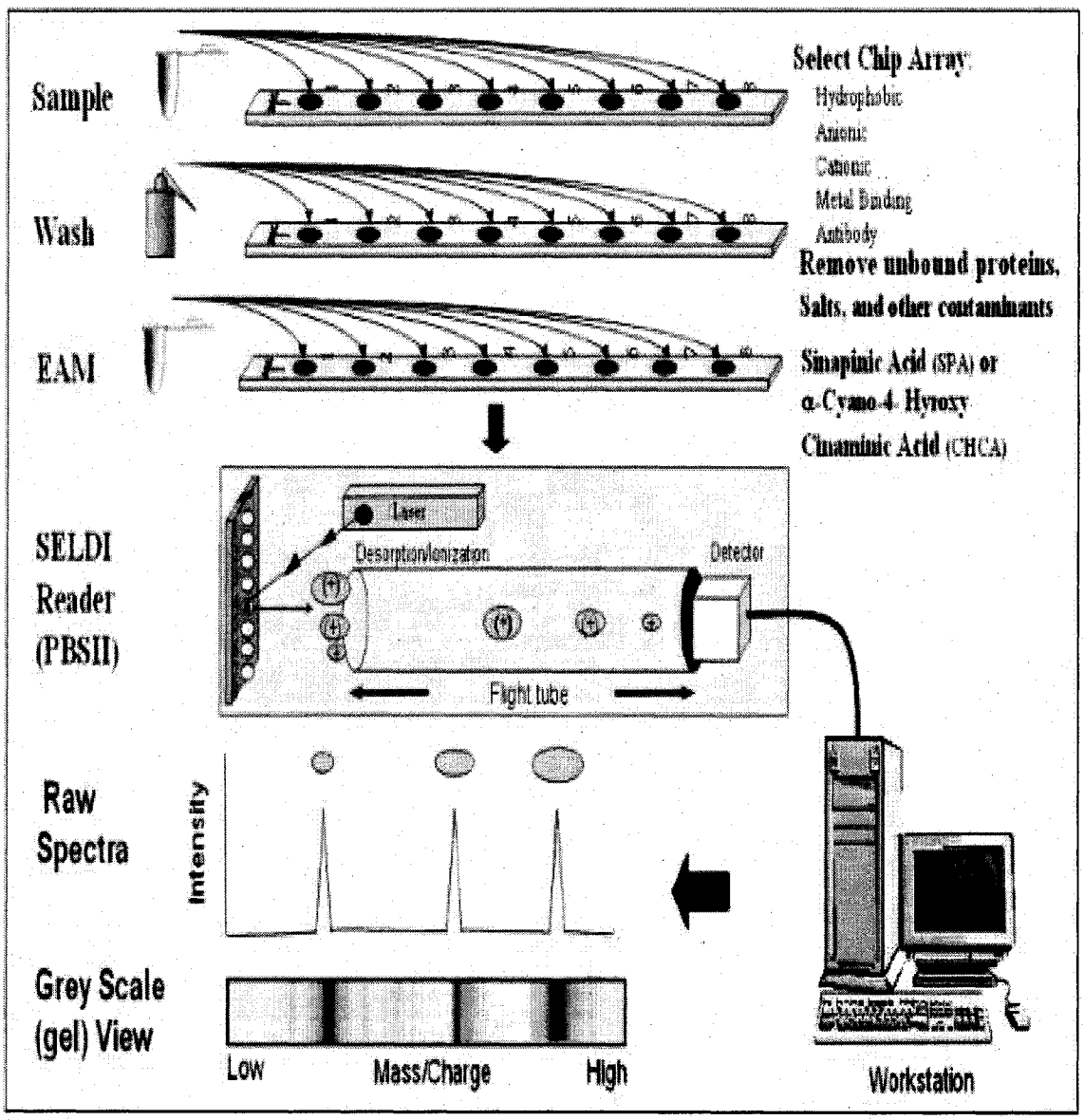


Figure 2. SELDI-TOF MS Technology. Serum samples are first incubated with a ProteinChip having a unique surface chemistry. Following A series of washes the unbound proteins would be removed, along with other contaminants. EAM matrix would be spotted and the samples read by SELDI PBS II instrument.

for patterns or peaks specific to disease, as shown by mass to charge ratios, and sample sets could be stratified using different algorithm-based data analysis programs. One large disadvantage of this approach is the inability to routinely determine the identity of the differentially expressed peaks observed in the sample sets. The instrument does not allow for direct fragmentation of the peptide of interest for database searching, which must be identified using other techniques (25, 26).

Other complications with the SELDI technology, in particular the lack of reproducibility of results obtained from different groups and institutions, surfaced around the time it was gaining the most popularity as a biomarker discovery tool (9, 27, 28). Since this time, validation studies have helped to alleviate some of these problems, which now seem to be more attributable to the study design and individual research institutions, than actual defects with the instrumentation (9, 27, 28). Yet better proteomic technologies are continually being developed, and in the time it has taken to validate many of the early SELDI studies, newer mass spectrometry platforms have virtually replaced the early SELDI instruments (9). Also contributing to the decrease in SELDI use has been the ongoing concerns regarding the mining of the dynamic range of the proteome. While SELDI seems to be very efficient from a global profiling perspective, it does not allow analysis of the less abundant proteins in clinical specimens. For these reasons there has been a recent transition into a more targeted, specific approach to biomarker discovery. One instrument with the capability to aid this transition, which was designed for large-scale clinical biomarker studies, is the Bruker Daltonics MALDI-TOF/TOF instrument with ClinProt analysis software. With better instruments that have higher detection sensitivities, the field of cancer-specific proteomic biomarker discovery has emerged as an area with the promise of improving the early detection of cancers, by identifying disease specific proteins within clinical specimens (9). Putative cancer biomarkers could be tumor-associated proteins or host-response proteins to the tumor, which could then be incorporated into a clinical assay, such as a blood test for cancer detection or recurrence (9).

1.3 Overview of Biomarker Discovery

Over the past ten years, proteomic profiling and biomarker discovery as applied to cancer research has produced many potential biomarkers for various cancers including prostate, pancreatic, nasopharyngeal and ovarian (29). In addition, the future implications of this research are tremendous, and biomarkers could serve as the bridge to personalized medicine (5). The roots for this type of research can be traced to the mid 1960's with the discovery of carcinoembryonic antigen (CEA), a protein thought to only be in fetal tissues, surprisingly found elevated in the blood of patients with colon cancer (5). In the 1970's and 1980's, discovery of other biomarkers continued including CA 19-9 for colorectal cancer and pancreatic cancer, CA15-3 for breast cancer and CA-125 for ovarian cancer (5). Although elevations of the levels of these biomarkers in serum can be used to diagnose an individual with cancer, they are not necessarily specific for the type of cancer. Currently, there are several early detection tests in use within the clinical setting. Two of these include prostate specific antigen (PSA) for prostate cancer and alpha-fetoprotein (AFP) for the occurrence of Hepatocellular Carcinoma (HCC) (26, 30). Although clinical use of both markers has led to success in decreasing the overall death associated with disease, they are still limited when used alone (26, 30). For example, AFP elevation is thought to correlate with HCC occurrence, yet AFP concentration is not a consistent indicator of disease, as it is known that it can be elevated in non-cancer related cases and in patients without serious disease. Furthermore, whether AFP can be used to detect the earliest cases of HCV derived HCC remains unclear (30). In relation to this thesis, an overview of the use of biomarkers like PSA for prostate cancers and AFP for HCC will be presented, including a discussion of their strengths and limitations for clinical use in early detection and prognostic assays.

1.4 Hepatocellular Carcinoma Incidence and Risk Factors

According to the American Cancer Society, over 18,510 new cases of liver cancer, or hepatocellular carcinoma (HCC), will be diagnosed in 2006 in the United States, accounting for over 16,200 deaths associated with disease. Furthermore, due to the long latency associated with cancer development, it is suggested that the number of new HCC cases will steadily increase, producing an anticipated 3-4 million new cases each year over the next 20 years (31-

35). Although the specific mechanisms leading to liver cancer remain undefined, there are several known risk factors associated with disease development including chronic infection either with hepatitis B (HBV) or hepatitis C (HCV) viruses, and incidence of chronic alcohol abuse. Individuals with either HBV or HCV infections account for greater then 80% of all HCC cases worldwide (36, 37). All of these risk factors produce an inflammation/cirrhosis of the liver, the key symptom associated with HCC development (33). Because the research project of this thesis involves a serum cohort consisting of HCV infected patients, HCV pathogenesis will be discussed in more detail in the following section.

1.5 Hepatitis C Virus (HCV) Pathogenesis

Over 170 million people currently suffer from chronic HCV infections, which along with HBV infection, is the main risk factor for cirrhosis and liver cancer (36, 38). HCV is a single stranded, enveloped RNA virus from the Flaviviridae family, originally isolated from the serum of a person with non-A, non-B hepatitis in 1989 (36). The virus is transmitted by contact with contaminated blood, and high risk behaviors include blood transfusions prior to 1992, intravenous drug use, high risk sexual activity, hemodialysis, occupational exposure, birth to an infected mother and intranasal cocaine use (36, 38). Currently in the U.S., the largest populations with hepatitis C are those individuals involved with IV drug use and/or high-risk sexual behaviors (36).

Information on the HCV viral life cycle has been impeded by the lack of a replication competent in vitro cell culture model; therefore most of what is known has been elucidated in related cell model systems (39). From these models, it is thought that HCV replicates similar to most positive strand RNA viruses in the following steps: viral attachment and entry, translation of viral RNA, and release of viral particles. HCV is thought to gain access to its host target cell, the hepatocyte, by two glycoproteins within the HCV envelope, E1 and E2 (40). Currently there are several cellular receptors proposed to be involved in this process, including the CD81 receptor, the scavenger receptor class B type I (SR-BI) receptor and the low the low density lipoprotein (LDL) receptor, all of which are associated with lipid and lipoprotein metabolism (39).

Once HCV gains entry to the hepatocyte, the virus will begin to replicate in the cytoplasm. This replication is thought to occur at a rapid rate, although levels of detectable viral proteins and

RNA in the infected tissue will remain low for some time (36, 40). Upon completion of this viral replication cycle, the patient is said to have an acute HCV infection. Interestingly, the majority of acute HCV infections will be asymptomatic, and 20-30% of adults will experience mild symptoms including malaise, weakness or anorexia and jaundice (36). Elevation of serum alanine aminotransferase levels and the detectability of HCV RNA in a patient's serum is a good indicator that they have an acute HCV infection (36). The antibody to HCV will not become positive until the patient presents with symptoms, which usually occurs 1 to 3 months following exposure (36). Although it is possible to clear the virus at this stage of acute infection, this occurs in only about 20% of the cases, while the other 80% of patients cannot clear the virus and thus develop chronic hepatitis (36). Furthermore, this viral clearance is difficult to study, since, as discussed, most acute cases are largely asymptomatic (41).

Chronic HCV infection is defined as the persistence of HCV RNA in the blood for at least 6 months following acute infection, and could be defined as a state of viremia (41). The development of chronic HCV is variable, and can take as long as 20 years to develop. The variability of this stage is dependent upon many different factors (36). Contributing factors include speed of virus production and cell-to-cell spread, and presence or absence of a T cell immune response to HCV antigens. Other factors not immediately related to the virus, but rather the patients, are also important and include amount of alcohol consumption, age of initial infection, degree of inflammation of liver, and the presence of HBV or HIV co-infection (36). Roughly 10-15% of patients with chronic HCV will develop cirrhosis, which in most cases leads to liver cancer. Diagnosis of HCC is performed by liver biopsy, which measures the number of mononuclear inflammatory cells present in and around the portal area of the liver, and the number of necrotic hepatocytes. Fibrosis of the liver suggests possible progression to cirrhosis, and more advanced stages of liver disease will have "bridging fibrosis" as opposed to more benign conditions of confined fibrosis (36).

1.6 Proteomic profiling of Serum for the Detection of HCC progression

HCC has a poor survival rate, especially when diagnosed at advanced stages where effective therapies are lacking. Therefore surveillance of patients at highest risk for developing

HCC (patients with cirrhosis) has been an important strategy to decrease the cancer related mortality rate (31, 34, 35). Abdominal ultrasound is currently the most common imaging modality used for surveillance of HCC, and while it has been successful in increasing the survival of patients with cirrhosis by allowing earlier disease management strategies, it still has limitations. For example, often times tumors within the liver can remain undetected until they reach a size of 3cm, which normally occurs only in the latest stages of disease. Further limitations of ultrasound include the subjectivity of the operator or clinician, which can differ from place to place (42, 43).

Alpha-fetoprotein, or AFP, is the biomarker currently used for the detection of hepatocelluar carcinoma (44). In normal circumstances, AFP is a glycoprotein produced by the liver of human fetuses beginning when they are around five weeks old. AFP levels can increase to levels of 4mg/ml at 12 to 16 weeks of gestation, and after birth, levels progressively decline to concentrations of less then 10 ng/ml, where they usually remain for the rest of an individual's life (33, 44). Because increases in AFP levels can signify liver cancer in an adult, an AFP concentration exceeding 20 ng/ml is used to diagnose an individual as having HCC (45). Between 30-60% of patients diagnosed with HCC will have AFP levels greater then 50ng/ml, and thus can readily be detected with an AFP ELISA test. However, 40% of cases will be missed, highlighting the problems of AFP as the sole early detection diagnostic marker for HCC. For this purpose, AFP tests suffer from low sensitivity and specificities, and AFP levels have been shown to fluctuate in different ethnic groups, as well as chronically infected individuals. Furthermore, AFP is not a good indicator of small tumors (45). Therefore, only a handful of liver cancers are detected in the early stages using AFP, thus new efforts have emerged to find additional serum markers.

One candidate, des-gamma carboyxprothomin (DCP), also known as prothrombin induced by vitamin K absence-II, is an abnormal prothrombin protein shown to be elevated in the serum of HCC patients. In some studies DCP has been shown to be better than AFP at differentiating HCC from nonmalignant chronic liver disease, and high levels of DCP in late stage tumor tissues have been linked as poor prognostic indicators for patients with HCC. Another candidate serum marker, GP73, is a novel type II Golgi membrane protein of unknown function

expressed in the hepatocytes of patients with adult giant cell hepatitis. Increased expression of GP73 in hepatocytes appears to be a hallmark feature of advanced liver disease(30).

Proteomic technologies are continuously being applied to identify biomarkers in serum indicative of advanced liver disease, specifically for viral hepatitis and HCC. Surface-enhanced laser desorption time of flight mass spectrometry is one recent technology within the proteomics area with the potential to identify novel disease biomarkers. Initial applications of this technology to clinical fluids showed great promise for the early detection of prostate, breast, head-neck, ovarian, pancreatic, bladder and liver cancers (46-52). The immediate goal of proteomic profiling is to identify tumor-associated proteins or host-response proteins to the tumor as biomarkers in the development of novel diagnostic cancer assays, such as a blood test for cancer detection or recurrence. Novel technologies such as proteomic analyses should allow the identification of new diagnostic markers as well as therapeutic and preventive targets (53).

1.7 Prostate Cancer Pathogenesis and Hormonal Regulation

The prostate is a small gland found only in men, which is located in front of the rectum and beneath the urinary bladder. The prostate is made up of secretory epithelial cells or glandular cells responsible for producing the fluid for semen (54). Most prostate cancers originate in these glandular cells, and are referred to as adenocarcinomas. Adenocarcinomas begin growing slowly in a confined manner within the prostate gland. When prostate cancers are detected in this confined state there is a high survival rate. In contrast, when prostate cancers undergo continued growth without detection, the tumor cells can invade the prostate gland eventually spreading to the tissues surrounding the prostate (54). The cancer will metastasize to the pelvic lymph nodes eventually reaching the bone and more distant tissues such as the brain, liver, lungs. When cancers metastasize to these distant organs the treatment options and survival rates decrease, thereby highlighting the necessity for a better understanding about the metastasis of prostate cancers (54). There are several potential oncogenic signaling pathways thought to be involved in the origination and spread of prostate cancers, which involve hormones and growth factors, cytokine pathways and neuropeptide signaling cascades. Detailed descriptions of each of these pathways are beyond the scope of this thesis. However as follows, an overview of the hormone-

regulated pathways will be presented since they are a critical part of current therapeutic strategies and could also benefit from better prognostic biomarkers (54).

The circulation of male hormones or androgens, including testosterone and dehydroepiandrosterone produced by the testis and adrenal glands, regulates the physiological activities of the prostate gland (54, 55). At birth and early adolescence androgen levels are high, yet over the lifetime of an average male, these hormones begin to fluctuate such that with increasing age, older males will possess higher estrogen levels than androgen. Interestingly, this transition has been shown to correlate with prostate cancer incidence (54, 55). Androgens regulate the prostate gland by maintaining a healthy prostatic epithelium. Therefore, the androgen receptor is a critical component for the signaling cascades of the prostate. 5-a-Dihydrotestosterone is responsible for binding to the androgen receptor, resulting in activation of many androgen dependent genes, including those involved in the growth of prostate tumor cells including: PSA, c-fos, Drg-1 and caveolin (54, 55). Highly metastatic prostate cancers can often transition to an androgen independent state. As the tumor cells adapt to androgen, they become more highly metastatic by being able to activate cancer genes in the absence of androgen. Concomitant with this drop in androgen dependence is an increase in estrogen dependence, since the levels of estrogens rise in relation to this decrease in androgen levels (54-56). Removal of androgens or blocking of androgen action are the main therapeutic options for treating advanced prostate cancer, however, the development of androgen independence ultimately limits the treatment effectiveness and frequently leads to terminal disease outcomes (56, 57).

Androgens are essential for the growth and maintenance of the healthy prostatic epithelium, a process largely regulated by the stroma cells of the prostate. In response to androgen exposure, the stroma cells maintain a healthy epithelium by the production of growth factors known as Andromeda's (58).

1.8 Early Detection of Prostate Cancer

According to the American Cancer Society, 27,350 men will die in the year 2006 from prostate cancer, which accounts for 10% of all cancer related deaths in men. Although the definitive pathways leading to prostate cancer occurrence remain unclear, there are numerous

risk factors known to increase the probability of cancer incidence. These include advanced age, poor diet, and a positive family history of prostatic disease (54). When prostate cancers are detected in their earliest stages, individuals are predicted to have a five-year survival rate of 100%, yet when prostate cancers are not found until the latter stages of disease, the five-year survival rate plummets to 34% (59). Clearly, early detection of prostate cancer is an important strategy for maintenance, treatment and overall life expectancy of men with disease (9, 59).Therefore, there is great interest in the clinic to use of biomarkers capable of detecting prostatic disease in its earliest form (9, 60). Since gaining FDA approval in 1994, prostate specific antigen (PSA) has become the most commonly used marker for the early detection of prostate cancers. PSA is a 33-kDa-serine protease produced by the prostatic epithelial cells secreted into seminal fluid to maintain liquefaction of the semen (55, 61). Within the seminal fluid, the PSA concentration can reach levels of 0.5-3 ng/ml (55, 61, 62). Any damage to the prostate, specifically to the prostatic basement membrane, can result in the release of PSA into the circulation thereby producing detectable levels of PSA in blood or serum (61). Elevation of serum PSA levels strongly suggest prostate cancer, yet high levels of PSA are also seen in benign conditions such as benign prostatic hyperplasia, which is a common occurrence in older men (61). PSA levels have greater success in determining the probability of clinical recurrence for carcinoma patients whom have recently undergone radiotherapy (59).

Although PSA is the "gold standard" diagnostic marker in clinical use for the early detection of prostate cancers, it does have some limitations (9, 26). For example, elevation of serum PSA levels greater than 4ng/mL and up to 10ng/mL are indicative of prostate cancers, yet when using these values alone, a clinician will miss greater than 25% of prostate cancers (9, 63). Clearly this is of great concern for clinicians, since PSA values are critical for determining whether a biopsy is necessary. It also puts pressure on individual clinicians to use their own discretion to decide whether it is best to adhere strictly to the PSA level guidelines and potentially miss cases, or lower the PSA level thresholds to detect more cases, which could potentially result in a higher incidence of unnecessary biopsies (63). In truth, neither PSA threshold nor a single marker for cancer detection is perfect, yet possibly a combination of markers or a range of parameters for

early detection could help clinicians decide on the biopsy question. For these reasons there has been much interest in development of additional parameters to be utilized with PSA. For example, it is now widely accepted that PSA should be used along with DRE (digital rectal examination) to increase the early detection of prostate cancers (60, 64). Also, it is recommended that more patient criteria should be taken into consideration including assessment of patient age, as well as other specific characteristics of PSA including: PSA velocity, PSA density and the proportion of free to total PSA ratio (60, 63, 64).

Assessment of patient age is a critical strategy for the maintenance and early detection of prostatic disease since increasing age is a key risk factor for development of prostate cancers. It is therefore recommended that men the age of 50 years or older with an anticipated lifespan of 10 or more years, as well as men 40-50 years old with a family history of prostate cancer receive testing. These men would be labeled as high risk, and be recommended to receive a digital rectal exam as well as PSA test every year, and if one of these tests were abnormal they would be referred for a biopsy (64). As discussed earlier, 25% of men with localized prostate cancers will have PSA values outside of the measured ranges, and it is therefore important to individualize each case by additionally assessing patient age and family history.

Other correlates of PSA levels such as PSA velocity, age-adjusted PSA, free to total PSA and PSA density have also begun to be incorporated in this testing (64). PSA velocity is the change in PSA over time and is assessed using three measurements spanning 1 to 1.5 years. The acceptable range is .75ng/ml/year, and it is thought that this test may save up to 30% of unnecessary biopsies among men with elevated PSA and prior negative biopsies (64). The inherent advantage of using PSA velocity is its ability to take patient age into consideration, based on the assumption that PSA levels naturally increase with age. This therefore will raise the upper limit of normal for older men to reduce unneeded biopsies, while not compromising its ability to detect cancer cases. Even though it seems promising, currently PSA velocity is thought to miss up to 5-20% of cancers (64). Another parameter, PSA density, is calculated by dividing the serum PSA level ng/ml by the estimated prostate volume in cubic cm as calculated on transrectal ultrasound (TRUS). PSA density shows the most promise in decreasing the rate of

false biopsies of men with PSA values which would otherwise suggest cancer, in the 4-10ng/mL range (64).

The free to total PSA ratio is measured as the proportion of free-floating PSA in serum compared to the amount of PSA complexed with other proteins. In serum the majority of PSA is found complexed with alpha-1-antichymotrypsin, leaving only a small amount of PSA in a free-floating state. This ratio has shown great promise in discerning men with benign prostatic diseases from those with cancer since the amount of free PSA is generally lower in men with prostate cancer (64). Unfortunately, like total PSA testing, this ratio is affected by factors other than the presence of cancer. For example free PSA levels correlate with large prostate sizes and increasing age. Therefore, many of the same problems that affect total PSA testing also impact this ratio.

1.9 Immune Response to Cancer

Many tumors have the ability to mimic the normal regulatory and signaling functions associated with the immune system, properties that generally serve to mask detection of tumors by the immune system (65). The immune system is a complex network of regulated effector cells that work in a collaborative and synergistic manner to defend the body from invading pathogens or foreign agents (65). Cumulative mutations in DNA during tumorigenesis affect the signaling pathways that regulate the processes of tissue homeostasis, cell survival and cell death, ultimately leading to un-regulated growth of the tumor (65). The immune response to the tumorigenesis process is equally complex and continues to be a critical area of cancer biology research (65). In the following paragraphs, a brief summary of the function of the immune system is provided first, followed by an overview of the immune response to cancer and the problems associated with this process that can lead to tumor evasion. Finally, the strategies used in current immunotherapies which seek to accentuate the response of the immune system toward detection and elimination of tumor cells are described.

The immune system is composed of two distinct branches termed innate immunity and adaptive immunity. Both branches have specific roles in the process of antigen/pathogen detection and elimination (65). The innate immune response is the body's first line of defense to a

pathogen, requiring no prior memory of infection, and occurring at the moment the body first recognizes a foreign non-self invader (65). Pathogens are recognized by macrophages and natural killer (NK) cells which initially rush to the site of infection to destroy the foreign agent. Following this recognition, the macrophages and mast cells will release cytokines, chemokines and matrix remodeling proteases, to recruit leukocytes into the damaged tissues to assist in the destruction and containment of the pathogen or foreign substance (66). If the innate immune response fails to completely contain and destroy the pathogen, it will be necessary for the adaptive immune response to mount a more specific and exhaustive attack of the pathogen (66).

Activation of the adaptive immune response occurs via the antigen presenting dendritic cells, which take up foreign antigen and migrate to lymphoid organs where they present the antigen to naive B and T cells (65). Upon activation by the antigen presenting cells, the B and T cells will be committed to this specific antigen and undergo clonal expansion, producing more B and T cells with the same recognition capability (65, 66). In addition, the adaptive response will generate memory B cells to be present in case the same antigen is found again. In this way under normal circumstances the different components of the immune system provide a unified defense against invading pathogens (65, 66). However, in the context of cancer, the immune response can be compromised primarily due to tumor's ability to mimic signaling components of the immune system. The tumor alters signaling cascades to create a more suitable environment for tumor growth (65, 66).

Tumors are largely heterogeneous, and often composed of fibroblasts, epithelial cells, innate and adaptive immune cells, cells that form blood and lymphatic vasculature as well as specific mesenchymal cell types (65). Tumor heterogeneity is advantageous to the cancer cell and provides the essential components for remodeling the tumor environment into one suitable for tumor propagation (65). From its initial site of disease, cancer cells can spread to distant sites of the body causing organ dysfunction eventually leading to death of the cancer patient. When tissue homeostasis is disrupted, interactions between innate and adaptive immune cells may also be altered, which can lead to the cancers ability to escape immune detection (65). How a tumor can escape immune detection is still being heavily researched. It was originally thought that

tumors did not produce sufficient levels of detectable antigens, such that the APCs (antigen presenting cells) did not have enough stimulatory signal to secrete the appropriate cytokines to mount an efficient immune response (65). Current research has focused more on the tumor cells themselves in regards to their ability to alter their microenvironment to reject the immune response, and overcome immunosurveillance (65).

Tumor immunosurveillance refers to the hypothesis that the immune system is capable of protecting its host against the development of cancers of non-viral origin and was first hypothesized in 1957 by Burnet and Thomas (67). This initial hypothesis was widely criticized because there were few examples of the immune system overcoming cancer. However, cumulative studies conducted between the mid 1970's and 1990's have provided experimental evidence to support the surveillance hypothesis (67, 68). Currently it is accepted that the immune system does attempt to constrain tumor growth, yet tumor cells may escape or attenuate this immune function (67). The duality of the immune system for both host protection and tumor progression has re-emphasized the immunosurveillance hypothesis, which has evolved into the cancer-immunoediting hypothesis (67). The immunoediting process consists of a series of three sequential phases, elimination (protection), equilibrium (persistence) and progression which allows a tumor to escape surveillance (65, 67).

In the elimination phase the transformed cells are recognized by the innate and adaptive immune system, which immediately attempt to eliminate the transformed cells (67). This achieved first by natural killer (NK) cells of the innate immune response. Following activation of the adaptive immune response from antigen presenting cells, $CD4^{+}T$ and $CD8^{+}T$ cells can also recognize tumor specific or tumor associated antigens via their appropriate MHC class II and I molecules and destroy them. In addition B cells produce antibodies that recognize antigens at tumor cell surfaces (67). Tumor cells, not eliminated in this phase, will continue to grow and proliferate into the equilibrium phase. In this phase, further expansion of these persisting tumor cells can still be prevented by the impending adaptive immune response. Failure to contain and eliminate the tumor cells in the equilibrium phase allows the cancer to persist and enter the escape/progression phase (67).

The escape phase can occur when the balance between the immune response and the tumor shifts in favor of tumor growth (67). Tumors can escape immune detection by creating an environment that does not allow leukocyte extravasation or local recruitment of immune cells, which is the beginning step of inflammation. In addition proliferating tumors can secrete cytokines, such as IL-10, which is an anti-inflammatory cytokine capable of inhibiting synthesis of the pro-inflammatory cytokines such as IFN γ and TNFα (69). Tumors may also secrete colony stimulating factor, vascular-endothelial growth factor and IL-6, which promote the accumulation of a heterogeneous population of immature myeloid cells as a result of continuous JAK2-STAT3 signaling. These myeloid cells possess specific immune suppressive properties. For example, these cells can suppress T-cell immunity by depletion of arginine and elaboration of reactive oxygen species (ROS) and nitric oxide (NO) (69). In addition to all of these potential mechanisms of immune evasion, the tumor microenvironment also supports the formation of regulatory T cells, which specifically suppress T cell function (69). An additional feature that assists immune escape is the overall genetic instability of a tumor, which increases with tumor progression. The end result is a tumor composed of cells that are genetically and epigenetically different from one another and very unstable (69, 70). This instability allows tumors to escape immune detection by down regulating the tumor antigens or the antigen processing machinery (69).

In addition to the ability of the tumor to directly influence the inhibition of immune detection, intracellular communication occurring between cancer cells and stroma cells can also contribute to this process (71). The stroma is composed of immune, inflammatory, endothelial, and fibromuscular cells and contains numerous growth factors and their receptors (69). Manipulation of growth factors and their receptors could have additive affects resulting in feedback loops which could propagate aggressive cancer growth (71). In addition, there are also many soluble factors that can be secreted by the stromal cells including chemokines and cytokines, which could negatively influence the immune response (71). Stroma derived soluble factors can also play a role in tumor invasion. For example during tumor invasion the ECM or extracellular matrix barriers are compromised, and most of the enzymes responsible for this process are not derived from tumor cells but rather the stroma cells (72).

Clearly the ability of a tumor to escape immune detection presents a large obstacle for treatment and survival of cancer patients. Therefore development of immunotherapies capable of treating the disease and assisting the immune response are being assessed (68). The underlying goal of immunotherapy is to enhance the immune response to cancer, to prevent tumor from escaping immune detection and allow for clearance of the disease. There are two primary tumor immunotherapy approaches currently in use. The first approach is based on the hypothesis that immune response elements are missing in the tumor microenvironment of patients with tumors (68). Therefore, the immunotherapy strategy is to supply or supplement the missing components. These treatments include injection of cytotoxic T lymphocytes, DC vaccinations, injection of activated natural killer cells, and administration of effector cytokines. The second branch of tumor immunotherapy is based on the hypothesis that immune tolerance is dominant in tumor patients. Therefore, reversing or blocking the suppressive mechanisms to the immune response is the goal of these therapies, specifically blocking the activities of regulatory T cells or dysfunctional dendritic cells and suppressive cytokines (68). Another attractive therapeutic approach is in targeting the normal stroma cells. Since the stroma cells surrounding the tumor stroma are diploid, genetically stable they have limited proliferative activity targeting the stroma is thought to decrease the incidence of immune evasion (73). It is thought that targeting therapies to recognize tumor stroma components might be a promising immunotherapy approach and one which needs further research (73).

It has been suggested that proteomics could offer an alternative approach to monitoring the immune response to immunotherapy, which could complement existing immunological assays (74). In this way biomarkers that aid in the early detection of cancers may also provide information about the quality of the immune response occurring within the cancer patient. Similarly, newly identified biomarkers are useful for early detection cancer biomarkers could be immune system components, as well as aid in choosing the most appropriate therapies for a particular individual (74).

1.10 Altered Glycoproteins and Cancer

Glycosylation is one of the most common post translational modifications of secreted proteins like those found in blood or on the membrane proteins of the extracellular surface (6, 75-78). Glycosylation events can be extremely diverse, and are known to influence many cellular processes including cell adhesion, signaling, stabilization of protein structure and protein trafficking (79-81). Alterations in glycoslylation patterns are strongly associated with oncogenesis and disease development (75-77, 82). Malignant associated glycosylation events include: increased glycan branching, greater lewis antigen expression and increased sialylation (80, 83). Therefore due to this longstanding relationship between aberrant glycosylation and oncogenesis, there has been great interest in determining whether glycoproteins within clinical samples could be used as tumor markers or biomarkers of disease (6, 77, 84-86). It is thought that either the glycoproteins themselves, as well as the differences in oligosaccharide moieties and/or enzymes responsible for the formation of the oligosaccharide structures attached to the glycoprotein of interest may be indicative of a specific disease state (6, 42, 79).

For example, altered activity of N-acetylglucosaminyltransferase V (Gnt-V), the enzyme responsible for the formation of branching asparagine-linked oligosaccharides, has been linked to tumor invasion and metastasis. Gnt-V activity, is reported to be increased in hepatocellular carcinomas, breast cancers and colon cancers (87-90). This has been supported by Handerson et al. who found an increase of ß1-6 branched oligosaccharides present on the surface of breast carcinoma cells within metastatic lymph nodes (88). When present, this was highly indicative of poor prognosis (88). The role of sialylated oligosaccharides in cancer has also been widely researched. In 2001, Dwek et al. evaluated the role of sialylated oligosaccharides within primary breast tumors, finding an overall reduction in the diversity of sialylated and neutral oligosaccharides with disease progression (81). In addition to assessing the overall glycosylation profiles of cancer and normal samples, it is also important to note that many of the current biomarkers in clinical use are well characterized secreted glycoproteins (91-94).

For example, multiple glycoforms of PSA have been described, and in prostate cancers, the glycoforms of PSA from seminal plasma have been shown to differ from those of PSA

secreted by the metastatic tumor cell line LNCaP (91-93). Other studies evaluating the difference between PSA glycoforms from seminal plasma and serum from healthy control and prostate cancer patients have also been performed (94), although these will need to be re-assessed using a larger set of clinically relevant samples to discern whether they could be potential markers. AFP is another glycoprotein useful for early detection. Like PSA, multiple glycoforms of AFP exist which again emphasizes the potential of using these PSA and AFP glycoforms for the diagnosis of disease state, as well as the implications of adapting these methodologies to other known serum components, (42, 95-97), characterization of these glycoforms represents a targeted proteomic strategy amenable to quantitative mass spectrometry strategies, which could be an effective future biomarker discovery strategy (98, 99).

1.11 Glycoproteomics and Lectins

Glycoproteomics is a newly emerging branch of proteomics that focuses on characterizing the carbohydrate residues added post-translationally to proteins (30). Structural elucidation of the oligosaccharides on mammalian glycoproteins has long relied upon the use of lectins, a class of proteins found in plants and animals known to bind specific sugar moieties (100). Unlike antigen-antibody binding affinities, the affinity constants for the binding of monosaccharides and oligosaccharides to most lectins are in the low micromolar range, but can be millimolar (101, 102). These affinities are still sufficient to allow use of lectins for effective chromatography separations based on oligosaccharide binding, as well as capture of a specific classes of glycoproteins (101). Recently, lectin affinity approaches have been utilized to study disease progression in a variety of cancers (82). However, few studies have applied lectin capture to characterize serum glycoforms indicative of disease. This is a relatively new area of research, called targeted glycoproteomics, which coupled with mass spectrometry is currently being used to identify serum glycoproteins specific to disease stage. This is an emerging strategy in glycoproteomic biomarker discovery studies since the glycoproteins that are selectively captured from the serum of a diseased patient may have altered expression of the glycoprotein, or possibly contains an altered oligosaccharide structure on the glycoprotein (42).

There have been many research groups utilizing numerous different strategies in an effort to define the cancer serum glycoproteome as summarized in Figure 3. Fortunately, lectin based fractionations are adaptable to many different approaches involving capture of total glycoproteins or glycopeptides upfront without additional strategies to deplete the most abundant serum/or plasma proteins such as albumin. There have also been studies focusing specifically on characterizing the glycan structures alone by the use of an enzyme, which specifically cleaves N-linked sugars (PNGase F) (103). Clearly there are many potential ways to achieve glycoprotein information and numerous ongoing research efforts exploring these possibilities. For example, Block et. al has recently found a panel of proteins that become hyperfucosylated with HCC progression (30, 37, 104). The increase is seen in the level of core α-1,6 linked fucosylation of AFP. This fucosylation event has also been seen in testicular cancers, and can be seen as a change from a common biantennary glycan to a 1,6 linked core fucosylated biantennary glycan (37). In addition, the analysis of the total serum of patients infected with HCC found additional proteins which were hyperfucosylated including α-1-acid glycoprotein, ceruloplasmin, alpha-2-macroglobulin, hemopexin, Apo-D, HBsAg, kininogen and GP73. Also, within the same patients, an overall decrease of fucosylation on haptoglobin was shown. Furthermore, when the fucosylated glycoforms of GP73 and hemopexin were used to stratify serum samples with HCC vs. healthy subjects, HBV carriers or HBV cirrhosis, the sensitivities and specificities were greater than 90% (37). Although larger validation studies will be needed to confirm these findings, these results highlight the potential of the glycoproteome as a gold mine for biomarker discovery for cancer detection.

Also, lectin panels can be used to increase the amount of glycoproteins selectively captured, as demonstrated by a recent study by Zhao et. al. This study looked specifically at sialic acid glycoproteins within a pancreatic serum cancer cohort by using a mixture of sialic acid binding lectins. Using this strategy, a sialylated plasma protease C1 inhibitor was found to be down regulated in serum from pancreatic cancer patients (105). This may be due to the glycosylation event on the protein as it relates to pancreatic cancer progression. In addition, a cancer specific glycosylation event on Asn-83 of alpha-1-antitrypsin was detected, which could

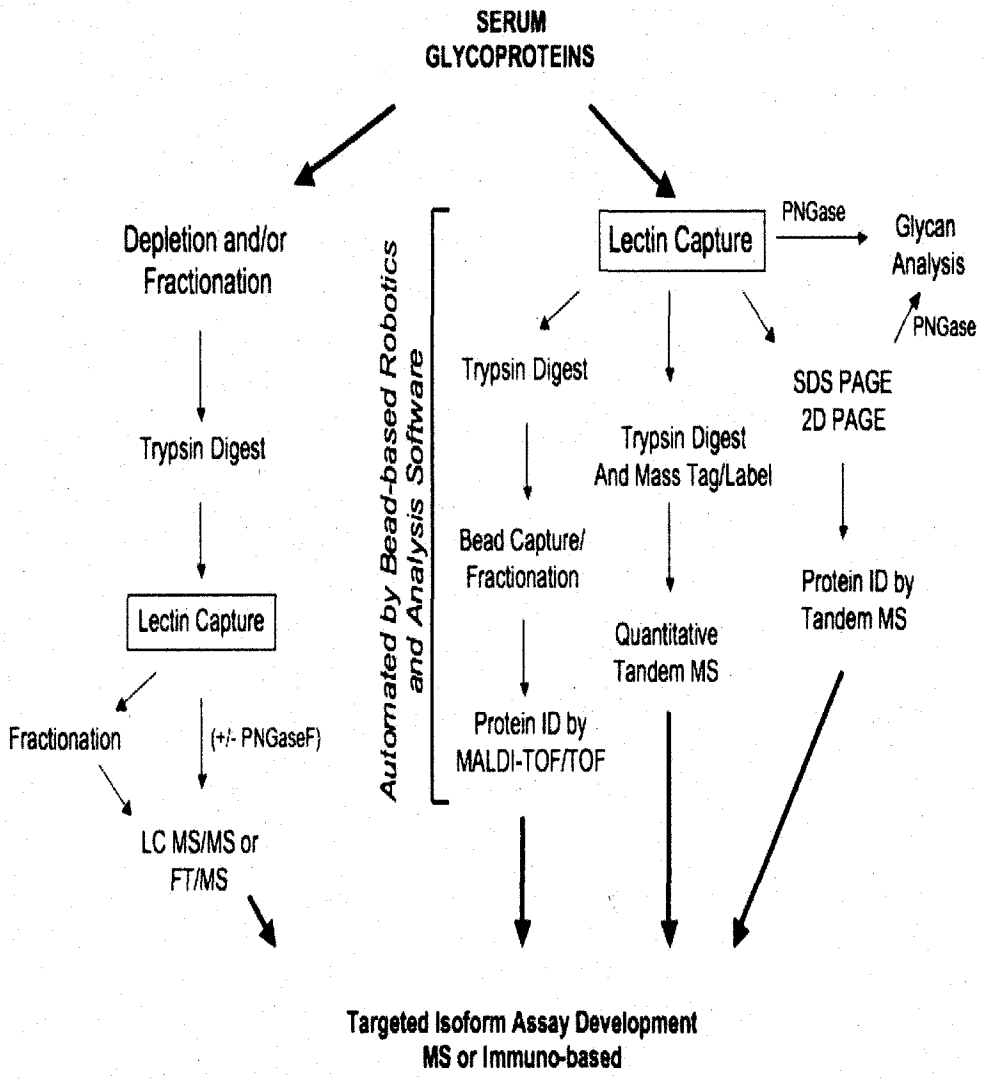


Figure 3. Diagram of the Different Strategies for Glycoproteomic Analysis.

also be a potential marker of pancreatic cancer (105). In addition to the aforementioned studies, lectin capture of serum glycoproteins following enzymatic digestion has been utilized to characterize more low abundance proteins within the proteome, thus allowing a greater dynamic range of the proteome to be assessed (24, 99, 106). The common strategy is to use trypsin to digest total serum proteins, isolate the glycopeptides with one or more lectins linked to a support resin, elute and deglycosylate the bound peptides with Protein N-Glycanase F (PNGaseF). The sequence and protein identities of these peptides are then determined by tandem mass spectrometry, (24, 99, 106), or Fourier transform ion-cyclotron resonance (FT-ICR) mass spectrometry (107-109). These strategies have been useful for identifying low concentration serum glycoproteins, though they are not amenable to high throughput clinical studies. Balancing the number of clinically relevant samples to be analyzed with depth of analysis is often the trade off in current proteomic profiling experiments.

1.12 Dynamic Range of Clinical Fluids

Proteomic profiling for disease biomarkers is widely performed in clinical fluids such as serum and plasma, since these fluids are presumably rich in proteins that leak from diseased tissues or are secreted from a tumor, and thus indicate disease. (110, 111) One downfall of using these protein rich fluids is they possess a wide dynamic concentration range. For example, the concentration of albumin, the most abundant plasma protein, is 10 orders of magnitude higher than the lowest abundant proteins within plasma (110, 111). In addition, the top 22 most abundant proteins in plasma account for 99% of the total protein content, making it difficult to analyze low concentration proteins that may be masked by the highly abundant ones (112). Also, many prospective biomarkers identified by SELDI and MALDI technologies are highly abundant known serum/plasma proteins, suggesting that when used with one-step affinity capture methods, only highly abundant proteins will be detected (105, 110, 111). Due to these concerns, there has been a concerted effort by the scientific community to resolve these issues, including development of more pre-fractionation approaches to reduce overall sample complexity prior to mass spectrometry, as well as the establishment of protocols to guide or standardize sample handling and collection (9, 110, 111, 113)) .

These strategies include depletion of the most abundant proteins, as well as pre-enrichment methodologies to capture specific subsets of proteins (112). Currently there are numerous different strategies for depletion of the most abundant proteins within serum and plasma samples. These techniques range from antibody based albumin and IgG removal to high-end, expensive immunoaffinity kits capable of removing the top 12 and top 20 most abundant serum/plasma proteins (114, 115). The depletions are performed prior to mass spectrometry analysis, which will theoretically allow visualization of lower abundant proteins. Because these abundant proteins function primarily as carrier proteins, there is a possibility that with these depletion approaches other proteins will also be lost non-specifically in the process (116, 117). This possibility was recently evaluated by Gong et. al, whom analyzed the bound fraction of albumin depleted from human plasma. From this research 57 unique proteins were identified in the albumin bound fraction. These proteins, which would have otherwise been discarded, could be of interest, since albumin functions as a transporter of many substances in the body including drugs, hormones and fatty acids (108, 116, 118). Furthermore, recent biomarker discovery groups have found many of the high abundant serum proteins, which are thought to mask low molecular weight proteins of the proteome to be differentially expressed in cancers. For example, alpha-1 acid glycoprotein, a known acute phase protein is thought to be down regulated in breast cancers, while levels of transthyretin, a thyroid binding protein is decreased in ovarian cancers. Also, haptoglobin has been shown to be over expressed in late stage ovarian cancers as well as cancers of the pancreas and lung (29, 38, 105, 119, 120). Another strategy to reduce the complexity of the proteome has been to perform a selective capture of a specific subclass of proteins, such as glycoproteins, to analyze the serum or plasma glycoproteome (29, 38, 42, 105, 113, 119, 120). Inherent in this approach is the high abundance of serum proteins thought to be glycosylated, in fact over 60% of this proteome is thought to be glycosylated (78). This has been achieved by the use of lectin capture, which allows selective isolation, and enrichment of the serum/plasma glycoproteins. This technique offers an upfront-targeted biological affinity approach, which not only captures glycoproteins, but lectins generally have minimal binding to abundant non-glycosylated serum proteins like albumin. Lectin capture as a first fractionation step

therefore has the potential to significantly decrease the levels of albumin in subsequent protein fractions comparable to levels obtained with high-cost depletion kits (113).

1.13 Dissertation Rationale and Specific Aims

The overall focus of this research project is on the use of proteomic profiling for the discovery of novel protein biomarkers indicative of the early detection or progression of cancers. Inherent in this approach is the assumption that these biomarkers will be cancer specific because either they were shed directly from a tumor cell, or they were derived from cells or other factors in the tumor microenvironment. Clearly this allows for a diverse range of protein classes with the potential to be a biomarker, including proteins with altered posttranslational modifications or truncated protein products (4, 9). **Therefore, the specific hypothesis of this dissertation is that the tumorigenic progression of liver and prostate cancers produces distinct disease-associated biomarker proteins derived from the tumor or surrounding environment that are detectable in serum or plasma.** It will also be determined whether protein patterns obtained during this analysis can be used to predict and monitor cancer occurrence and progression. Since newly discovered tumor associated proteins could be further assessed for the development of novel diagnostic cancer assays, such as a blood test for cancer detection or recurrence, the biological significance of this project is immense. The hypothesis was evaluated by the following specific aims.

Specific Aim I: Determine whether protein profiling of serum using SELDI-TOF-MS can distinguish patients with different stages of liver disease ranging from chronic hepatitis C to HCV-HCC. This will be assessed by:

A. A serum cohort with varying stages of liver disease will be profiled by SELDI-MS using copper-coated immobilized metal affinity capture (IMAC-Cu) protein chips (Ciphergen Biosystems, Fremont, CA) to assess whether differential protein patterns exist between the disease groups.

B. Data analysis will be performed on the spectra to determine whether there are differentially expressed protein peaks across the sample cohort. This will be accomplished by generation of p

values and peak lists across all sample groups for each detected peak for construction of classification algorithm regression trees (CART).

C. A serum sample cohort blinded to the investigator will be profiled identically to the initial test set run from A. Following the data analysis, the blinded samples will be classified using the decision trees derived in B and scored on their ability to be correctly classified.

D. The serum concentrations of other known HCC markers such as AFP or GP-73 will be compared and included with the SELDI data to determine whether these values improve or change classification of disease groups.

Specific Aim II: To use a targeted multi-lectin affinity approach to identify and differentiate glycoproteins present in serum from subjects with benign prostatic disease and those with prostate cancers. This will be determined by:

A. Lectin capture of serum glycoproteins will be performed with a panel of bead-based lectins with varying oligosaccharide-binding specificities. The bound glycoproteins will be eluted and separated by SDS-polyacrylamide gel analysis.

B. All glycoproteins of interest will be identified with peptide mass fingerprinting on a Bruker MALDI-TOF/TOF coupled with an appropriate protein database search engine.

C. Any putative glycoprotein markers identified will be validated using specific ELISA or other immunologic methods within the different prostate serum groups.

D. The structures of the oligosaccharide released from glycoproteins of interest will be assessed using MALDI-TOF analysis.

Specific Aim III: Determine whether use of lectins as an initial pre-fractionation step is a reproducible approach for profiling the entire dynamic range of the serum/plasma proteome.

A. Determine to what extent the amount of albumin within a serum and plasma sample is decreased using combinations of albumin depletion kits and lectin fractionation followed by SDS-polyacrylamide gel analysis.

B. Determine if lectin fractionation is reproducible and adaptable for automation by utilizing magnetic bead based lectins.

C. Compare lectin fractionation and depletion of albumin with commercially available kits as a pre-fractionation approach to decrease the sample complexity of serum and or plasma.

D. Determine whether albumin depletion prior to lectin fractionation allows capture of different glycoproteins when no albumin depletion step is done.

E. Use lectin fractionation followed by albumin depletion to determine whether this allows a greater number of proteins to be identified.

CHAPTER II

AIM I: TO DETERMINE WHETHER PROTEIN PROFILING OF SERUM USING SELDI-TOF-MS CAN DISTINGUISH PATIENTS WITH DIFFERENT STAGES OF LIVER DISEASE RANGING FROM CHRONIC HEPATITIS TO HCV- ASSOCIATED HCC

2.1 Introduction

Proteomic profiling of serum to identify new biomarkers indicative of disease severity and progression is an emerging technique important for the maintenance and early detection of cancers. Surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS), is a profiling technology introduced by Ciphergen Biosystems in 1999 specifically designed for high throughput global proteomic profiling of clinical samples such as serum and plasma (25). Initial SELDI-TOF experiments emphasized the enormous potential of this approach in the discovery of new diagnostic markers of cancer. The potential biological significance of this research is immense since new biomarkers could improve detection of cancer in its earliest form, as well as provide clues in the diagnosis of disease stage and existing tumor grade (9). To assess the utility of this approach and establish appropriate protocols, there were many early studies conducted (7, 25, 28, 121). One of these studies sought to determine whether SELDI-TOF MS profiling of serum could reproducibly detect prostate cancer patients from those with benign disease. This was assessed by comparing the protein peak selection and sample stratification at six different institutions, and showed that SELDI-TOF MS could reproducibly generate identical protein profiling spectra for quality control sera, and correctly distinguish healthy from prostate cancer subjects based on serum protein profiles (28). In addition, SELDI-TOF MS technologies have successfully identified biomarkers capable of distinguishing multiple cancers including: prostate, bladder, ovarian, lung, colon, breast and pancreatic (46-48, 50, 51, 122-128).

Using similar experimental protocols to those standardized in this multi-institutional study, and in an effort to build upon the earliest successes of this technology, (28). The objective of aim 1 reported herein was to determine if protein profiling of serum using SELDI-TOF-MS could accurately distinguish patients with different stages of liver disease, especially those associated

with hepatitis C virus (HCV) infections, ranging from chronic hepatitis to HCV-associated HCC. This study addresses a significant clinical problem since the current early detection marker of HCC; alpha-fetoprotein (AFP) suffers a poor sensitivity and specificity rate and can be expressed at normal levels in over one third of patients whom have small HCC tumors (129, 130). When HCC is not detected until it is in an advanced stage, the patient's five-year survival rate is roughly 50% following surgical resection (130, 131). Furthermore, it is possible that a panel of biomarkers may be a more attractive diagnostic assay than one single marker. Therefore, we also assessed whether inclusion of other known biomarkers of HCC in combination with the protein peaks found by SELDI-TOF MS could improve detection of disease. In this way it could be determined whether information from other known diagnostic markers could be combined with the SELDI peaks to improve HCC detection, or vice versa. Incorporation of known markers along with SELDI peaks had not been previously shown, and was a novel aspect of the approach outlined here.

These additional markers include: des-gamma carboxyprothrombin (DCP), or prothrombin induced by vitamin K absence-II, GP73 or type II Golgi membrane protein, and AFP. In previous studies, DCP was shown to be better then AFP in differentiating HCC from non-malignant chronic liver disease in serum obtained from patients in the United States. Also, high DCP values in serum or late-stage tumor tissues have been linked as poor prognostic indicators for patients with HCC (132, 133). Although GP73 function remains unknown, it is expressed in the hepatocytes of patients with adult giant-cell hepatitis, and it is over expressed in the hepatocytes of HCC patients. A recent study has shown that expression of GP73 can discriminate individuals with cirrhosis from those with HCC with a sensitivity of 69% and specificity 75% (134). Therefore, a comparison of the performance of these serum markers (AFP, DCP, GP73) singly, grouped, or in combination with the SELDI protein peaks will also be presented to determine if this approach increases sensitivity or specificity of detection.

2.2 Materials and Methods

Patient Specimens

All patients were enrolled from the liver and liver transplantation clinics at the University of Michigan Medical Center between September 2001 and May 2002 with Institutional Review Board (IRB) approval. Written informed consent was obtained from each patient. Four groups of consecutive subjects were enrolled. The first group included subjects with no history of liver disease and normal liver biochemistry, no risk factors for viral hepatitis, and alcohol consumption less than 40g/wk (135). The second group consisted of subjects with histological confirmed chronic hepatitis. The third group consisted of patients with histological proven cirrhosis and compensated liver disease (Child-Turcotte-Pugh score <7). A fourth group consisted of patients with histologically proven HCC (135). A 20-mL blood sample was drawn from each subject for AFP and DCP testing more then 2 weeks after liver biopsy was performed. Blood samples were spun and serum was aliquoted and stored at -80° until testing. Each sample used for proteomic profiling had not been thawed more that once. Blood samples from HCC subjects were drawn before initiation of treatment.

Serum Liver Biomarker Assay

AFP was tested using commercially available immunometric assays using enhanced chemiluminescence at the University of Michigan Hospital Clinical Diagnostic Laboratory. The upper limit of normal was 8ng/mL. DCP levels were measured using an enzyme-linked immunosorbent assay kit (Eitest PIVKA-II, Eisai Co., Tokyo, Japan) per the manufacturer's instructions and were performed in duplicate. Levels of GP73 in serum were determined by Western Blot analysis (136, 137). Equal volumes of patient sera (0.5 μL/lane) were separated via SDS-PAGE on 4% to 20% polyacrylamide gels. For normalization, each gel also included a lane containing 0.5 μL of serum from a pool of sources negative for HCV and hepatitis B virus (Sigma, St. Louis, MO). GP-73-specific signals from the 73-kd species were quantified from X ray film using AlphaInnotech FluorChem CCD camera with AlphaEase spot densitometry software (both from Alpha Innotech Corp., San Leandro, CA) and were expressed as integrated intensity units

relative to the GP73 signal detected in Sigma control serum standard. Values were calculated as the mean of duplicate or triplicate determinants for each serum sample.

SELDI Processing of Serum Samples

Aliquots of each serum specimen at the University of Michigan Medical Center were mixed in a 2:3 ratio of serum to 8 mol/L urea, 1% CHAPS, and were frozen at -80°C before being shipped to the Center for Biomedical Proteomics at Eastern Virginia Medical School. Serum samples were processed robotically on a Biomek 2000 liquid handling system (Beckman Coulter, Fullerton, CA) in a 96-well format for SELDI analysis in the following manner: A further dilution (1:5) of the thawed serum in the 8-mol/L urea buffer was made in 1mol/L urea, 0.125% CHAPS, and phosphate-buffered saline. Diluted serum was randomly spotted in duplicate onto copper-coated immobilized metal affinity capture (IMAC-Cu) protein chips (Ciphergen Biosystems, Fremont, CA) for SELDI-TOF analysis with the aid of a 96 –well bioprocessor. The sample was allowed to bind to the protein chips for 30 minutes at room temperature, followed by washes of phosphate-buffered saline and water. The IMAC-Cu chip arrays were allowed to air dry and a saturated solution of sinapinic acid in 50% (vol/vol) acetonitrile, 0.5% (vol/vol) trifluoroacetic acid was added to each spot. A detailed diagram of this process is included in Figure 4.

SELDI Data Analysis

The protein chip arrays were analyzed using the SELDI ProteinChip System (PBS-II; Ciphergen Biosystems). The spectra were generated by the accumulation of 192 shots at laser intensity 220 in a positive mode. The protein masses were calibrated externally using purified peptide standards. Spectra were analyzed with Ciphergen ProteinChip software (version 3.1) and normalized using total ion current. Peak clustering in the 1.5 to 20-kd range was performed using Biomarker Wizard Software ProteinChip software (version 3.1) and normalized using total ion current. Peak clustering in the 1.5 to 20-kd range was performed using Biomarker Wizard Software (Ciphergen Biosystems) at settings that provide a 5% minimum peak threshold, 0.2% mass window, and 2% to 3% signal/noise determination. Intensity values for each peak were then averaged for each duplicate sample pair analyzed and input into BioMarker Patterns software (Ciphergen Biosystems) for classification tree analysis as described previously (50, 51,

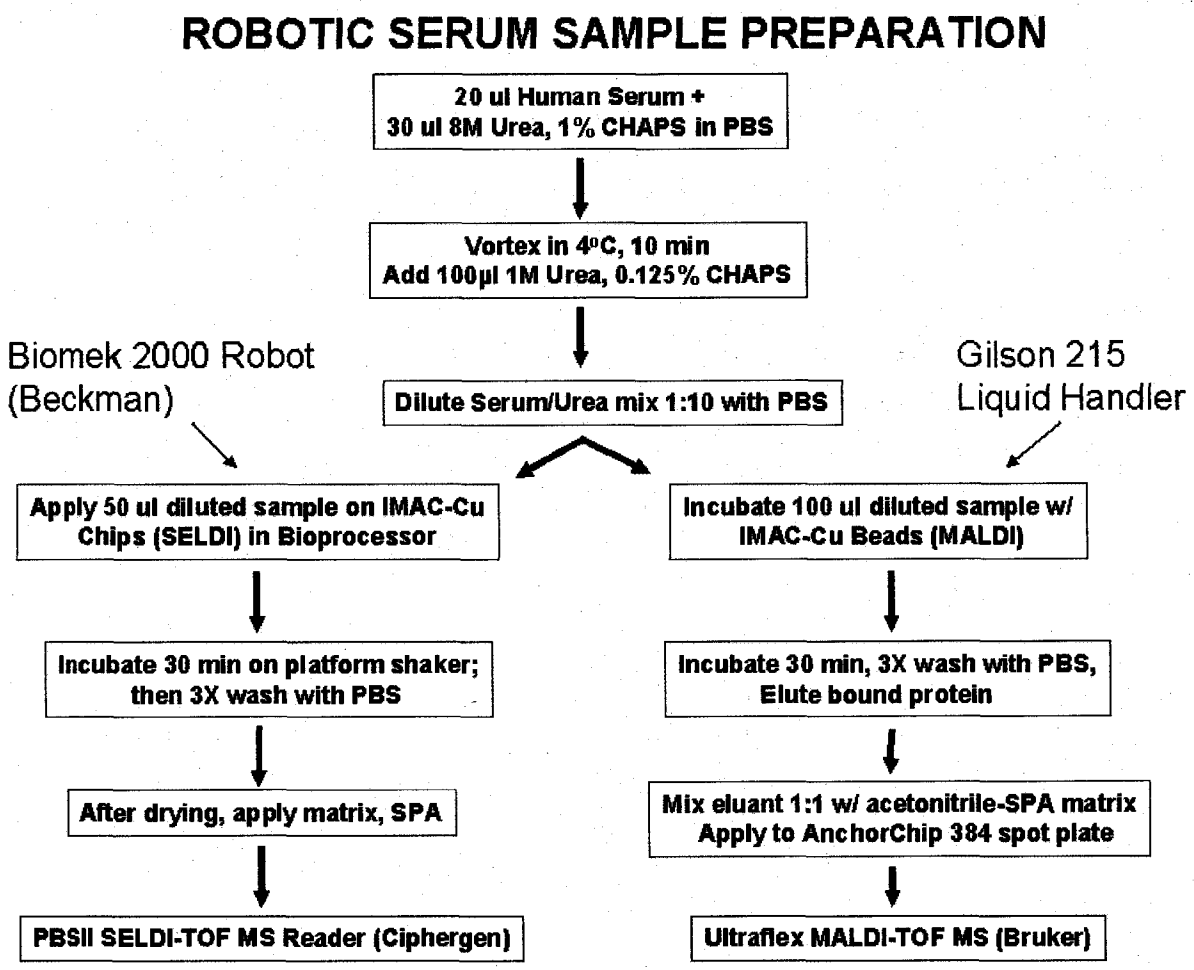


Figure 4. Robotic Serum Sample Preparation for SELDI-TOF MS.
20 µl of human serum is incubated with 8M urea and CHAPS and vortexed for 10 minutes. This is followed by addition of 1M urea and .125% CHAPS. This was further diluted and added to IMAC-Cu chips and spotted with SPA matrix.

138-140). Briefly, classification trees split the data into two nodes, using one rule at a time in the form of a question. The splitting decisions in this case were based on the normalized intensity levels of peaks from the SELDI protein expression profile. Each peak or cluster identified from the SELDI profile is therefore a variable in the classification process. The process of splitting is continued until terminal nodes are reached and further splitting has no gain in data classification. Multiple classification trees were generated using this process, and the best performing tree was chosen for testing. During the analysis, a pruning step occurs in which branches are removed and the cost of the removal is determined to establish a minimal tree size. This is referred to as the "learning set." Second, the decision tree was subjected to cross-validation. In this step, the data is portioned such that randomly selected samples are categorized with the decision tree being tested to ensure that the decision tree is valid. Only these cross-validated values are presented in the data tables herein. The nine SELDI peaks that formed the main splitters of the tree(s) with the highest prediction rates in the cross-validation analysis were selected for further analysis with the different serum markers.

Principle Component Analysis

Principle component analysis (PCA) was performed by CiphergenExpress Software. PCA analysis is a multivariate analysis method useful for reducing the dimensionality of the data. It is used in CE to visualize two and three dimensional graphs. The correlation matrix setting was utilized, which weighs each variable (sample intensity at an individual peak) equally. The graphs are plotted against three components. The CE software picks the components automatically based on the peaks detected in biomarker wizard, and clustered according to specified groups (normal vs. control).

Statistical Analysis

Specificity was calculated as the ratio of the number of negative samples correctly classified to the total number of true negative samples. Sensitivity was calculated as the ratio of the number of correctly classified diseased samples to the total number of diseased samples. Comparison of relative peak intensity levels between groups was calculated using the Student t test.

2.3 Results

Sample Processing and SELDI analysis

The SELDI-TOF approach was applied to serum from 170 patients as summarized in Table 1, which shows statistics on the samples used within the study including: patients with no liver disease (n =39), liver diseases not associated with cirrhosis (n = 36), cirrhosis (n = 38), or HCC (n = 57). Of these, a subset with only HCV-related disease was identified: HCV-HCC (n = 28), HCV cirrhosis (n = 22), chronic HCV (n = 27). Each serum sample was applied to copper-coated immobilized metal affinity chips (IMAC-Cu) in duplicate. All sample loading, processing and analysis steps were fully automated to minimize sample-processing errors. Following baseline subtraction and normalization using total ion current peaks present in all of the samples were labeled and clustered. The peak intensity values of 39 differentially expressed peaks identified in all samples in the 1.5 to 20-kd mass ranges were used for further analysis.

Comparison of SELDI spectra

Representative spectra in the 6,000 to 12,000 –Da range is shown in Figure 5. This particular spectra is from four individual samples as indicated by the sample numbers. This initial data was encouraging as it clearly showed the potential of detecting differential peaks across the four different samples as shown by the boxes drawn to illustrate the differences in intensities for a given peak (5,808, 8,939, 9,501, 11,735 m/z). To determine if these features were reproducible across the sample sets, multiple samples from each condition were assessed. The gel-view representation of five sample spectra from four groups (healthy, HCV-no cirrhosis, HCV-cirrhosis, and HCV-HCC) in the 5,000 to 12,000-Da range is presented in Figure 6. The four different distinguishing mass values for the given peaks are included. Again, in order to assess these patterns in the entire serum cohort, which was run, the intensities of the four peaks were assessed in all of the samples analyzed. This was achieved by plotting the peak intensity of each sample for the m/z of interest. Scatter plots of the intensities of these same four peaks in all of the HCV related samples analyzed are shown in Figures 7-10. Using the mean intensities of each sample as indicated by the bar, the proteins represented by the 5,808 and 11,735 m/z

Table 1. Demographic Information and Etiology of Liver Disease

	Group 1 (n = 39)	Group 2 (n = 36)	Group 3 (n = 38)	Group 4 (n = 57)
Sex (Male/Female)	21:18	20:16	22:16	35:22
Age (yr)	51 ± 11	50 ± 6	52 ± 8	54 ± 13
Ethnicity % NHW/AA/H/Asian	90/10/0/0	78/16/4/2	69/11/9/11	88/6/4/2
Etiology %				
HCV		75	58	49
HBV		10	11	4
Alcohol		0	2	13
Autoimmune		6	4	2
Cryptogenic		6	20	26
Other		3	5	6
MELD score	4 ± 0.7	5 ±0.8	7.2 ± 1.3	8 ± 2.3
ALT (IU/mL)	28.6 ± 9	67 ± 41	112 ± 124	81 ± 49
AST (IU/mL)	22 ± 5	53 ± 36	94 ± 85	109 ± 59
Bilirubin (mg/dL)	0.4 ± 0.2	0.5 ± 0.4	0.9 ± 0.6	1.2 ± 0.9
AFP (ng/mL)	2.94 ± 1.6	10.8 ± 23	19.7 ± 38	11,788 ± 60, 359
% < 20	100	88	77	55
% 20-200	0	12	23	24
% > 200	0	0	0	21
DCP (mAU/mL)	25 ± 4	31 ± 8	36 ± 12	1,925 ± 235
GP73 (mAU/mL)	5.1 ± 3.4	6.9 ± 5	10.3 ± 9	16.6 ± 8
TNM stage % (I/II/III/IV)	NA	NA	NA	16/22/21/13

Note: All data are presented as the mean ± SD.

Abbreviations: NS, non- significant, MELD, model for end-stage liver disease; NHW, non-hispanic white; AA, African American; H Hispanic; HBV, hepatitis B virus; ALT, alanine aminotransferase; AST, aspartate aminotransferase;
TNM, primary tumor/lymph node/distant metastasis; NA, not applicable

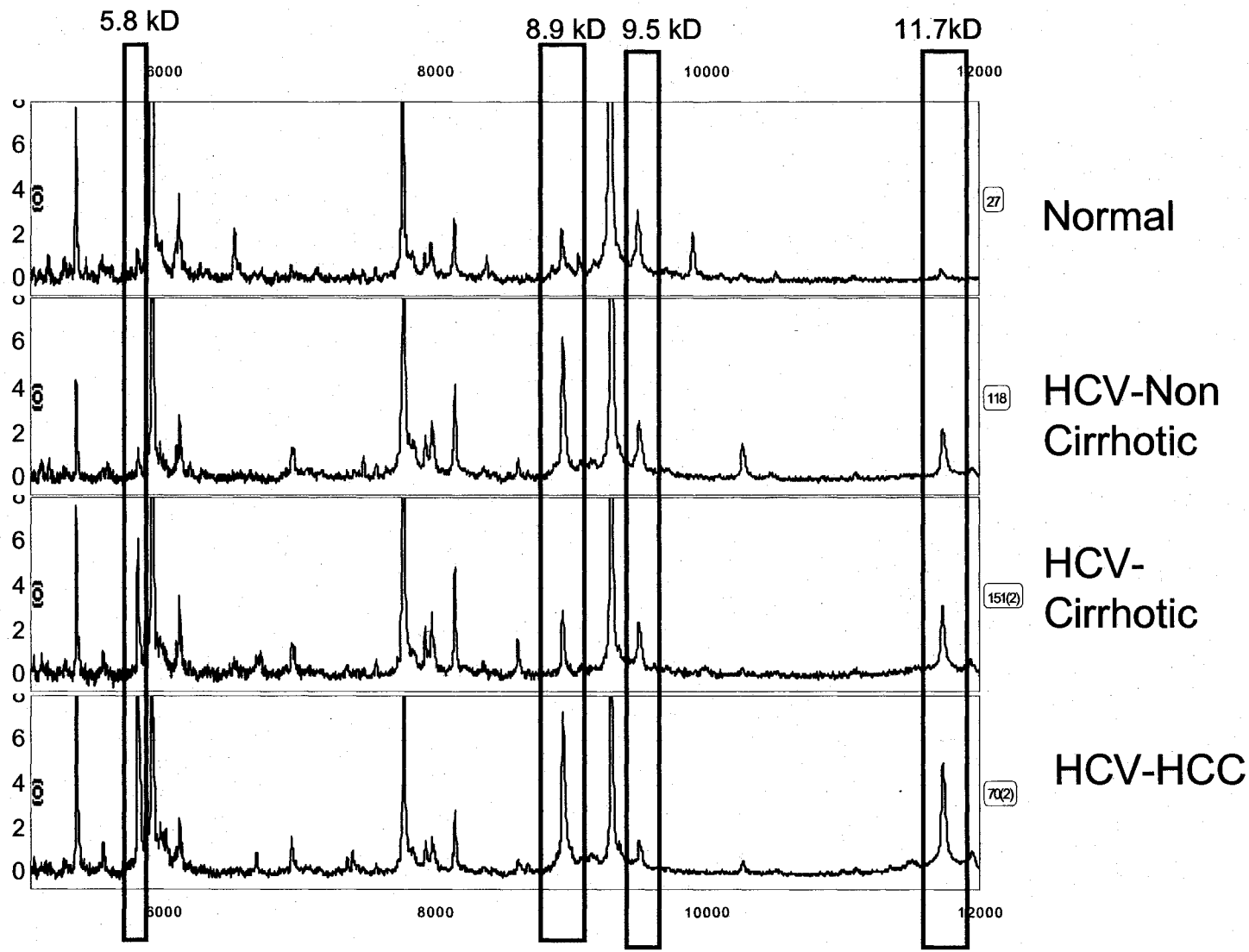


Figure 5. Representative SELDI spectra. Comparison of serum samples from healthy, HCV-non-cirrhosis, HCV-cirrhosis, and HCV-HCC subjects in the 5,000-12,000 m/z range. Each boxed region identifies differentially expressed peaks within the sample groups.

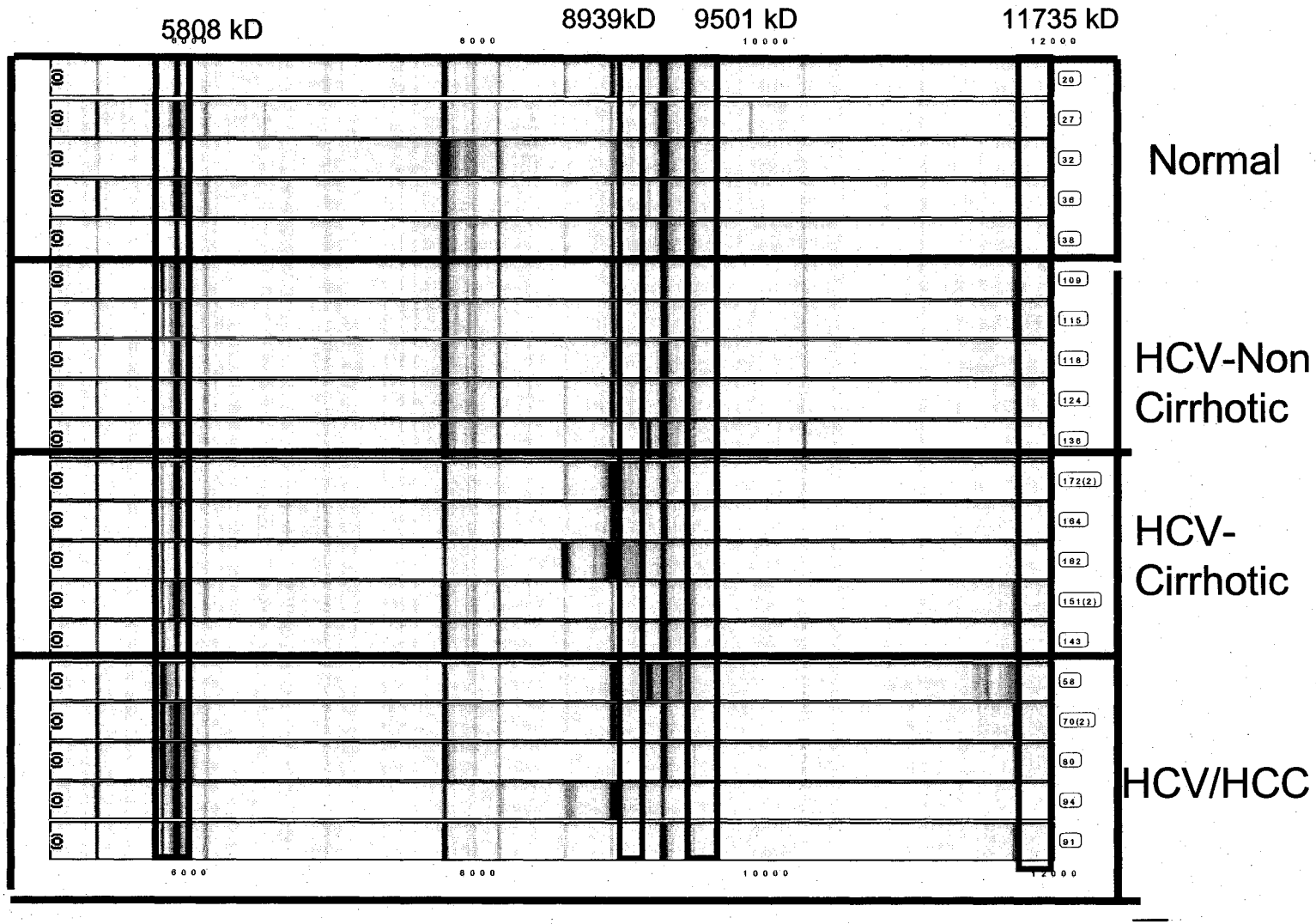


Figure 6. Representative SELDI gel view. Gel view comparison of serum samples from Healthy, HCV non-cirrhosis and HCV-HCC subjects (5 samples per condition) in the 5,000-12,000 m/z range. Each box represents a differentially expressed peak across the 4 groups.

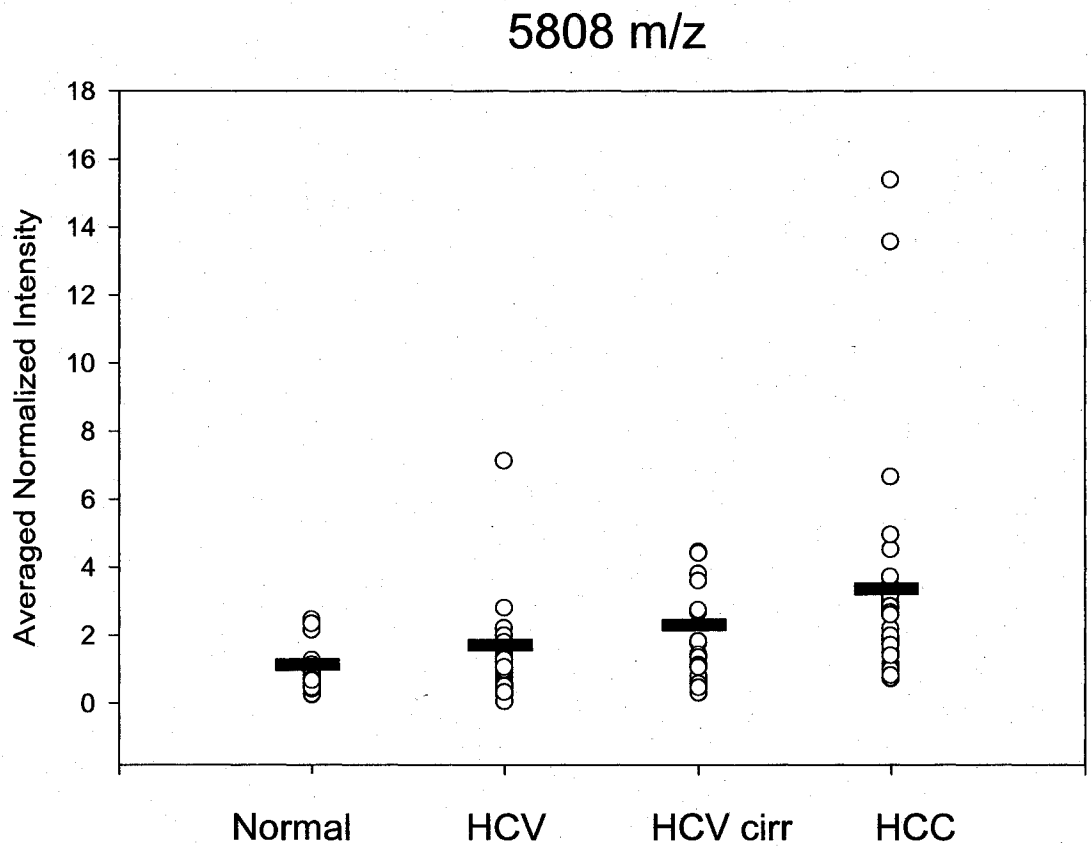


Figure 7. Expression levels of the 5,808 m/z protein. For each sample in all of The indicated HCV disease stratified data sets. Black bars indicate the mean normalized intensity;open circles represent values of individual samples. HCV, Hepatitis C virus infection but no cirrhosis; HCV cirr, HCV-associated cirrhosis; HCC, HCV associated hepatocellular carcinoma.

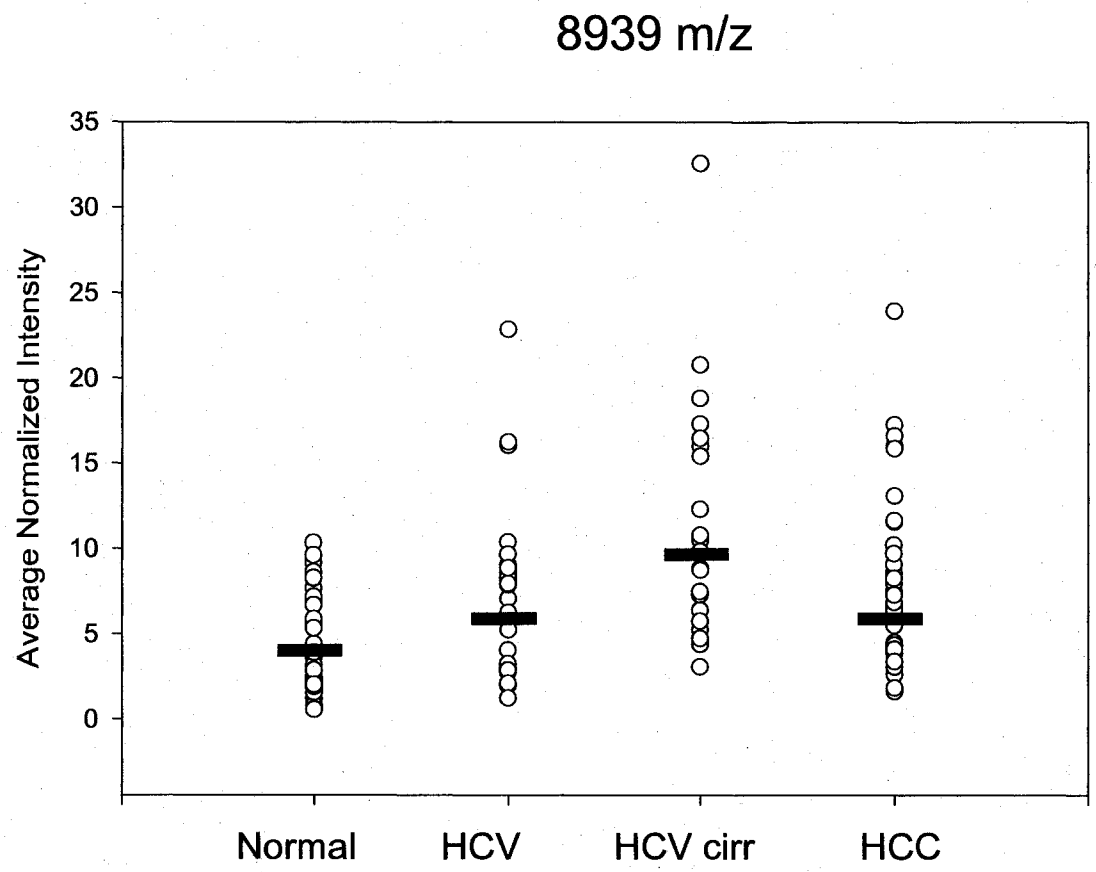


Figure 8. Expression levels of the 8,939 m/z protein. For each sample in all of The indicated HCV disease stratified data sets. Black bars indicate the mean normalized intensity;open circles represent values of individual samples. HCV, Hepatitis C virus infection but no cirrhosis; HCV cirr, HCV-associated cirrhosis; HCC, HCV associated hepatocellular carcinoma.

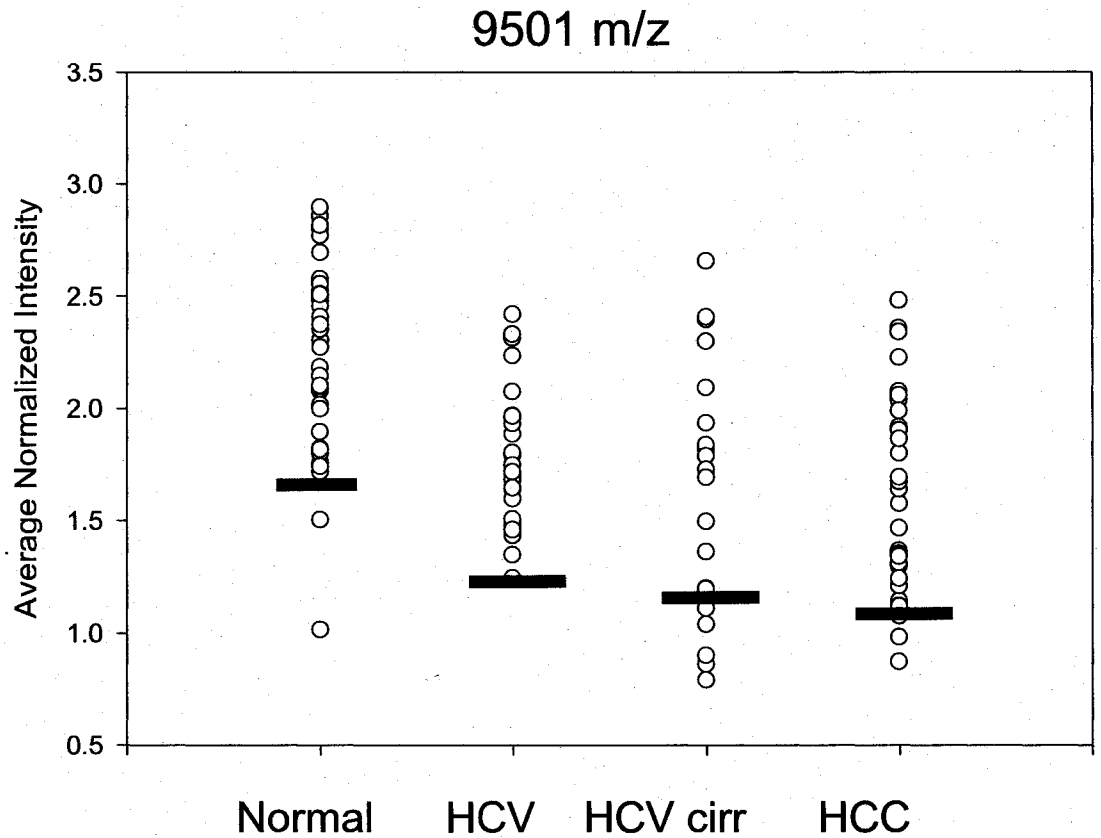


Figure 9. Expression levels of the 9,501 m/z protein. For each sample in all of The indicated HCV disease stratified data sets. Black bars indicate the mean normalized intensity;open circles represent values of individual samples. HCV, Hepatitis C virus infection but no cirrhosis; HCV cirr, HCV-associated cirrhosis; HCC, HCV associated hepatocellular carcinoma.

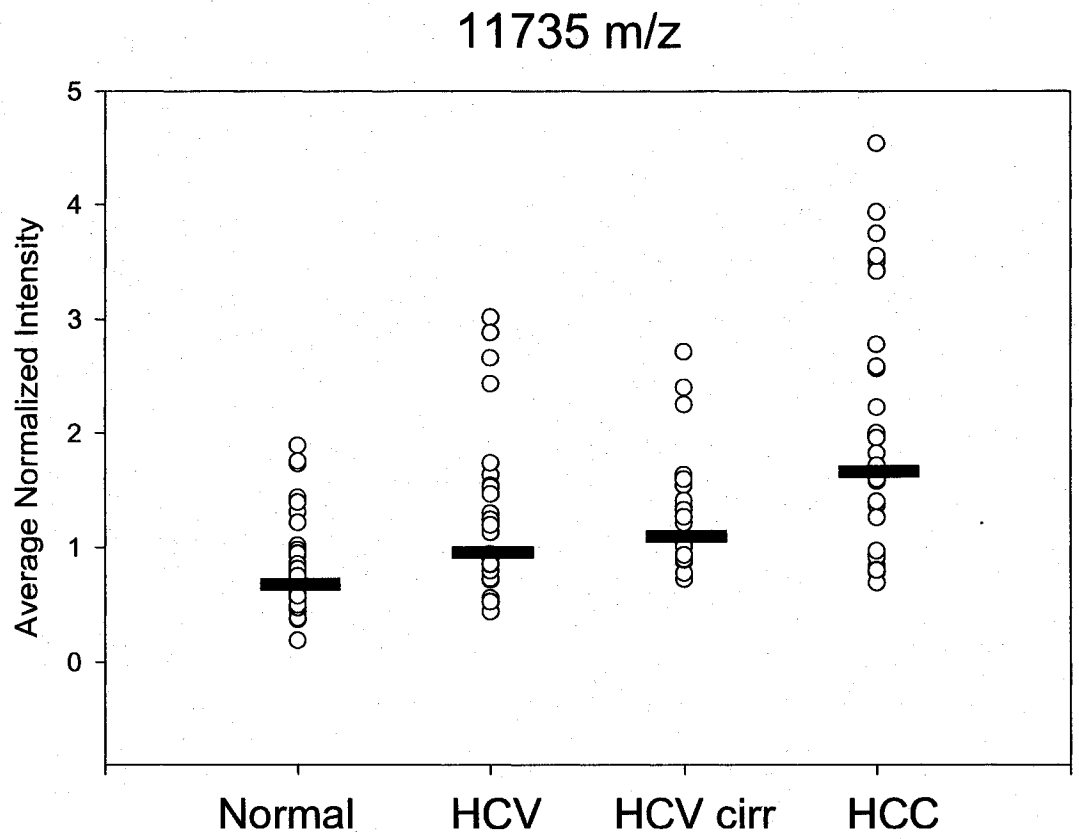


Figure 10. Expression levels of the 11,735 m/z protein. For each sample in all of The indicated HCV disease stratified data sets. Black bars indicate the mean normalized intensity;open circles represent values of individual samples. HCV, Hepatitis C virus infection but no cirrhosis; HCV cirr, HCV-associated cirrhosis; HCC, HCV associated hepatocellular carcinoma.

peaks increased with disease severity. The levels of the 8,939-m/z protein were also increased following HCV infection and were actually highest in the serum from the HCV-associated cirrhosis subjects. Conversely, the intensity of the 9,501-m/z protein decreased in all HCV-associated serum samples relative to the healthy subjects. Table 2 shows p-values of the mean intensities of each of the four peaks for six pair wise comparisons between the different patient groups. When comparing normal samples to liver diseases, the p-values decreased with disease severity. The changes were variable when comparing differences among chronic hepatitis, cirrhosis, and HCC samples, although the 5.8 and 11.7m/z markers were able to distinguish HCV patients with cirrhosis from HCV-associated HCC. The cumulative data illustrate how changes in multiple biomarker proteins can be used as fingerprint patterns reflective of disease state, even though any single protein would not be sufficient for classification.

Principle Component Analysis

Classically principle component analysis or PCA is a multivariate analysis method that attempts to reduce the dimensionality of the data. It is used in CiphergenExpress software to visualize two and three-dimensional graphs. Unfortunately, the software automatically picks the three components by which it stratifies the samples so they are unknown, what is known is the analysis uses the peaks detected in the software and clusters them by disease group i.e. HCC or Normal. In this way we only utilized this analysis as a snapshot for the ability of the data to group together according to disease group. The PCA analysis for HCC and normal is included in Figure 11 and the HCV cirrhosis vs. HCC in figure 12. The PCA analysis was our initial clue this dataset might have the ability be stratified according to severity of disease. Interestingly, the HCC and Normal analysis appears to differentiate between the classes as shown in figure 11. However, figure 12 demonstrate the difficulty and overlap between the later stages of disease, re-emphasizing how difficult it is to distinguish HCV-cirrhosis from the development of HCC.

Decision Classification Tree Analysis

For each sample analyzed, intensity values for each peak in the 1.5 to 20-kd range were averaged for duplicate samples and input into the BioMarker patterns software (Ciphergen Biosystems) for classification tree analysis as described in Materials and Methods. The

Table 2. Comparison of P Values for the Indicated Markers in Different Sample Pairs

Marker (m/z)	Normal vs. HCV	Normal vs. HCV-Cirrhosis	Normal vs. HCC	HCV vs.HCV Cirrhosis	HCV vs. HCC	HCV Cirrhosis vs. HCC
5.8	.72	$<5 \times 10^{-5}$	$<5 \times 10^{-5}$	.077	$<5 \times 10^{-5}$	.002
8.9	.0002	$<5 \times 10^{-5}$	$<5 \times 10^{-5}$	.109	.91	.050
9.5	$<5 \times 10^{-5}$	$<5 \times 10^{-5}$	$<5 \times 10^{-5}$	.037	.02	.800
11.7	.0015	$<5 \times 10^{-5}$	$<5 \times 10^{-5}$	.151	$<5 \times 10^{-5}$	.001

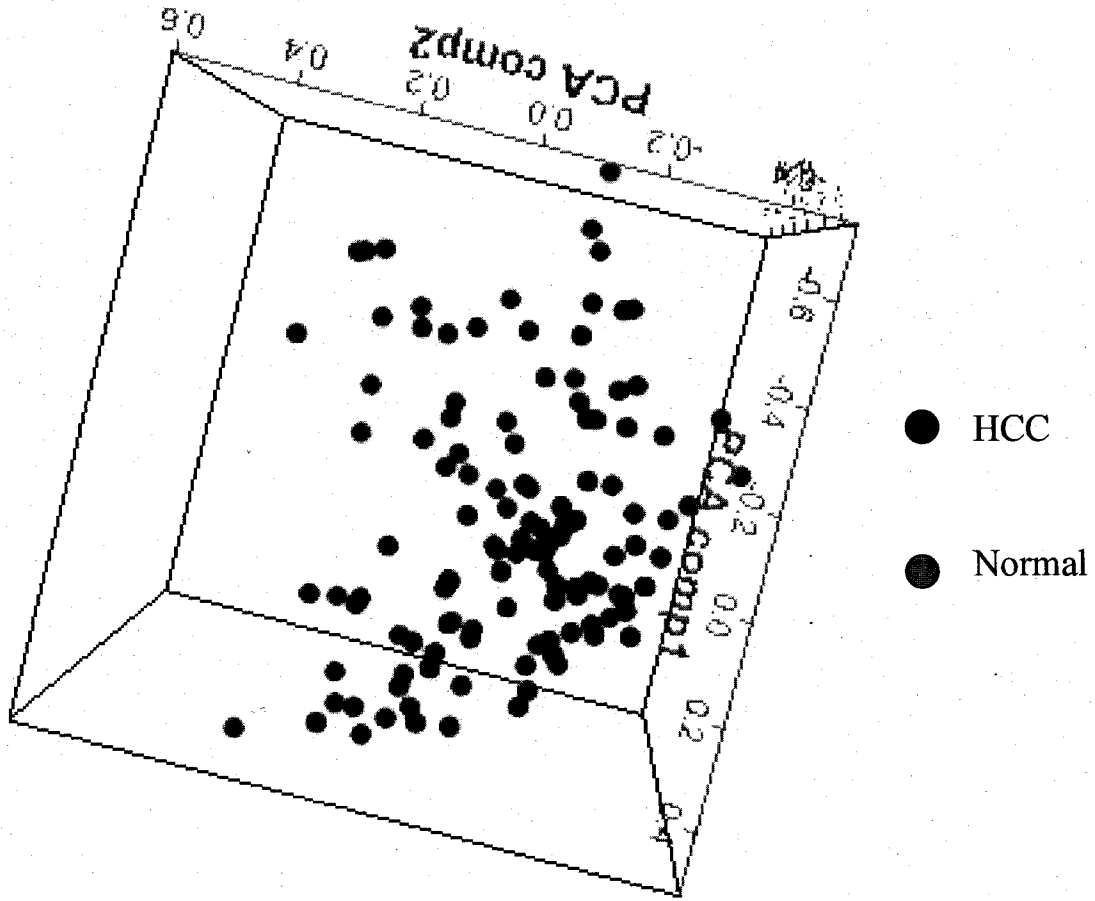


Figure 11. Principle component analysis of HCC Vs. normal. As shown in the three dimensional chart, the normal samples appear to form a tight cluster with themselves, whereas the HCC samples appear more scattered.

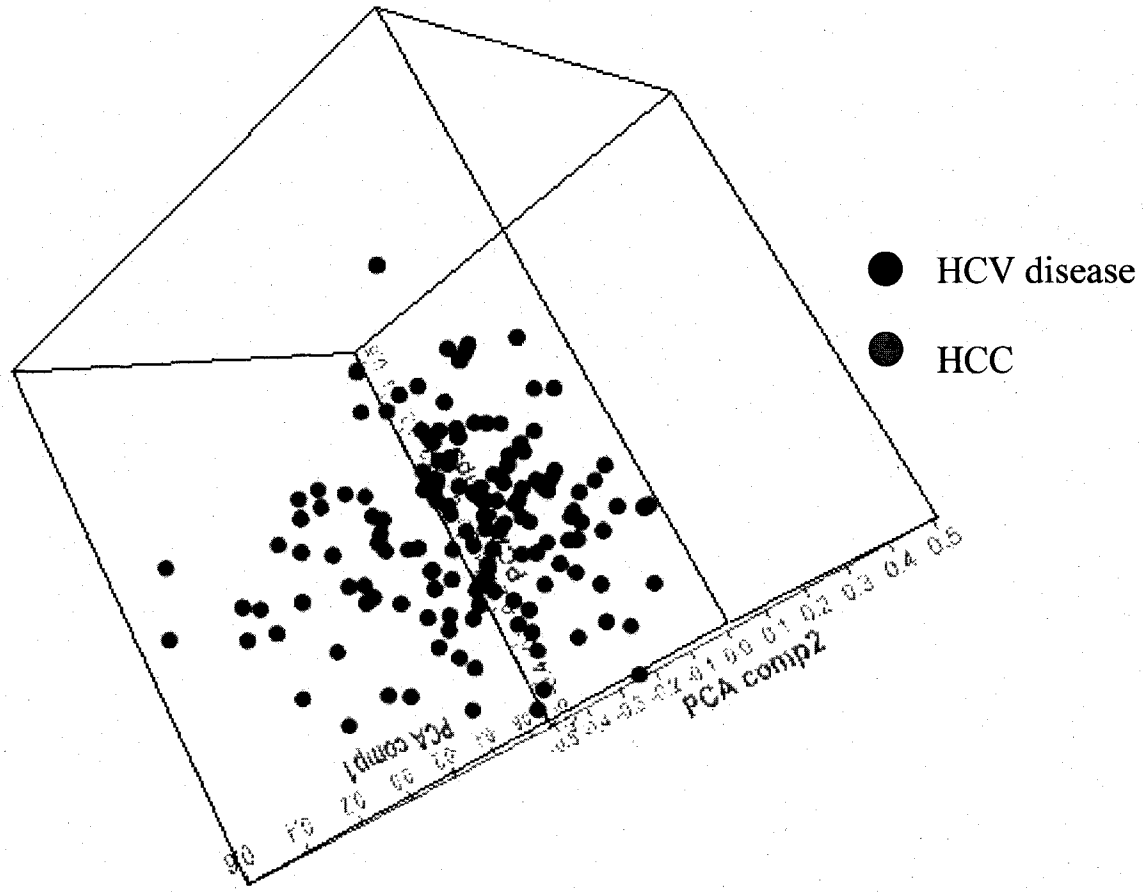


Figure 12. Principle component analysis of HCC Vs. Disease. In contrast to figure 11 all the samples seem to over lap and it is hard to see any stratification.

classification trees split the data into two nodes, using one rule at a time in the form of a question. The splitting decisions in this case were based on the normalized intensity levels of 39 shared peaks from the SELDI protein expression profile across each sample, such that each peak was used as a variable in the classification process. An internal 1/10-sample exclusion, cross-validation process was done automatically for each decision tree generated, and the most significant and best performing tree for each condition was chosen. A separate blinded sample set of sera from healthy individuals (n = 42) and those with HCC (n = 56) (Table 3) was also evaluated for the validation of the classification trees. The clinical characteristics of these sera were analogous to those described in Table 1 for the learning set sera. Initially, serum profiles from healthy subjects were compared individually with serum profiles generated from non-cirrhosis, cirrhosis or HCC samples, including the subset of HCV-infected samples (Table 4). A representative diagram of a decision tree for the comparison of all HCC versus healthy patients is shown in Figure 13, the result corresponding to group III.A. in Table 4. A similar process was used to generate a decision tree for each sample pair presented in Tables 3 and 4. Only the cross-validation results for each decision classification analysis are presented. Following generation of the SELDI spectra for the blinded sera on the IMAC-Cu protein chips, the obtained peak clusters were applied to the previously optimized decision tree (Figure 13). A correct classification of 91% (51 of 56) for HCC and 76% (33 of 42) for healthy patients was obtained.

The results in Table 4 indicate that the SELDI peak profiles were progressively more effective at distinguishing normal samples from non-cirrhosis, cirrhosis and HCC conditions as the severity of the disease increased, regardless of etiology. A separate stratification of these samples was performed based only on an HCV-associated etiology and analyzed separately. The sensitivities of HCV-associated diseases versus normal were better compared when all other liver disease were combined as presented. For example, correct classification of HCV-cirrhosis conditions increased to 91% compared with 72% in the cirrhosis sample set that included other types of liver disease .The ability to correctly classify the more clinically relevant scenarios for surveillance of HCV disease progression from chronic hepatitis to cirrhosis to HCC was examined using the HCV disease stratified sample set used in the initial training set. The results are

Table 3. Demographic Information of the Validation of the Classification Tree

	Healthy (n = 42)	HCC (n = 56)	P Value (p > .05)
Sex (Male/Female)	28:14	38:18	NS
Age (yr)	53 ≤ 11	55 ≤ 13	NS
Ethnicity (%)			
NHW/AA/H/Asian	88/12/0/0	90/4/6/0	NS
Etiology (%)			
HCV		56	<.001
HBV		3	NS
Alcohol		10	<.001
Cryptogenic		18	.01
Other		13	NS
MELD score	5 ≤ 0.7	8 ≤ 2	<.01
ALT (IU/mL)	21 ≤ 9	75 ≤ 40	<.001
AST (IU/mL)	18 ≤ 5	111 ≤ 65	.003
Bilirubin (mg/dL)	0.3 ≤ 0.2	1.5 ≤ 1	.07
AFP (ng/mL)	1.8 ≤ 0.6	234,871 ≤ 3,592	.003
% < 20	100	32	
% 20 - 200	0	45	
% > 200	0	23	
TNM stage % (I/II/III/IV)	NA	13/24/38/25	

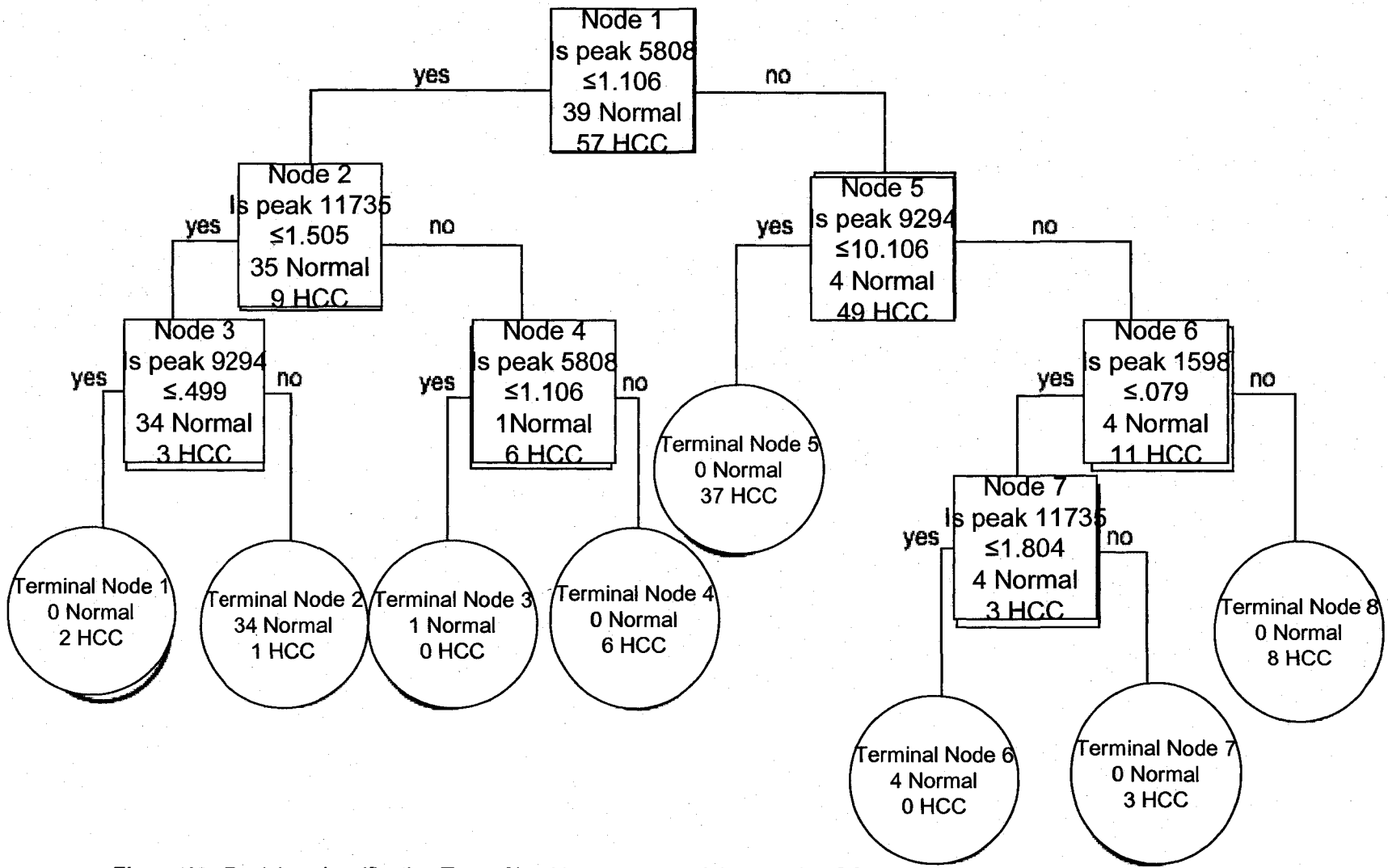


Figure 13. Decision classification Tree of healthy patients and those with HCC. The squares are primary nodes and the circles are terminal nodes. The mass value in the root nodes is followed by the intensity value or less. For example at each node a question is asked to stratify the data into its appropriate group based on SELDI peaks.

Table 4. Decision Tree Class Results for Normal Samples Vs Different sets of Liver Disease

Condition	Sensitivity	Specificity
I.A. All non-cirrhosis liver disease vs. normal	67% (25/38)	66% (26/39)
I.B. HCV non-cirrhosis vs. normal	74% (20/27)	74% (29/39)
II.A. All cirrhosis vs normal	72% (26/36)	72% (28/39)
II.B. HCV-cirrhosis vs normal	91% (20/22)	91% (35/39)
III.A. All HCC vs. normal	82% (47/57)	90% (35/39)
III.B. HCV-HCC vs normal	89% (25/28)	95% (38/39)
III.C. All HCV disease vs. normal	79% (65/82)	82% (32/39)

reported as sensitivity or ability to correctly classify disease cases, and specificity or ability to correctly classify the patients without disease. Chronic HCV samples could be distinguished from HCV-HCC samples with a sensitivity of 71% and specificity of 64% (Table 4). When both chronic hepatitis and cirrhosis samples were combined and compared with HCV-HCC samples, sensitivities decreased to 61%, but specificity increased to 76% (Table 4). To test whether the accuracy of classifications could be increased by including clinical data in the analysis, values for AFP, DCP and GP73 were included in the classification decision tree analysis with all of the SELDI peaks. These values were considered by the algorithm to be additional "peaks" to be used with those from SELDI analyses to build classification trees. As seen in Table 5, inclusion of these marker protein values increased the correct classification of disease states to 79%/86% and 75%/92% sensitivity/specificity, respectively, for both sample sets.

2.4 Discussion

Proteomic analysis of tissue or serum derived from HCC subjects is an emerging technique for the identification of biomarkers indicative of disease severity and progression (37, 49, 141, 142). To date, new HCC biomarkers have been found primarily by using differential two-dimensional gel separations of tumor tissues, (141), or serum together with mass spectrometry for protein identification (37, 131, 142). Although these approaches have greatly increased the ability to quantitate the expression of individual proteins between case and control samples, they are not useful for identifying proteins with masses less than 10 kD and are not amenable to high throughput analysis (9). Using a clinically defined serum sample set reflective of liver disease progression to HCC, we have obtained results that support the use of SELDI-TOF profiling as a surveillance tool to follow disease progression. We also show that known serum markers or other clinical test data can be incorporated into SELDI peak analysis. The SELDI analysis was comparable to the performance of AFP, the only clinically used marker in this pilot study. The combination of the SELDI-derived protein peaks with known serum marker data has also been reported to increase correct classification of a pancreatic cancer serum cohort (143).

Another HCC serum proteomics study using SELDI and immobilized metal affinity protein chips was reported by Poon et al (49). This study took a different approach in that the serum

Table 5. Analysis of All 38 SELDI Peaks and Determined Serum Levels of 3 Marker Proteins (AFP, DCP, GP73)

Condition	Sensitivity	Specificity
Chronic HCV vs. HCV-HCC	71% (20/28)	64% (14/22)
Chronic HCV vs. HCV-HCC + (AFP/DCP/GP73)	79% (22/28)	86% (19/22)
HCV disease vs. HCV-HCC	61% (17/28)	76% (37/49)
HCV disease vs. HCV-HCC + (AFP/DCP/GP73)	75% (21/28)	92% (45/49)

samples were initially pre-fractionated into six separate components before application to the copper affinity chips (or a weak cation chip), and a neural network type cluster analysis algorithm was used for classification following peak selection. The best sensitivities and specificities of 92%, and 90%, respectively, were obtained for distinguishing chronic liver diseases (n = 20) from late-stage HCC, (III/IV, n = 24) samples. Our study differs significantly in that more early-stage HCC samples (I/II) were evaluated, no prefractionation of serum was done, a larger sample size was used, and the clinical serum set analyzed was more fully stratified for the different liver diseases. Also, we chose to use a non-fractionation of the serum prior to protein chip analysis, since this allows a higher throughput and minimizes reproducibility problems for large numbers of serum samples. Use of a decision tree classification algorithm further enhanced our studies by allowing the inclusion of other serum marker data. Even though two distinct sample preparation and algorithm analyses were applied, the Poon et. al. study and our study demonstrate that SELDI protein profiling of serum could provide discriminating protein peaks in classifying chronic liver diseases and HCC (49).

Although the decision classification analysis here is suitable for analyzing two disease groups, it is not effective for intraspectral analysis; its use was limited in that it is able to compare only two conditions when four sets required analysis. Still, the methods described in this research project do demonstrate the potential of this type of platform in the development of a direct clinical assay, because the process involves minimal processing of serum, a fully automated loading and chip-binding procedure, and enough peak features to determine differences in disease states. This approach emphasizes peak/protein selection prior to classification and can allow the subsequent identification of potential biomarkers of known molecular mass. There is also enough flexibility in the process to allow inclusion of other clinical data from known tumor or disease markers (e.g. AFP, DCP, GP73). The ability of SELDI-TOF MS technologies to use protein fingerprints or patterns to discriminate case from control in a high throughput manner is the real strength of this methodology. The largest drawback to this approach is in its inability to identify any of the proteins, which make up the patterns, without the use of an additional purification method.

At the time of publication, it was originally hypothesized that the discriminating peaks and subsequent protein profiles detected in our study by SELDI software could itself be diagnostic, and that this utility would not be dependent upon identification of the proteins in discriminating peaks. The original hypothesis was based on the supposition that the proteins are shed directly from the tumors since serum and plasma remain in continuous contact with tissues and cells. Thus, serum from a healthy individual would consist of different proteins than serum a from cancer sample, which in theory would contain tumor derived products (49). Since this time, this hypothesis has been met with much skepticism, resulting in two basic theories as to the true identity of these proteins detectable by SELDI profiling. The first theory is that these are not tumor proteins at all, but acute phase reactants produced in response to injury of local inflammatory cells such as neutrophil granulocytes and macrophages. Since HCC primarily effects the liver causing inflammation it is possible that this could account for the identified SELDI peaks, yet if this is the case the discriminatory peak would have to be there from the earliest time of inflammation, at pre-cirrhosis, and not arise in the late stage HCC. In our study differential protein patterns were obtained across the three liver disease states, which suggests that these peaks are more than just acute phase reactants.

The second theory is that the detected proteins are actually cleavage products from proteases secreted from the tumors (144). There have been numerous studies designed to investigate whether peptides bound to circulating carrier proteins could be fragments of proteins associated with cell growth and oncogenesis (45, 104). One study, which supports this protease cleavage hypothesis, is the finding of differentially expressed proteins within HCC tissues that are ER resident proteins, which were identified to have lower masses then expected. It was determined that these proteins are actually cleavage products of their intact versions, and included such ER resident proteins as calreticulin and calnexin (131).

In our serum-processing protocols, protein fragments attached to carrier proteins or cleaved by tumor specific proteases would be stripped off their carriers by the urea/CHAPS dilutions, which is one explanation for the presence and detection of the disease-specific peptides. In addition, it is possible that the lower-mass polypeptides found in our study represent fragments of

acute phase proteins, generated from tumor proteases as mentioned above. Therefore, the pattern of their appearance and, ultimately, identification of the proteases that generated them, could provide direct insights into disease pathogenesis.

One recent study has identified an 8.9-dalton peak capable of discriminating patients with HCV-cirrhosis and HCC to be complement C3a (124). We speculate that this 8.9 is the same as the 8.9 peak identified in our study. Biologically, C3a is a complement protein resulting from the activation of the complement pathway and subsequent cleavage of C3 into C3a and C3b. The C3b product often remains with the pathogen, while C3a, the smaller portion of C3, can diffuse from the site of activation. Clearly this could be an indication of disease to find an activated innate immunity protein within the serum of a patient with HCC. However, within this study C3a was not found to correlate with any definitive clinical feature of HCC and has been identified as a marker of breast and colon cancers (145, 146). Also of concern is that C3a is a high abundant serum protein, which adds yet another highly abundant serum protein to the list of potential cancer serum biomarkers.

The observations seen with C3a highlight the issues associated with the dynamic range of concentrations of serum and plasma proteins (111). As discussed in section 1.11, the most abundant proteins (albumin and immunoglobulin) in serum and plasma are ten orders of magnitude higher then the lowest abundant proteins (cytokines) (111). Technologies such as SELDI, do not have the capability of profiling the entire dynamic range of the proteome. In a recent study by Zhou et. al, normal serum was fractionated using a SELDI ProteinChip with WCX affinity coupled to tandem mass spectrometry for identification of all proteins captured (45). Over 300 proteins were captured and sequenced, most of which would be considered high abundant serum proteins including apolipoproteins, haptoglobin and ceruloplasmin. Consequently, none could be considered low abundant proteins within serum or plasma (45). These results subsequently highlight the weakness of our HCC SELDI analyses, especially in regards to the identities and concentrations of the potential disease biomarkers of interest (9).

To address the deficiencies of technologies such as SELDI-TOF, there has been much recent technological advancement including the development of better fractionation approaches

to target the lower abundant proteins with serum and plasma (9, 45). Though there are many different strategies, most of them include an upfront depletion of the highest abundant proteins within serum and plasma to decrease the sample complexity of the starting material (114). Another area which needs to be assessed is the role of posttranslational modifications in regards to serum biomarkers, as has been described for apolipoprotein AII (26). N-linked glycosylation is a common posttranslational modification previously associated with disease development, (75-80, 113), which has recently been assessed for its role in HCC progression (30). For example, fucosylated alpha-fetoprotein has been shown to be a highly specific tumor marker for HCC. In addition Nakagawa et. al has shown direct evidence for fucosylation as a regulation mechanism allowing secretion of hepatic specific glycoproteins into bile ducts (147), Although these markers would only be found in bilary secretions in normal physiological situations, in HCC these proteins could be otherwise targeted to secretion into serum and other liver compartments.

There is also evidence for these disease specific glycosylation events in HCC serum. For example. Block et. al has developed a targeted glycoproteomic approach to assess whether glycoproteins could serve as serum markers of HCC (30). This approach allowed for the discovery of GP73, which had elevated expression in animals with HCC compared to normal samples when captured using a lectin recognizing fucosylated structures. This study therefore shows direct correlation between altered glycosylation, specifically fucosylation and progression of HCC (30). This was assessed further in a research effort by the same group, which assessed the oligosaccharide structures in serum from HCC patients as opposed to those with HBV infection only, or those without disease (37). This study found an increase in the level of core fucosylation in serum of patients with liver cancers as compared to those without HCC (37). Since the authors of this work are collaborators of this study described herein, the larger cohort of HCC, which is being collected from the SELDI study, will be assessed for glycosylation variants using the similar protocols developed by this group and adapted to the studies outlined in the following prostate chapter.

In conclusion, our initial HCC serum analysis made use of the strengths of the SELDI technology by including a relevant clinically defined serum cohort, as well as including additional

clinical diagnostic values, which was novel at the time. Within the Early Detection Research Network, our collaborator Dr. Jorge Marrero at the University of Michigan has been accruing over 450 new, prospectively collected HCC-related serum samples at five clinical sites. This cohort will be further assessed for the effectiveness of DCP and GP73 as cirrhotic and HCC biomarkers, as well as be made available to EVMS for follow-up serum proteomic studies. As alluded to earlier, it is clear that the future of proteomic profiling with mass spectrometry technologies will need to include multi dimensional approaches that should allow for identification of many more proteins (111). Assessment of posttranslational modifications in relation to disease progression is one area that we feel has the potential for the discovery of new biomarkers, which is the direction that was pursued following this initial SELDI exploratory phase.

CHAPTER III

AIM II: TO USE A TARGETED MULTI-LECTIN AFFINITY APPROACH TO IDENTIFY GLYCOPROTEINS PRESENT IN PROSTATE CANCER SERUM FROM BENIGN PATIENTS AS WELL AS THOSE WITH LATE STAGE PROSTATE CANCERS

3.1 Introduction

Glycoproteomics is a newly emerging branch of proteomics that focuses on characterizing any changes in the glycosylation patterns occurring on proteins. (30, 37, 42) As discussed in section 1.10 changes in glycosylation patterns may be of great significance, since they are known to influence many cellular processes (24, 30, 79, 80, 148). In addition, aberrant glycosylation of glycoproteins and glycolipids has been observed in many cancers including breast, liver, and prostate. (30, 37, 79, 80, 94) One recent glycoproteomic study by Comunale et al found an overall increase in the fucosylation of proteins in patients with advanced liver cancers. An increase in fucosylation of glycoproteins with progression to cancer has also been shown to occur on PSA in prostate cancers (94). These studies have generated a great interest in characterizing the serum glycoproteome to identify potential biomarkers for cancer occurrence and progression.

Consequently, many lectin based strategies to capture serum glycoproteins specific for cancer detection have been developed. Many of these approaches include an upfront antibody or affi-gel based depletion of the most abundant serum proteins such as albumin and immunoglobulins, followed by lectin capture and characterization (24, 148, 149). We present here a different approach that does not involve an upfront depletion. In the study reported herein we chose to utilize a panel of different lectins shown in Table 6, to enrich for serum glycoforms found in serum from prostate cancer. The experimental design for aim #2 can be seen in Figure 14. One unique feature of this study is the use of the prostatic serum sample groups listed in Figure 15, which were stratified by a pathologist as a relevant clinically defined sample set. By using these serum samples we sought to identify serum glycoprotein isoforms that could be correlated with specific prostatic disease pathologies. This type of clinical relevance has not

Table 6. Lectin Panel

Lectin	Specificity	Binding Buffer	Elution Buffer
AAA Anguilla anguilla (fresh water eel)	alpha 1,2 linked fucose	20 mM TBS pH 7 1mM manganese chloride,1mM calcium chloride,1mM magnesium chloride	200 mM fucose
AAL Aleuria aurantia (mushroom)	alpha 1,6 linked fucose	20 mM TBS ph 7 1mM manganese chloride, 1mM calcium chloride, 1mM magnesium chloride	200 mM fucose
ConA Concanavalin A (jack bean)	alpha mannose alpha glucose	20 mM TBS pH 7 1mM calcium chloride, 1mM manganese chloride, 1mm magnesium chloride	200 mM mannose
HPA Helix pomatia agglutinin (snail)	GlcNac GalNac Sialic acids	20 mM Tris-HCL 50 mM NaCL pH 7.6	200 mM GlcNac or GalNac
MAL Maackia amurensis (seed)	alpha 2,3 linked sialic acids	20 mM Tris-HCL 50 mM NaCl pH 7.6	100 mM glycine
SNA1 Sambucus nigra 1 agglutinin (tree bark)	alpha 2,6 linked sialic acids	20 mM TBS pH 7 1 mM manganese chloride, 1mM calcium chloride, 1mM magnesium chloride	100 mM lactose
WGA Wheat germ agglutinin	terminal GlcNac Sialic acids	.05 M $NaPO^4$ .02 M NaCl	200 mM GlcNac

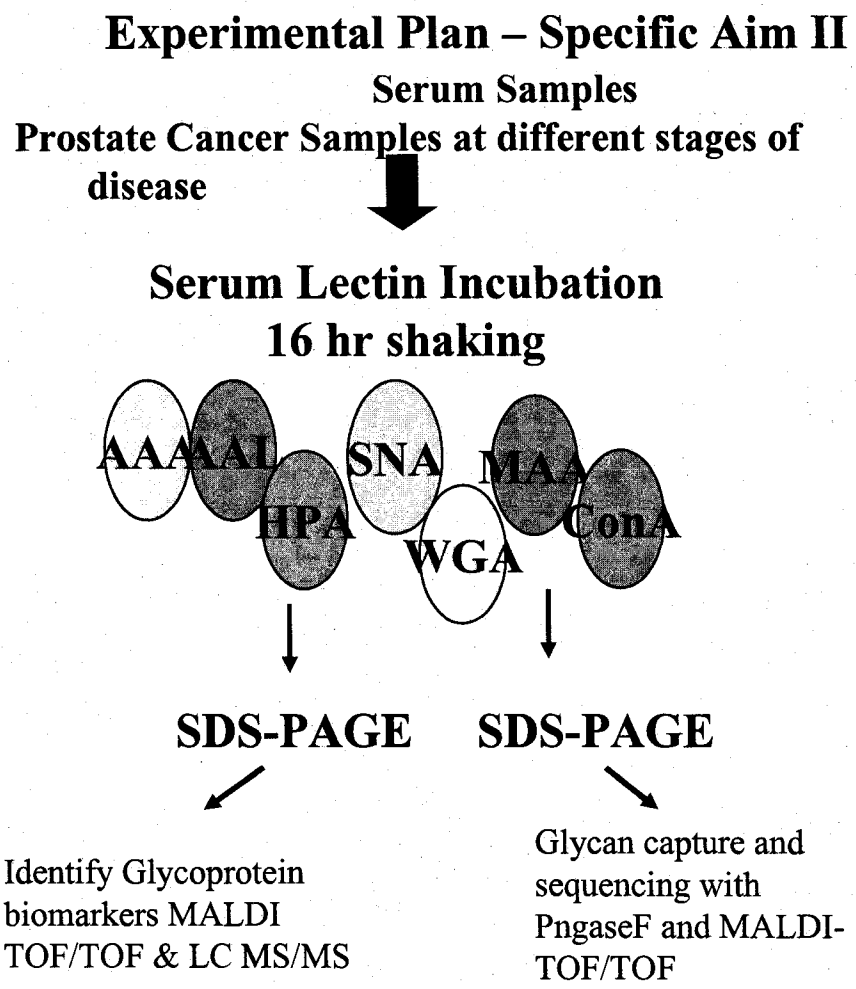


Figure 14. Experimental Design of Aim II.

Group A: Predicant Like study : Cancer vs. Non-cancer with normal PSA (2-4ng/ml)

A1 biopsy -, A2 biopsy +

Group B: Staging Groups

B1= 5 patients with early localized cancer T1,T2

B2 = 5 patients wih stage T3,T4 advanced localized cancer

B3 = 5 patients with + bone scan and PSA > 60 metastatic prostate cancer

Group C: Benign Disease vs. Cancer

C1 = healthy donors PSA<4.0 ng/ml, normal DRE

C2 = BPH 5 patients with benign disease PSA 4-10 biopsy negative

C3 = low grade cancer Gleason score < or = to 6

C4 = High grade cancer Gleason > or = to 7

Figure 15. Prostate Serum Cohorts Utilized in the Lectin Capture Study. Each group consisted of a pool of five serum samples.

been fully addressed in other studies, which have been more deeply concerned with method development (24, 115, 148, 149).

In addition, since glycoslation is a complex process resulting in different glycoforms, (37, 94), which cannot be visualized using 1D SDS-PAGE yet could be an important diagnostic feature, we also present a strategy for the capture and purification of glycans from both prostatic acid phosphatase (PAP) and IgG using MALDI-TOF/TOF.

3.2 Materials and Methods

Serum Samples

All serum samples utilized within this project were obtained from the Virginia Prostate Center Tissue and Body Fluid Bank. Blood samples collected under the same protocols from properly consented patients were obtained from the Department of Urology, Eastern Virginia Medical School (EVMS). All samples were collected following IRB approved protocols.

Lectin Capture

To determine differential glycoprotein profiles of each sample, bead-bound lectins with known oligosaccharide binding properties were utilized see (Table 6). The panels of lectins chosen have varying oligosaccharide binding patterns including: Anguilla anguilla (AAA: EY laboratory) which recognizes alpha 1,2 linked fucose residues, Concanavalin A (ConA: EY laboratory) binds alpha mannoses and alpha glucoses, Aleuria aurantia (AAL: Vector Labs) recognizes alpha 1,6 or 1,3 linked fucose, Maackia amurensis (MAA: EY laboratory) binds alpha 2,3 linked sialic acids, Helix pomatia agglutinin (HPA: EY laboratory) binds GlcNAcs and GalNAcs, Sambucus nigra (SNA1: EY laboratory) which binds sialic acid residues linked alpha 2,6 and Wheat germ agglutinin (WGA) which binds terminal GlcNAc and sialic acids. The lectin incubations were carried out as follows: 30 µl of patient sera was incubated with 100 µl of lectin beads, processed with the appropriate binding buffers which are detailed in Table 6. Incubation volumes were adjusted to 50 µl patient serum incubated with 150 µl of lectin beads and eluted with 300 µl of elution buffer for the MALDI-TOF/TOF analysis to enable peptide mass fingerprint analysis. The bead/serum mixture was allowed to shake and incubate in the cold for 16 hours. Following incubation all unbound material was collected; the beads washed of all remaining

unbound material, and the bound glycoproteins eluted. The eulates were either next separated using SDS-PAGE on 8-16% Criterion gradient gels (BioRad) and visualized by either Silver Stain (Invitrogen) or Coomassie Brilliant Blue (BioRad). Glycoproteins of interest were cut from the gels and then trypsin digested. All digested peptides were coupled to a MALDI TOF/TOF or LC MS/MS for rapid protein identification. Quantitation was performed using the Bradford assay, where appropriate.

Trypsin Digest and Peptide Mass Fingerprinting

Trypsin digests of Gel Slices For gel bands to be analyzed on the LC MS/MS, Promega trypsin was used, whereas, for gel bands to be analyzed on the MALDI-TOF/TOF, Roche trypsin was used. All gel bands were excised and placed into 1.5 ml Axygen Maximum Recovery tubes and washed with 100 mM NH_4HCO_3 to ensure correct pH. To destain silver stained gel pieces, these were incubated with a 1:1 mix of destain A and B reagents (Invitrogen) and allowed to shake for 20 minutes followed by two 5 minute washes with 100 mM NH_4HCO_3. Coomassie stained gel pieces were destained by incubation with a 50% acetonitrile (ACN), 50% 100 mM NH_4HCO_3 solution for 10 minutes. Following destaining, the gel bands were dehydrated with 50 ul ACN (HPLC grade Fisher) for ten minutes or until the gel slice appeared thin and grainy. Next, the gel bands were reduced with 60 ul of 20 mM DTT (BioRad) for 30 minutes at room temperature (RT). Following reduction, the proteins were alkylated with 60 ul of 50 mM iodoacetamide at RT in the dark. Following alkylation, the gel pieces were dehydrated three times with ACN, and rehydrated once with 10 mM NH_4HCO_3 for ten minutes. Next the gel pieces were dehydrated three more times with ACN and dried in under reduced pressure in a bench-top vacuum centrifuge. Roche trypsin (25 ug) was mixed 200 ul 100 mM NH_4HCO_3, diluted 1:1 with acetonitrile (ACN) and added to the gel pieces. Promega trypsin was mixed to a final concentration of 20 ug with 100 mM NH_4HCO_3 and added to the gel pieces. The digest was carried out at 37° C for no longer than 16 hours. Following digestion, the samples for MALDI analysis were spotted directly, while the samples for LC-MS/MS analysis were subjected to two incubations with peptide extraction buffer (50% ACN and 0.1% trifluoroacetic acid (TFA) with 1

minute spins between additions. Extraction eluates were pooled and dried in the bench-top vacuum centrifuge.

Trypsin digests of ConA Eluates 4 ul (100 µg) of eluate from ConA fraction of samples A1 and A2 was digested with Promega trypsin (1:20), 100 ug protein was digested 16 hours at 37°, following digestion samples were further fractionated with Imac-Cu^{+} Bruker beads according to manufacturer protocol (Bruker) and spotted on an Anchor plate with a ClinProt liquid sample handling robot. Eluates were prepared with α-cyano-4-hydroxycinaminic acid (10mg/ml) and analyzed using an Ultraflex MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) read in reflectron mode.

ELISA Analysis

Alpha-1-antitrypsin (AAT) ELISA analysis was carried out according to the manufacturer protocol. and was purchased from Genway (San Diego, CA). The 96 well plate was coated with AAT capture antibody diluted with .05M carbonate-Bicarbonate at pH 9.6. and allowed to incubate for 60 minutes. Next the plate was blocked using 1% milk in PBS for 60 minutes. Following blocking, the plate was washed 2 times using HPLC grade water (Fisher). The standards were prepared according to the manufacturer protocol and 100 ul of the lectin eluate was added its appropriate wells. The lectin eluates were run in duplicate and prepared in the following manner. WGA, Con A and SNA1 eluates were prepared for ELISA assay be bead capture previously described. Briefly, 50 ul serum was incubated with 150 ul lectin bead and allowed to incubate for 16 hours at 4°. Lectin beads were eluted by competitive inhibition using 200 ul of their appropriate sugar. Lectin eluates were further diluted 1:1 with 1% non-fat milk in PBS. 100 ul of sample was added to the coated plate. The sample was allowed to incubate for 60 minutes and then followed with 2 wash steps with HPLC grade water (Fisher). Next the HRP conjugate was added, it was prepared 1:1 and 100 ul was added according to manufacturer and incubated for 30 minutes. To this substrate 100 ul stopping solution was added and the plate was read at 450 nm.

Glycan Analysis

Analysis of Standards standards were purchased from Prozyme (San Leandro, CA). Standards were reconstituted in 20 µl HPLC grade water and spotted 1:1.8 with DHB 10mg/ml on the AnchorPlate.

PNGase F Gel Analysis this protocol was adapted from Prakash et al. (93)50 ul of lectin eluate was precipitated with 1mL cold acetone and the pellet was resuspended in PBS + SDS (35ul) PBS + 0.5 ul SDS (5%). The sample was next allowed to heat to 100°C for denaturation, and then cooled on ice to RT. When the sample had reached RT, 2.7 ul of 7.5% NP40 + 2 ul of PNGase F was added and the sample was allowed to incubate at 37°C overnight. For the ID PAGE analysis, the sample was acetone precipitated with 1 mL cold acetone, and resuspended in 30 ul rehydration buffer (4M Urea, 20mM DTT, 100 mM tris pH 8.0, 4% SDS and bromophenol blue) and loaded onto a BioRad 8-16% gel, visualized with Brilliant Blue Coomassie (BioRad).

Glycan Isolation Proteins with attached glycans of interest were first separated by 1D-SDS PAGE and visualized with coomassie brilliant blue staining (BioRad). Protein bands were sliced from the gel and washed with 20 mM $NaHCO_3$ two times for 30 minutes each. To this solution 40 µl of 45 mM DTT (BioRad) was added and the solution was incubated for thirty minutes at 60° C for 30 minutes. Next the samples were cooled to room temperature and 40 µl of 100 mM iodoacetamide was added and allowed to incubate for 30 minutes in the dark. The DTT and iodoacetamide (BioRad) was removed and the gel plugs were incubated in a 1:1 acetonitrile/20 mM $NaHCO_2$ pH 7.0 solution for 60 minutes.

PNGaseF digestion After incubation, the gel plugs were cut into smaller pieces and completely dried in a speedvac. For the in situ digestion 40 µl total volume was added of the following: PnGase F (Prozyme) (100 U/mL) in 20 mM NaHCO3 pH 7.0. All digests were covered with additional buffer (100 µl) and incubated at 37 º for 12-16 hours.

Glycan Extraction Glycans were extracted from the gel pieces by removing the incubation buffer and using 3 changes of 200 µl HPLC grade water (Fisher) with sonication for 30 minutes. The extracted glycans were purified using a Glycoclean H cartridge (Prozyme).

Glyoclean H cartridge clean up each cartridge was prepared according to the manufacturer protocol as follows. Each cartridge was washed with 3 mLs of NaOH, followed by 6 mLs of HPLC grade water (Fisher). Next 3 mLs of acetic acid was added (Fisher) followed by 3 mLs of HPLC grade water (Fisher). This was followed by 3 mLs of solvent A (50% ACN + 0.1% TFA) and then 3 mLs of solvent B (5% ACN + 0.1% TFA) that was allowed to completely drain. Each cartridge was placed over a collection vessel and the glycans added to the column and allowed to drip into the collection vessel. The glycans were further eluted by addition of 4 (0.5 mL) additions of solvent, A which were forced through the column by pressure. The collected glycans were completely dried down using a speed vac.

Glycan analysis on MALDI-TOF/TOF Glycan pellet was resuspended in HPLC grade water and spotted on a Bruker Anchorplate with saturated DHB matrix (9 HPLC water/1 ACN). The sample was spotted in a 1.2/2.0 ratio with matrix. All glycans were read using reflectron mode.

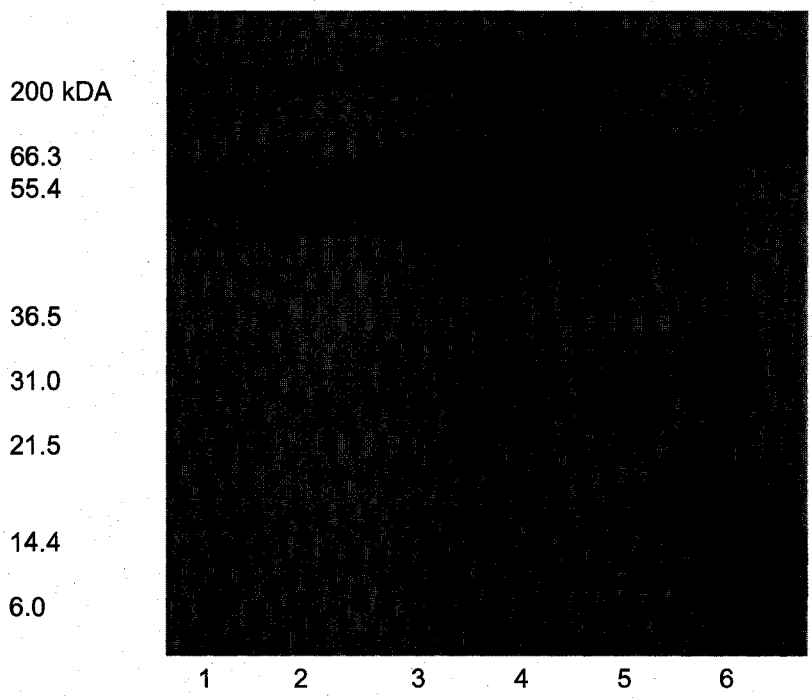
Western Blot Analysis

Roughly 40 ug of protein was loaded based on protein concentrations determined by the Bradford assay (Biorad). Seminal fluid samples were mixed 1:1 with sample loading buffer (4M Urea, 20mM DTT, 100 mM Tris pH 8.0, 4% SDS and Bromophenol Blue) and boiled for 5 min. prior to loading. Pre-cast 8-16% SDS-PAGE NuPage mini-gels (Invitrogen) were used for western blots. Proteins were transferred to a PVDF membrane (Immobilon P, Millipore) for 45 minutes at 400 mA at room temperature. Anti-PAP was incubated at 1:500. Krista Yaudes is responsible for the PAP western shown in Figure 39.

3.3 Results

Preliminary Quantitation and Assessment of Lectin Capture

To determine the extent to which the lectin beads could bind serum efficiently, we first performed a titration experiment with QC serum and WGA beads that is shown in Figure 16. In lanes 7-12 the bead volumes were increased while the serum levels remain constant. As indicated by the albumin band which runs at 66 kD, the beads reach a saturation point to the amount of protein they bind, despite increasing the bead volume. Following this experiment we determined that 30 ul serum and 100 ul of beads would be a sufficient volume to carry out serum



1. 30 µl serum 70 µl beads
2. 30 µl serum 100 µl beads
3. 30 µl serum 120 µl beads
4. 30 µl serum 130 µl beads
5. 30 µl serum 140 µl beads
6. 30 µl serum 150 µl beads

Figure 16. WGA QC Bead Titration. QC serum was added to different bead volumes to determine if a saturation point of bound material could be reached. Eluates were separated by 1D polyacrylamide gel analysis and visualized by silver stain. Roughly 10 ul of eluate was loaded.

glycoproteomic capture. Next, to determine if our lectin capture methodologies would be sufficient for glycoprotein capture from serum, we first performed the affinity capture using pooled normal and cancer prostate samples. To visualize whether lectin incubation resulted in a specific capture, the bound eluates from the lectin were run next to the unbound serum fraction and visualized by silver stain, as shown in Figure 17. Although there are some redundant proteins in both the bound and unbound fractions, there are also unique proteins which are enriched in the bound sample. For example the protein band at 20kD. Also interesting is the decrease in albumin as evident by the albumin band at 66 kD. It should be noted that for comparison reasons only 1µl of the UB fraction was run, compared to 25µl of the bound eluates, which further highlights the simplification/fractionation occurring with lectin capture. These preliminary results were promising, since they confirmed that glycoprotein capture was occurring yet it remained unclear as to how specific and uniform this approach could be.

Therefore, to determine the amount of protein being captured by lectin fractionation, and to assess the reproducibility of this approach consecutive bead experiments using the same quality control serum sample was performed. Three different beads were utilized including: AAA, SNA and WGA, roughly 30 ul of serum and 100 ul of beads were used with each bead experiment. The protein concentrations of the starting QC serum, eluates following lectin incubation, as well as the unbound fraction can be seen in Table 7. The numbers highlight the range of lectin specificities utilized in this project. For example in the first set of AAA captured serum the protein concentrations average 1.44 ug/ ul as opposed to the WGA lectin which increased to 6.9 ug/ ul, which is directly related to their oligosaccharide binding specificities since AAA binds fucose residues, which are less common, than the GlcNac or GalNac residues recognized by WGA. Figure 18 is a snapshot of each lectin utilized in this research project as applied to benign prostate hyperplasia samples versus prostate cancer serum. Clearly differences exist between benign and cancer groups, and it is clear that there is a definitive range of binding specificities dependent upon the lectin utilized. The two lectins, which bind the most proteins by gel analysis is Con, A which binds alpha mannoses, and WGA, which binds GlcNac and sialic

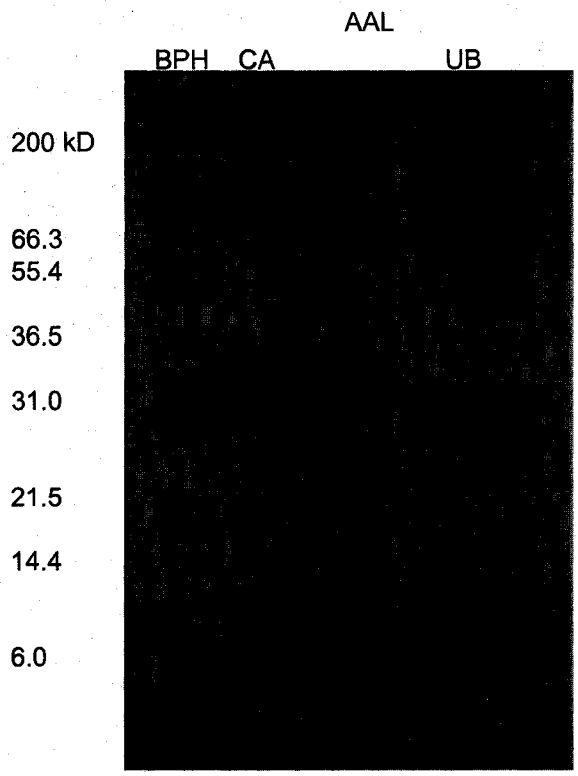


Figure 17. Representative 1D Gel Analysis of Bound and Unbound Lectin Capture. 30 μl of BPH or PCA serum was incubated with 100 μl beads for 16 hours and eluted with 100 mM Fucose. Samples were separated on 8-16% SDS-Criterion BioRad gels and visualized by silver stain. Roughly 20 μl (30u μg/μl) of AAL bound material was run in comparison to 1 ul of unbound sample.

Table 7. Lectin Binding Efficiencies.		
QC serum	----------	76 ug/ul
QC serum	AAA Unbound	60 ug/ul
QC serum	AAA 1	1.22 ug/ul
QC serum	AAA 2	1.44ug/ul
QC serum	AAA 3	1.01 ug/ul
QC serum	AAA 4	2.22 ug/ul
QC serum	AAA 5	1.33 ug/ul
	AAA Average ± std. deviation	1.44 ± .462
QC serum	SNA Unbound	53 ug/ul
QC serum	SNA 1	3.22 ug/ul
QC serum	SNA 2	3.66 ug/ul
QC serum	SNA 3	4.01 ug/ul
QC serum	SNA 4	3.53 ug/ul
QC serum	SNA 5	3.78 ug/ul
	SNA Average ±std. deviation	3.64 ± .294
QC serum	WGA Unbound	64 ug/ul
QC serum	WGA 1	6.64 ug/ul
QC serum	WGA 2	7.22 ug/ul
QC serum	WGA 3	6.55 ug/ul
QC serum	WGA 4	6.79 ug/ul
QC serum	WGA 5	7.44 ug/ul
	WGA Average ±std. deviation	6.928 ± .38

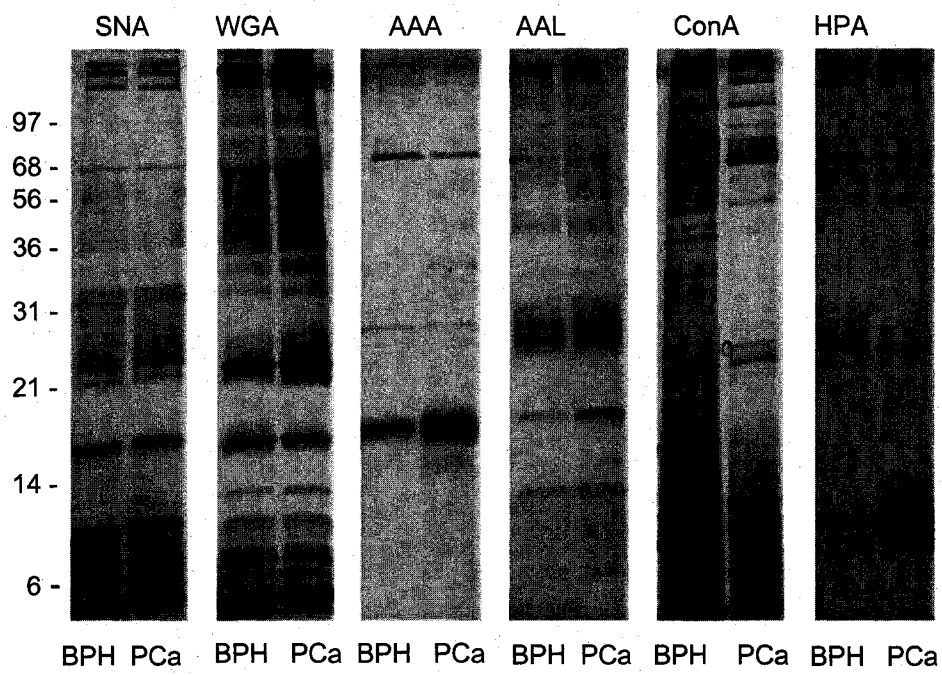


Figure 18. Differential Lectin Capture of Serum Glycoproteins. Pooled serum samples (30 μl) from BPH and PCa patients were incubated with the appropriate lectin agarose beads (100 μl) for 16 hours and eluted with the appropriate monosaccharides. Bound proteins were separated on 8-16% Criterion-SDS-PAGE gels and visualized by silver stain.

acid residues. In contrast, AAA and AAL recognize fucoses and are known to capture far fewer proteins as shown here by gel analysis.

To determine the proteins being captured using this strategy, gel bands from the AAA lectin capture of the BPH and PCa samples as shown in Figure 19 were digested and sequenced by LC-MS/MS analysis. Alpha-fetoprotein, myotubularin related protein, haptoglobin and transthyretin were among the proteins identified. These initial results were promising, since alpha-fetoprotein has been shown previously to hyperfucosylated with HCC progression (147). In addition, haptoglobin has been shown to have different isotypes as shown with 2D-PAGE analysis (150). Figure 20 shows three different BPH samples and three different PCa samples so the protein identities can be visualized in replicate samples. These preliminary experiments prompted us to apply this technique to a clinically stratified and relevant prostate serum cohort in order to determine whether the captured serum glycoproteins could stratify disease stage, and if more potentially relevant glycoproteins could be identified.

Lectin Capture and Identification of Serum Glycoproteins with LCQ-MS/MS or MALDI-TOF/TOF

As detailed in Figure 15, three clinically defined prostate serum groups were included for this study. Since PSA levels are not always a perfect indication of prostate cancers, group A included 10 patients with normal PSA values. Group A1 included five patients having a positive PCa biopsy and Group A2 included five patients with negative biopsy. Clearly, the biological significance here would be to find a marker that could assist and improve PSA in distinguishing these two groups. Group B represents the staging groups: B1 consisted of five patients with early localized cancer versus B2 and B3, which are more advanced stages of disease. Group C provides a range of disease stages including healthy (C1), BPH (C2) low grade prostate cancer (C3) and high-grade prostate cancer including some with metastatic disease (C4). Each lectin listed in table 6 was utilized and applied to each sample. Captured glycoproteins were separated by 1D PAGE and visualized using either silver stain or Coomassie blue. Figure 21 shows a representative gel picture of the lectin AAL run on each sample group and visualized by silver stain. At the beginning of this research project, the only instrument available for protein

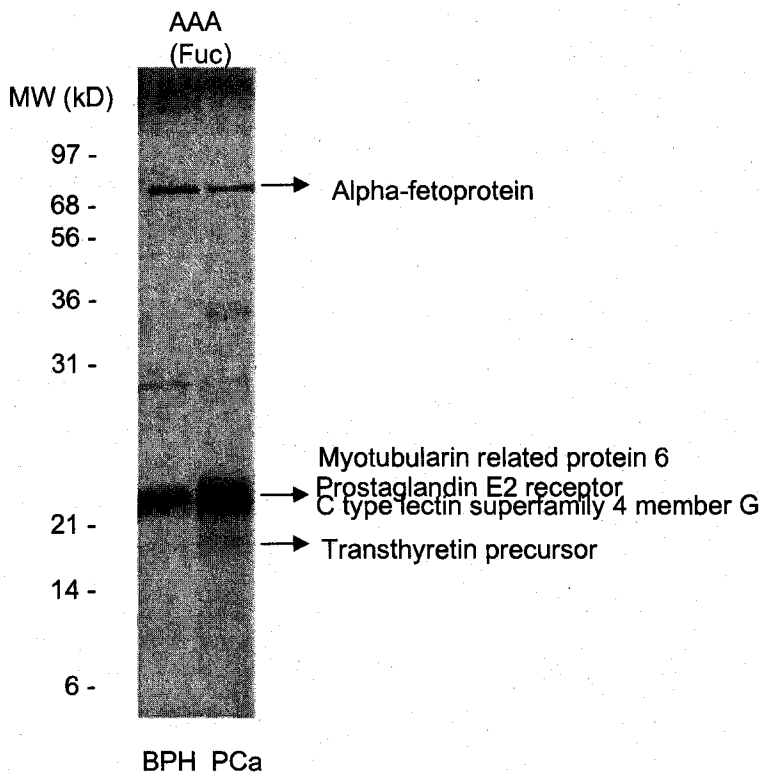


Figure 19. Protein Identities from AAA Lectin Capture. 30 µl of BPH and CA samples were incubated with 100 µl beads for 16 hours. Lectins were eluted and separated on 8-16% Criterion SDS-PAGE gels and visualized by silver stain. Gel slices were trypsin digested according to protocol and protein IDs obtained by LC-MS/MS analysis.

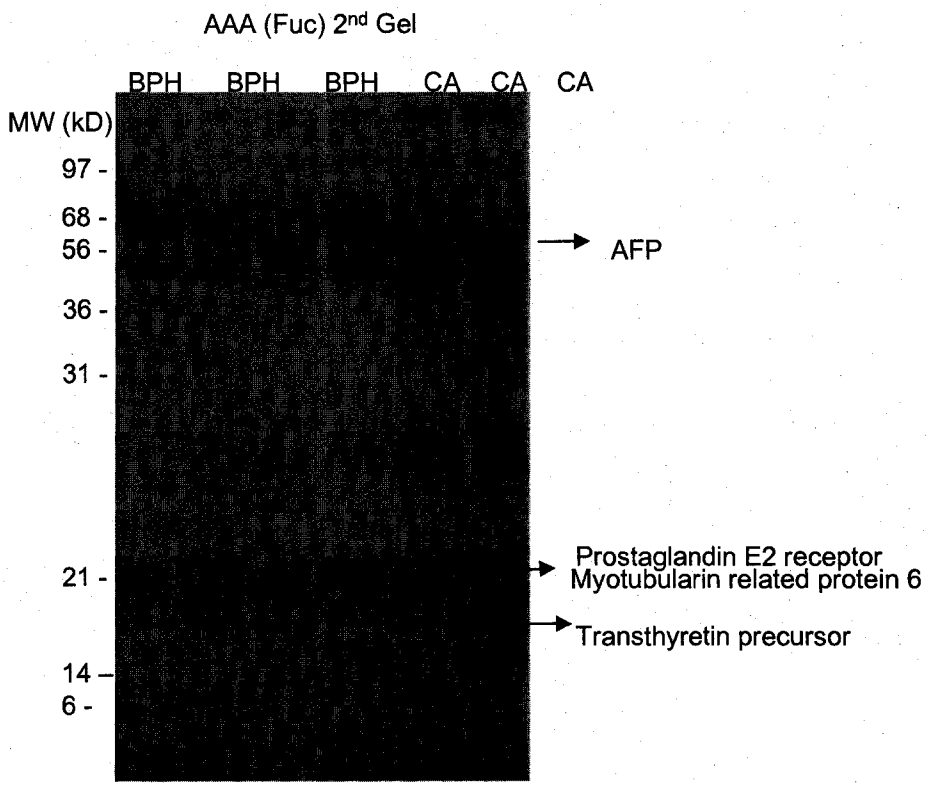


Figure 20. Lectin Capture of 3 BPH and 3 CA Serum Samples with Corresponding Protein Identities. 30 μl of BPH and CA serum samples were incubated with 100 μl beads for 16 hours. Lectins were eluted and separated on 8-16% Criterion SDS-PAGE gels and visualized by silver stain. Gel slices were trypsin digest according to protocol and the identities obtained by LC-MS/MS.

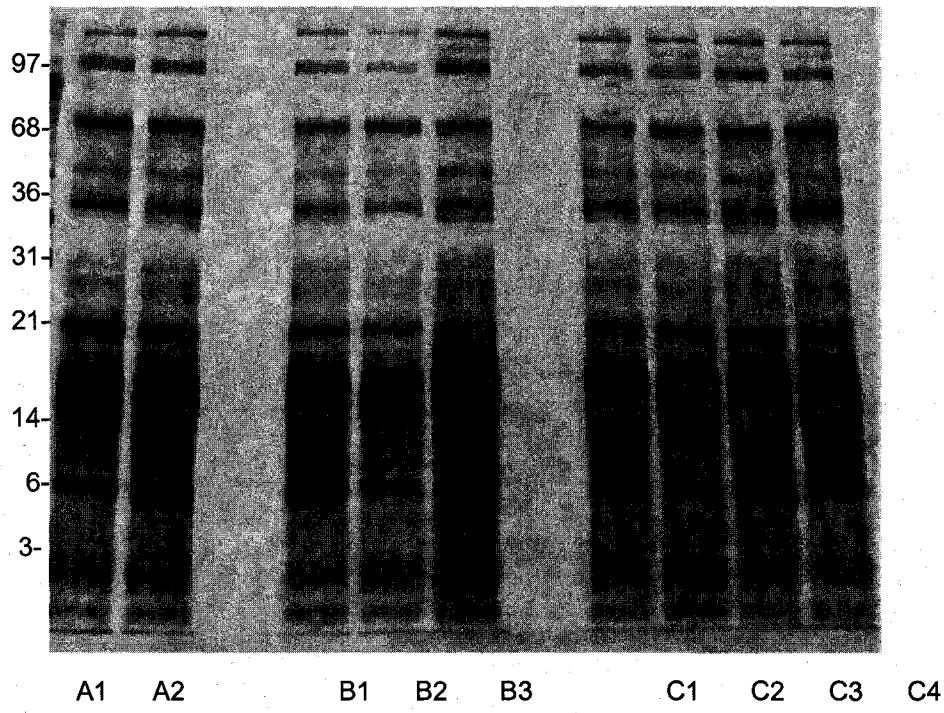


Figure 21. AAL Lectin Capture of Prostate Serum Groups. 30 μl of each serum group was incubated with 100 μl of AAL Lectin bead slurry and incubated for 16 hours. The samples Were eluted with 200 mM fucose and separated on 7.5% Criterion SDS-PAGE gels and visualized by silver stain.

sequencing was the LC-MS/MS. Therefore, many of the initial lectin capture experiments were carried out in this way. Figure 22 is the same gel from above with the numbers added to correspond to the gel slices that were digested and run on the LCQ. Figure 23 lists the identities which were determined from the LCQ run, some numbers are absent, since not every band gave an acceptable protein match. From this protein list it is clear that many of the proteins being identified here are very abundant serum proteins such as alpha 2 macroglobulin, serotransferrin, as well as the immunoglobulins. Whether they are present due to their abundance in the serum proteome or differentially glycosylated is unknown from this initial analysis. This could be determined by glycan sequencing which will be discussed later. Also interesting is the presence of albumin, which is non-specifically captured even though it is not glycosylated. Peptide scores above 60 with a high degree of sequence coverage determined by matching the peptide digest with possible protein identity was utilized to determine whether the protein hit was accurate.

Figure 24 is gel representative of lectin capture of groups A1 and A2 with lectins: AAA, HPA and SNA with the subsequent protein identifications in figure 25. The protein ID results did find both differences and redundancies by utilizing different lectins, for example alpha-1-antitrypsin, and complement factor B were both identified from the SNA eluate which recognizes sialic acids, proteins which had not been previously identified with the fucose recognizing lectin AAL. Yet serotransferrin was identified again but this time using SNA1. Figure 26 is ConA, MAA and WGA lectin eluates, which unfortunately only one provided one protein identification that was albumin. Since glycoproteins can be very large, the eluates were run on 7.5% gradient gels as well as 8-16% gels. Following visualization of these gels it did not seem to show us the breadth we had hoped, and the identifications from these gels were proteins we had previously seen on the original gradients such as Albumin and AFP, therefore the subsequent studies were run on the 8-16% gels. It is possible that a total in solution trypsin digest without gel separation could be a better alternative to this protein complex question, which will be discussed in more detail later.

Around this time, a Bruker MALDI-TOF/TOF instrument was purchased within our laboratory, and made available for subsequent protein profiling and sequencing studies. Due to the high throughput capabilities of this instrumentation, many of the protein sequencing reactions

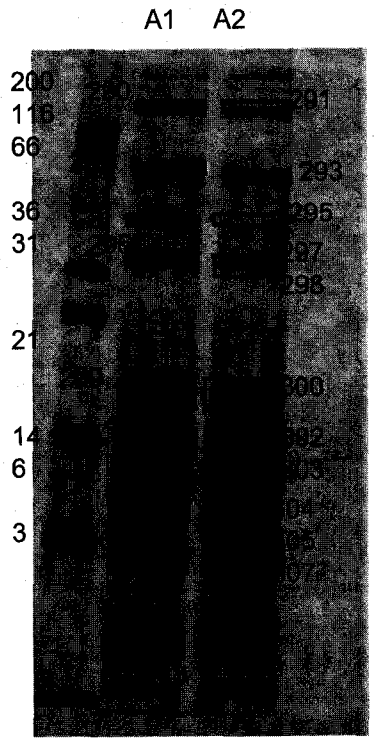


Figure 22. Capture from A1 and A2
Gel Slices Cut for Protein Identification.

Protein ID results (significant hits)	Score
290: Desmoglein 4 precursor 113	47
Slit homolog 3 protein precursor 167k	43
291: too many	
292: alpha 2 macroglobulin 163k	359
293: alpha 2 macroglobulin	359
294: versican core protein precursor 372k	41
295: Pericentrin 2 (pericentrin B) 377k	51
296: FUS interacting serine arginine rich protein 372 K	46
298: Ig Mu heavy chain diseae protein BOT	141
(43K), Ig mu chain C region (49K)	118
299: Serotransferrin precursor 77k	115
Ig mu heavy chain disease protein BOT 43k	111
300: Albumin	270
Ig mu heavy chain disease protein	102
301: Albumin	567
302: Albumin	92
303: Synaptojanin-1(172k)	29
304: Ig alpha chain C region (37k)	59
306: Ig gamma 2 chain C region (36k)	48
307: Ig gamma 2 chain C region (35k)	45

Figure 23. Protein Identities from AAL samples A1 and A2.
Protein hits with subsequent scores and molecular weight cut from the gel. The proteins were a result of MASCOT peptide database search using the spectra from LC-MS/MS run.

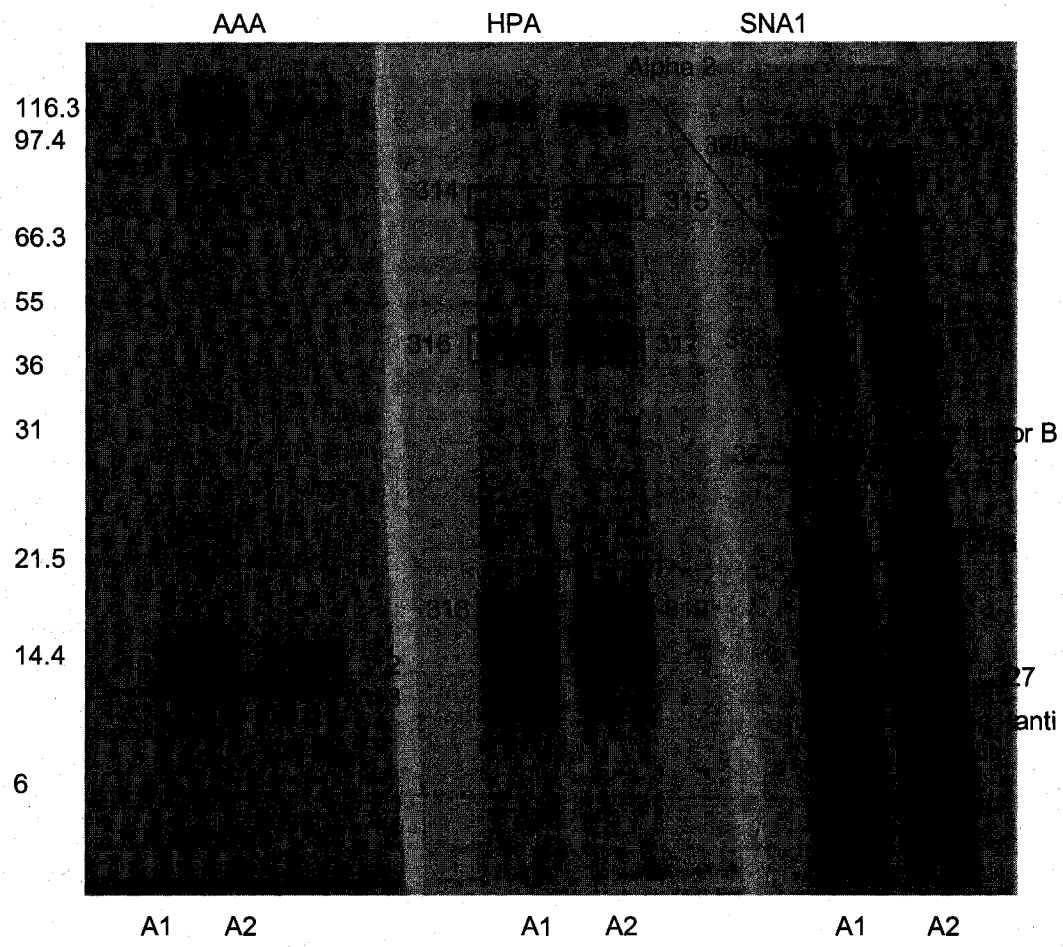


Figure 24. AAA, HPA and SNA1 capture of A1 and A2. 30 μl of each serum group was incubated with 100 μl of the appropriate bead slurry and incubated 16 hours. The samples were eluted and separated on 7.5% Criterion SDS-PAGE gels and visualized with silver stain. The gel slices are numbered with some of the protein identities labeled.

Protein ID results (significant hits)	Score
AAA	
310: 30s ribosomal protein	
311: keratin	311
312: 31 SNX 14 human sorting nexin 14	31
313: wall associated precursor	50
HPA	
314: keratin	65
315: Dna binding protein creA	38
316: Ig alpha 1 chain C region (37k)	71
317: NACHT containing protein (165k)	
318: Ig alpha 1 chain C region precursor	168
319: Ig alpha 1 chain C region (37k)	125
SNA1	
320: Apolipoprotein B-100 precursor 515k	316
321: Alpha 2 macroglobulin 163k	48
myotubularin related protein	41
322: kinase suppressor of ras2 93k	49
323:Complement factor B precursor 85k	65
324: Zinc finger protein 433 77k	29
325: Complement factor B precursor	90
326: Serotransferrin precursor 77k	356
327: Alpha 1 antitrypsin precursor 46k	286

Figure 25. Protein Identities from AAA, HPA and SNA Lectin Group A. Protein hits with subsequent scores and molecular weight cut from the gel along with sample group identified from. The possible proteins were a result of MASCOT peptide database search using the spectra from LC-MS/MS.

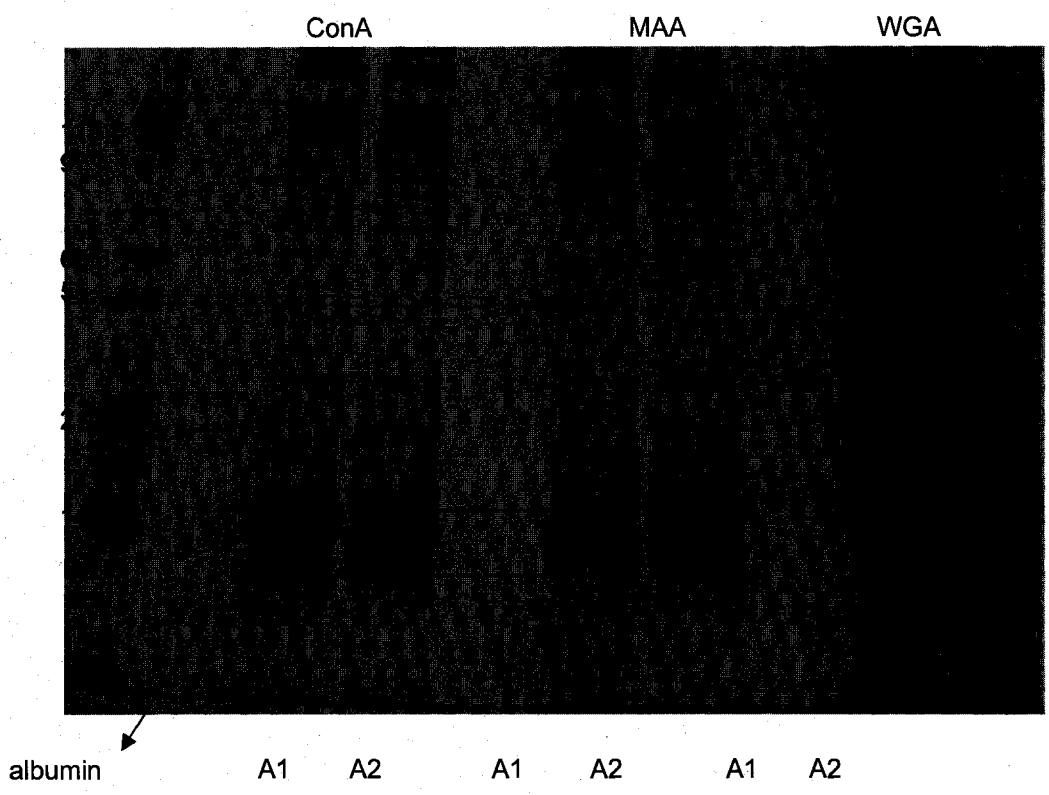


Figure 26. Con A, MAA and WGA Capture of A1 and A2.
30 μl of each serum group was incubated with 100 μl of bead slurry And incubated for 16 hours. The samples were eluted with appropriate elution buffers and separated on 7.5% SDS PAGE gels and visualized by silver stain. The gel slices are numbered with some of the protein Identities labeled.

from the lectin capture experiments were run on this instrument instead of the LCQ. Figure 27 is a gel picture of the sample groups A1 and A2 with the lectins AAA, AAL, ConA, HPA, MAA and SNA1. This gel was one of the very first gels trypsin digested and prepared for peptide mass fingerprinting using the Bruker instrument. Unfortunately, there are no protein identities, which correspond to this gel despite being fractionated and digested similar to previous gels run on the LCQ. The reason lies in the sensitivity of the MALDI instrument with regard to protein concentration on the gel. For example, the early gels were silver stained and subsequently identified with little problem. When this same strategy was applied to the MALDI the protein identification scores were too low to accept, which was the first indication that there would be many necessary optimization steps to reproducibly sequence proteins on this instrument. The first trouble shooting strategy we assessed was protein concentration, by switching from silver stain to Coomassie. Figure 28 is an example of a gel run with the same amount of protein which when silver stain was appropriate for identifications of the LCQ, yet when coomassie stained is barely visible. This finding prompted a switch from silver to coomassie, and subsequently an increase in the amount of protein loaded on all gels to be levels visible by Coomassie which for peptide mass fingerprinting experiments on the MALDI needed to be at least 50-100 ug, as compared to 20-50 for LCQ.

Figure 29 is a coomassie stained gel of group B with the lectins AAA, AAL and ConA. All proteins identified are labeled in the figure, which since this gel is coomassie stained there is not enough total protein loaded to visualize any proteins below 30 kD. At this point, any protein above or roughly 30 kD could be identified using the MALDI, whereas proteins below this molecular weight needed to be identified using the LC-MS/MS instrument. Figure 30 is an example of a peptide mass fingerprint spectra, which is sufficient to be matched to a protein in the database, since each peak has a high intensity and is labeled. This spectra was acquired from a standard for alpha-2-macroglobulin and acquired a mowse score of 300 for alpha-2. Numerous gels were run to identify as many proteins as possible; figures 31 and 32 are more examples, this time from group B and C. Figure 33 was purposely overloaded with 400 ug to try and identify the low molecular weight proteins at 30 kD and below and to search for low abundance proteins not

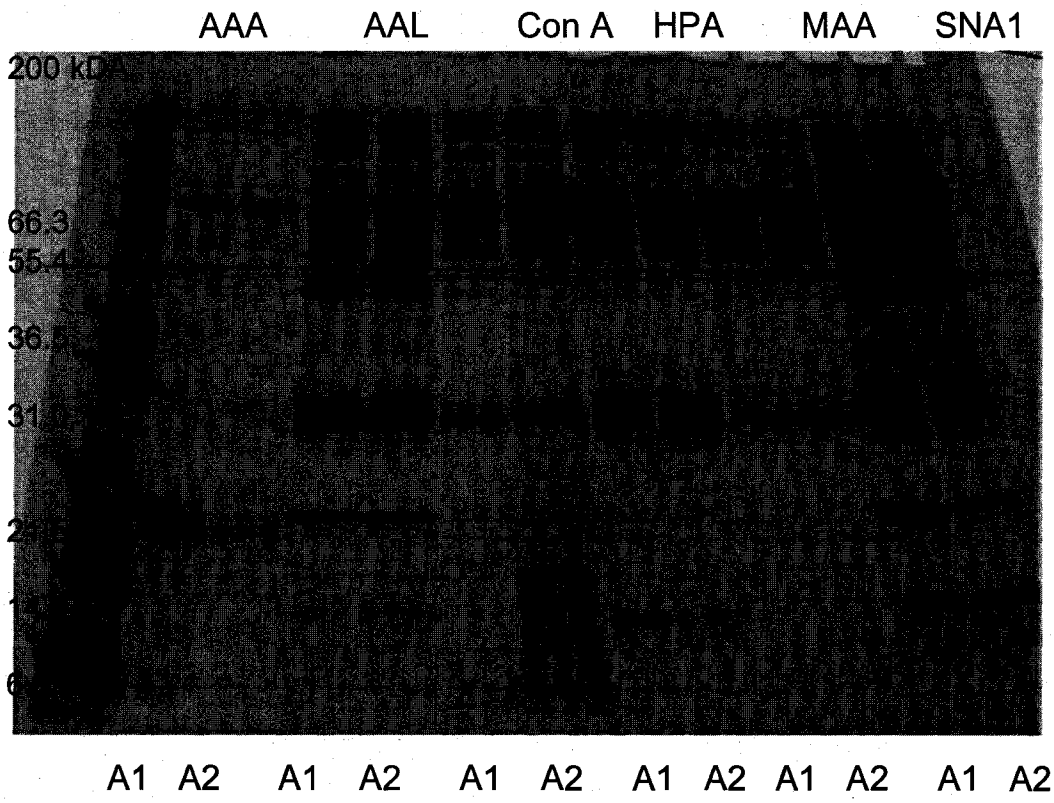


Figure 27. Differential Lectin capture of Group A.
30 µl of each serum group was incubated with 100 µl of each lectin bead slurry and incubated for 16 hours. The samples were eluted with appropriate elution buffers and separated on 8-16% Criterion SDS-PAGE gels and visualized by silver stain.

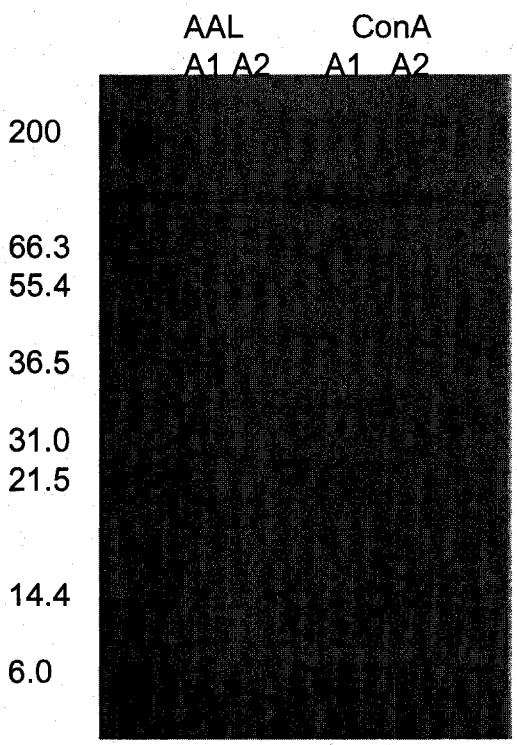


Figure 28. Coomassie Stain of Lectin Eluates. AAL and ConA. lectin incubations were carried out as previously described but stained with coomassie instead of silver. The lack of staining indicates the low amount of protein loaded on the gel.

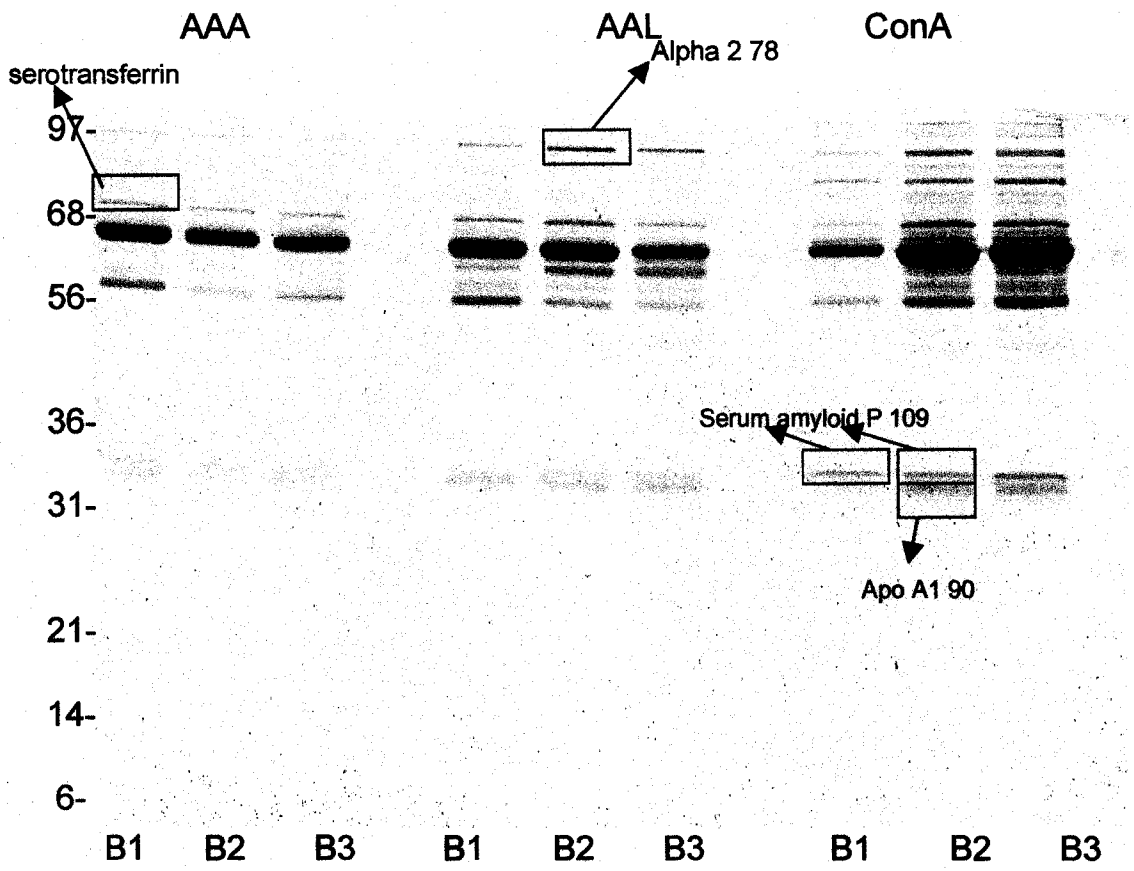


Figure 29. Lectin Capture with AAA, AAL and Con A of Group B. 50 μl of each serum group was incubated with 100 μl of each lectin bead slurry and incubated for 16 hours. The samples were eluted with appropriate elution buffers separated on 8-16% Criterion SDS-PAGE gels and visualized with coomassie. The subsequent protein identities from PMF are labeled.

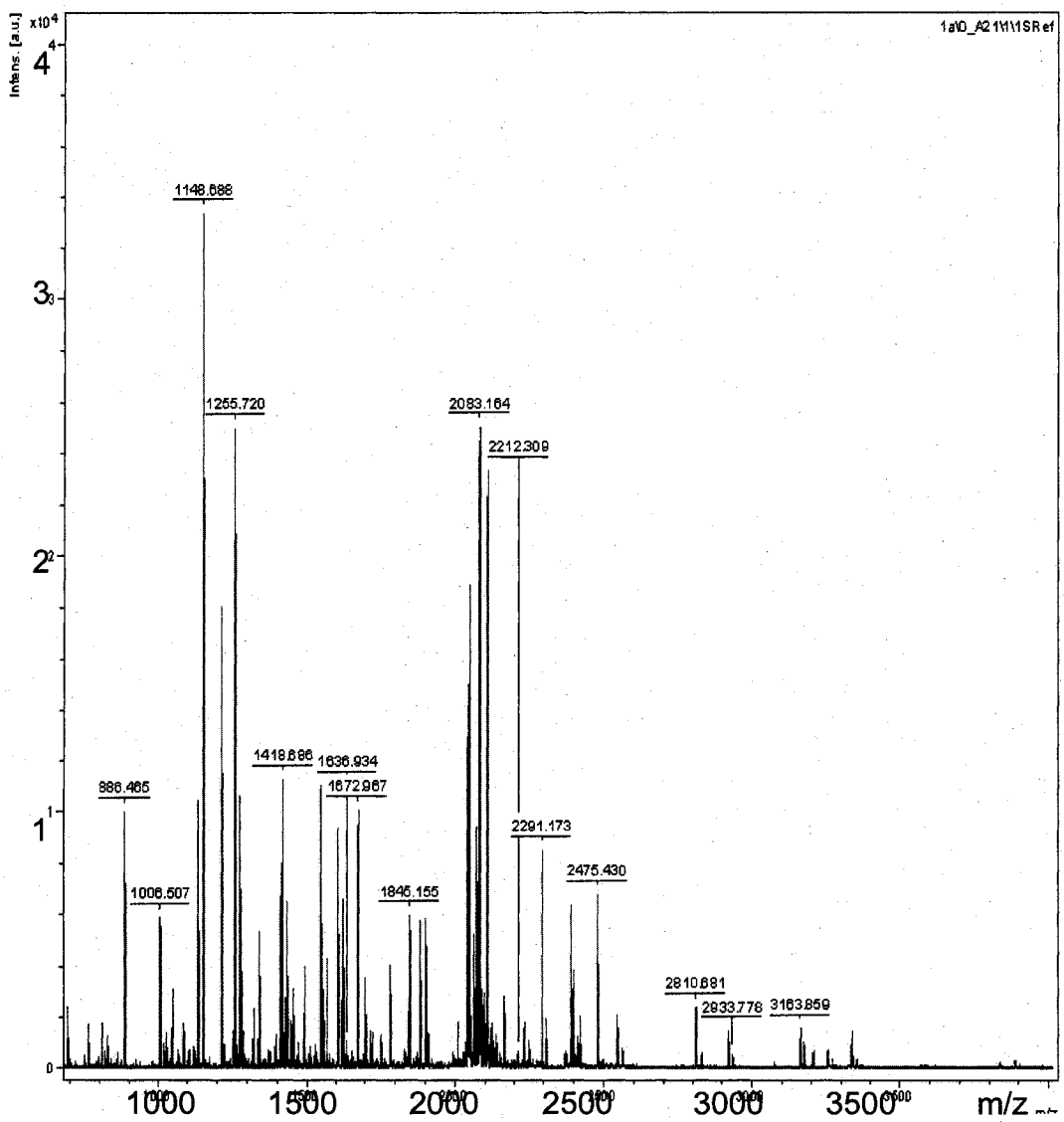


Figure 30. Representative Peptide Mass Fingerprint Spectra. Representative picture of a peptide mass fingerprint of alpha-2-macroglobulin, which yielded a score of 300. The spectra has a high intensity and the majority of peaks labeled which enables a good protein match.

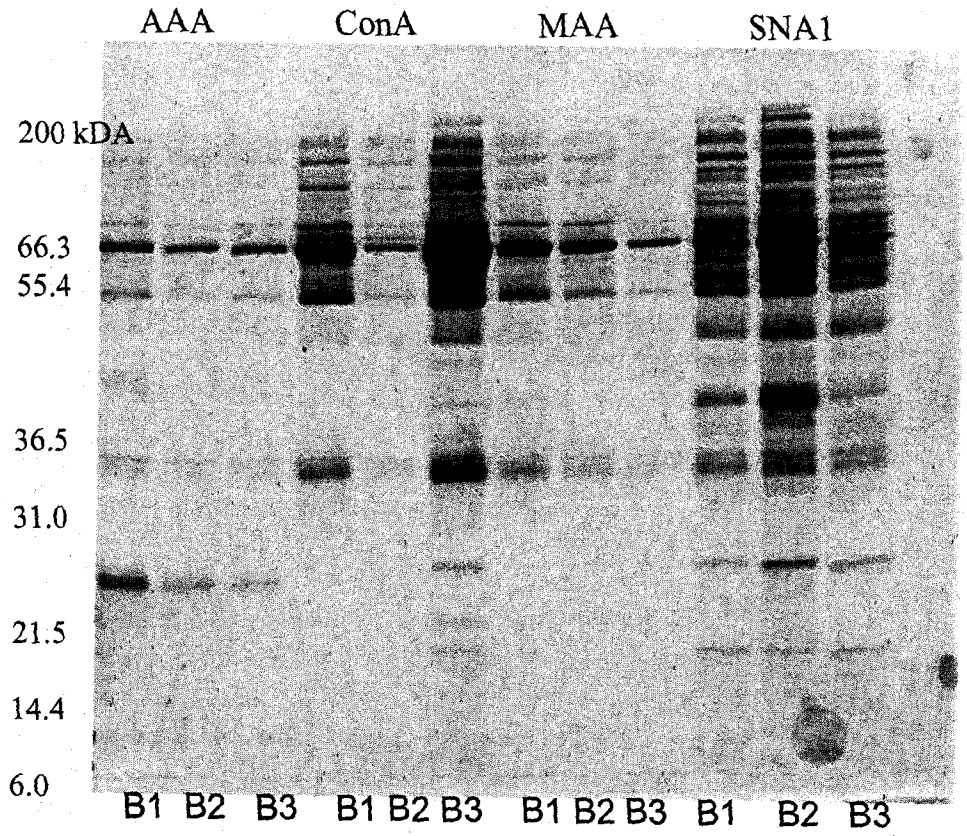


Figure 31. Lectin capture with AAA, Con A, MAA and SNA1 of Group B. 50 μl of each serum group was incubated with 100 μll of each lectin bead slurry and incubated for 16 hours. The samples were eluted with appropriate elution buffers and separated on 8-16% SDS gels visualized by coomassie.

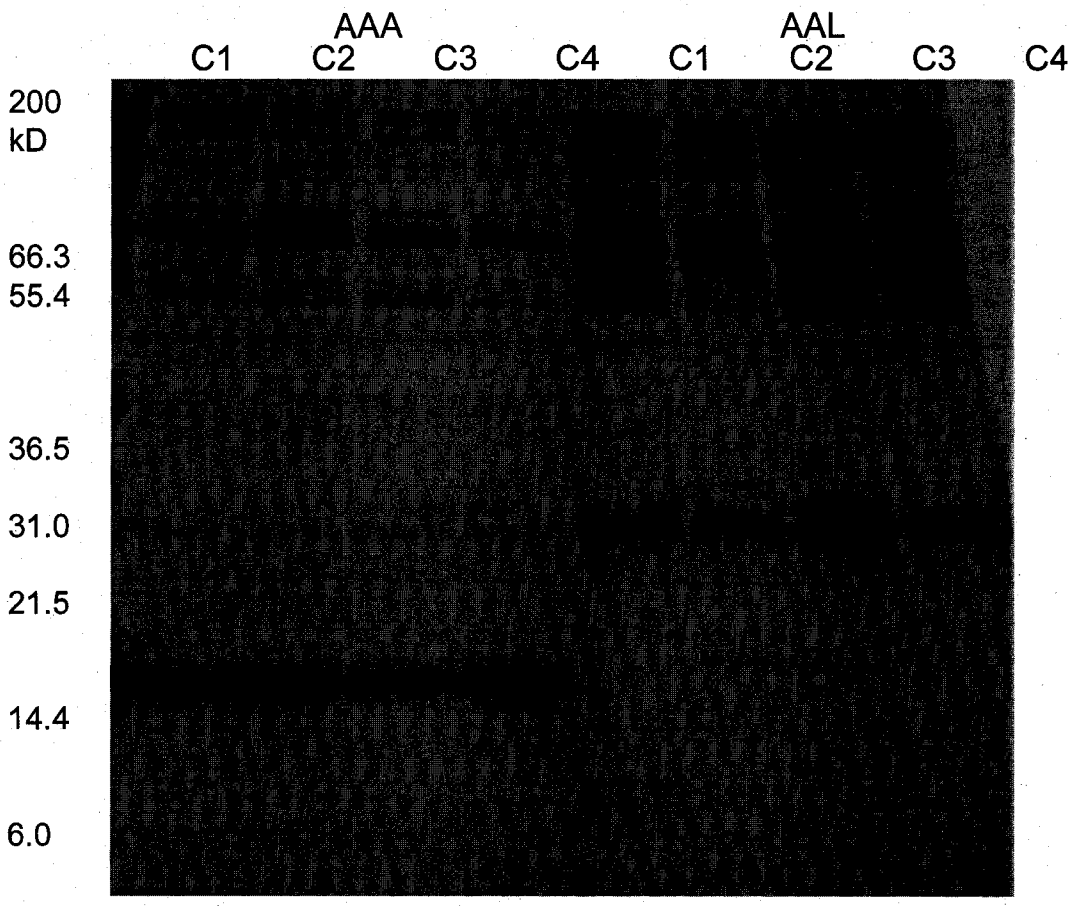


Figure 32. Lectin capture with AAA and AAL of Group C. 50 μl of each serum group was incubated with 100 μl of each lectin bead slurry for 16 hours. The samples were eluted with appropriate elution buffers and separated on 8-16% SDS-gels visualized with coomassie.

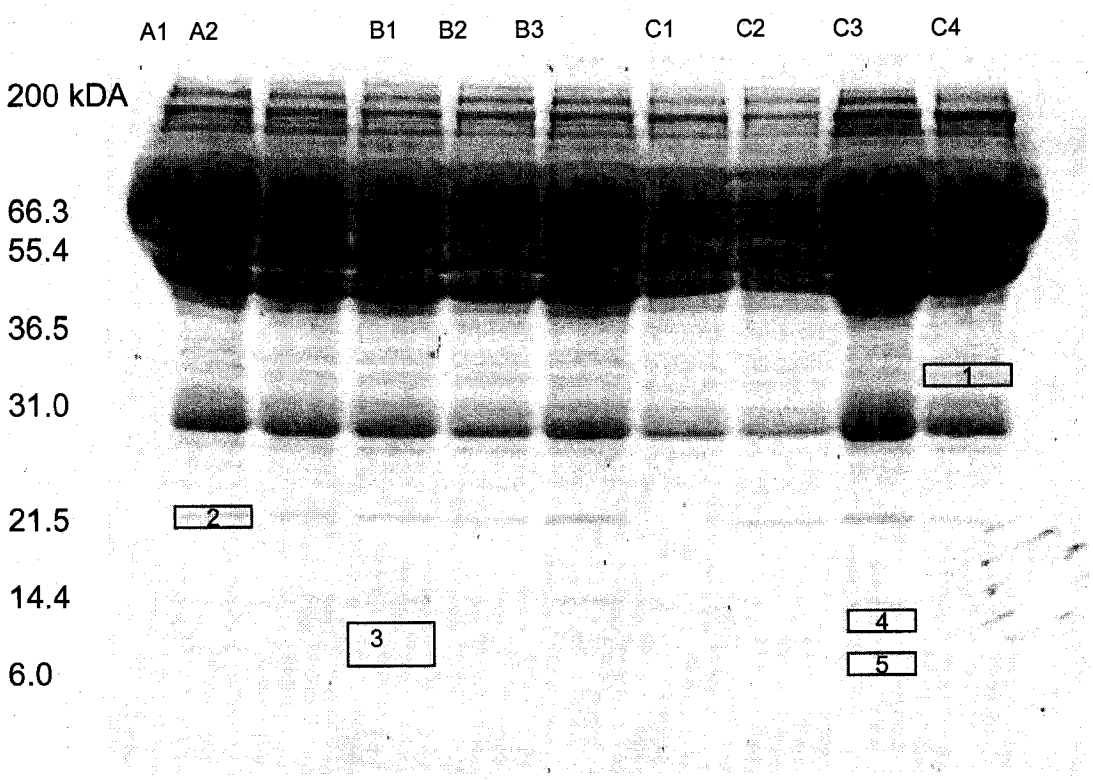


Figure 33. AAL Lectin Capture of all Groups 400 ug. To elucidate the low molecular weight proteins of 50 μl of serum was incubated with 150 μl of appropriate lectin and roughly 400 μg total protein was separated on 8-16% criterion SDS-PAGE gels. The numbers correspond to protein IDs listed in figure 38.

previously identified. Interestingly, we did identify a few haptoglobin fragments, which we had not seen previously. All of the proteins identified from each lectin capture and subsequent gel fractionation were compiled and listed in figure 34 corresponding to their appropriate lectin. The haptoglobin fragment was one potentially interesting identification, as well as AFP, which has been found fucosylated in other cancers as mentioned previously. Also, alpha-1-antitrypsin was identified in this study, which has recently been linked to pancreatic cancer (105). In the pancreatic cancer study a different isoform of alpha-1-antitrypsin was found to be specifically linked to metstatic cancer. Therefore, it will be necessary to assess whether the protein is differentially expressed or rather the glycan is changing due to prostate cancer progression. To determine whether these proteins have differential expression in relation to disease stage they were assessed by ELISA analysis.

ELISA Assay of Alpha-1-Antitrypsin

To determine whether alpha-1-antitrypsin is differentially expressed across the different prostate subgroups, we utilized ELISA assays. From the gel separation and identification experiments mentioned earlier, we identified AAT (alpha-1-antitrypsin) from more advanced stages of prostate cancers, therefore we utilized group C of our prostate serum cohort shown in Figure 15. The ELISA data can be seen in Figure 35. Interestingly, the expression of AAT is increased in WGA and SNA captured serum from prostate cancer group C3 (low gleason cancer) compared to the other stages C1, C2 or C4. The expression in WGA C3 is higher then SNA C3 which is most likely due to the nature of lectin specificity of each lectin. For example WGA binds sialic acids linked in every configuration while, SNA1 only recognizes sialic acids linked in the α 2,6 configuration. In each case the samples can be compared to a quality control serum (QC) sample which is a pool of normal serum which is less then each disease case. Con A eluates were also assessed for the level of AAT expression. Interestingly, the same pattern shown for WGA and SNA holds true, though the level of C3 is lower here. This data suggests that AAT is potentially differentially expressed in group C3 of the prostate serum samples, and will need to be further evaluated for its role as a biomarker of prostate cancer. ELISA assays for haptoglobin and

AAL
Albumin
a-2-macroglobulin
Immunoglobulin a Chain C
Serotransferrin precursor
AFP - alpha fetoprotein
Haptoglobin*
Complement C-3 precursor
Protocadherin precursor

Con A
Albumin
a-2-macroglobulin
ApoA-1
Serum amyloid P chain
Serotransferrin precursor

WGA
Apolipoprotein B-100 precursor
Apolipoprotein A-IV precursor
Apolipoprotein A-1 precursor
Alpha-1-antitrypsin*
Alpha-2-macroglobulin
Alpha-1-acid glycoprotein 2 precursor
Complement C1q component
Ig lambda chain C region

AAA
Albumin
ApoA-1
Apo C-III
AFP - alpha fetoprotein
Haptoglobin
Myotubularin protein
Transthyretin precursor*

HPA
Apolipoprotein C-III precursor
Haptoglobin precursor *
Transthyretin*
Haptoglobin*

SNA1
a-1-antitrypsin
ApoA-1
Complement factor B precursor
Myotubularin related protein
Transferrin

*Potentially differentially expressed in clinically stratified prostate serum groups, will be confirmed by ELISA.

Figure 34. List of all Proteins Identified . Comprehensive list of proteins identified from lectin capture.The asterick indicates potentially differentially expressed proteins across sample groups.

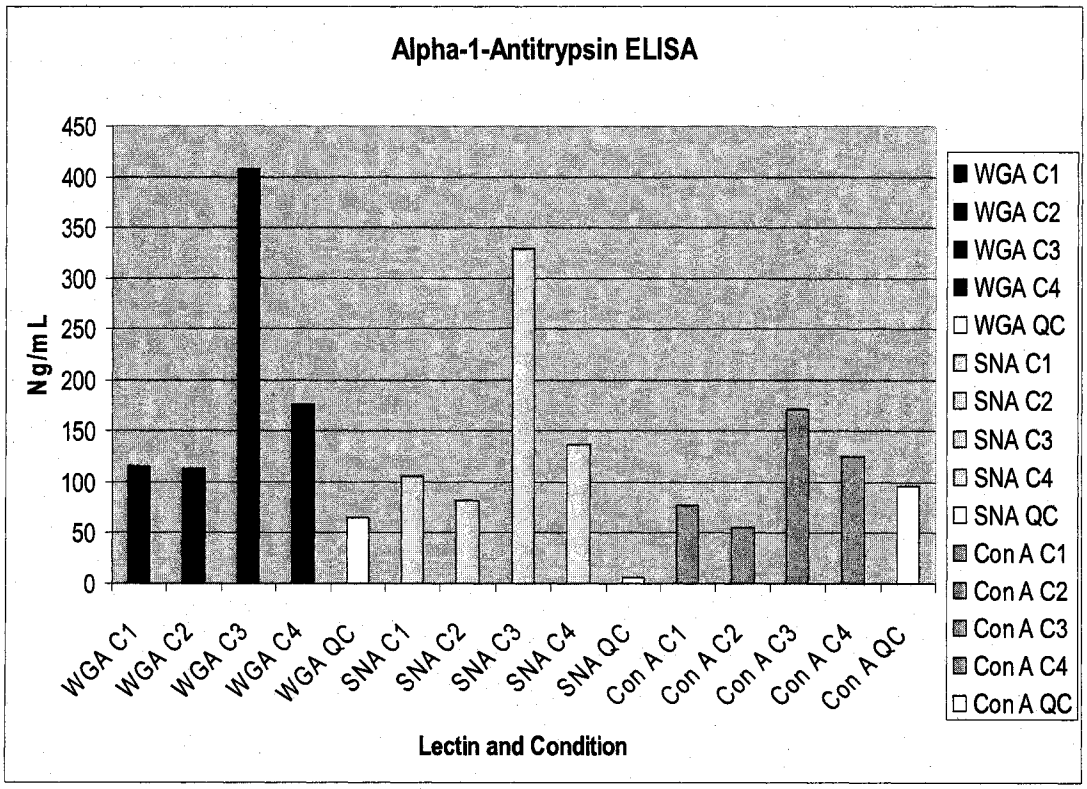


Figure 35. Alpha-1-Antitrypsin ELISA. Lectin eluates were run on A1AT Genway (San Diego, CA) ELISA to determine level of expression for A1AT protein when captured using WGA, SNA and Con A lectins. QC (quality control) is a pool of normal serum used here as a control.

AFP were also carried out simultaneous to the AAT experiment. Initial results suggests similar patterns as indicated with AAT, though they will need further optimization to confirm the findings.

In Solution Trypsin Digest To Visualize High Molecular Weight Proteins

To determine whether high molecular weight protein complexes or proteins captured using the lectin fractionation approach are being missed by separation with 1D PAGE we next performed a total in solution trypsin digest of the lectin eluates without gel separation. Roughly 100 ul of lectin eluate was incubated with trypsin (1:20) cleaned up with IMAC-Cu beads and spotted on a MALDI Anchorplate and read in reflectron mode. As shown in Figure 36, the digested spectra clearly have more peaks then the undigested spectra. The gel picture also shows a difference in the samples digested with trypsin vs. those without, since the digested samples show presence of many more low mass proteins then the undigested eluates. This demonstrates the potential of using this approach as an alternative strategy to visualize protein complexes. An analogous approach with the types of proteins identified will be discussed in the next chapter.

Glycan Analysis

Elucidation of the N-linked oligosaccharide structures of glycoproteins has relied on release of glycans using the enzyme PNGase F (peptide N-glycosidase F) followed by glycan purification, end labeling with fluorescent groups, and analysis using HPLC, mass spectrometry or nuclear magnetic resonance. PNGase F is an amidase that cleaves between the innermost GlcNAc and asparagine residues of high mannose, hybrid, and complex oligosaccharides from N-linked glycoproteins shown in Figure 37. Since the oligosaccharide chains may undergo alterations due to disease, which is not visible by 1D gel analysis alone, we were very interested in developing amenable protocols to perform glycan sequencing on the Bruker MALDI-TOF/TOF. There are many available forms of the enzyme PNGase F and like any enzyme it has varying degrees of processivity Therefore to assess its activity we initially performed N-terminus glycan cleavage on serum samples following SNA lectin capture. Figure 38 is a gel view representation of this experiment. The prominent band at 31kd (arrow 4) is the enzyme, yet in the PNGase F treated samples there is a recognizable decrease in protein mobilities corresponding to cleavage

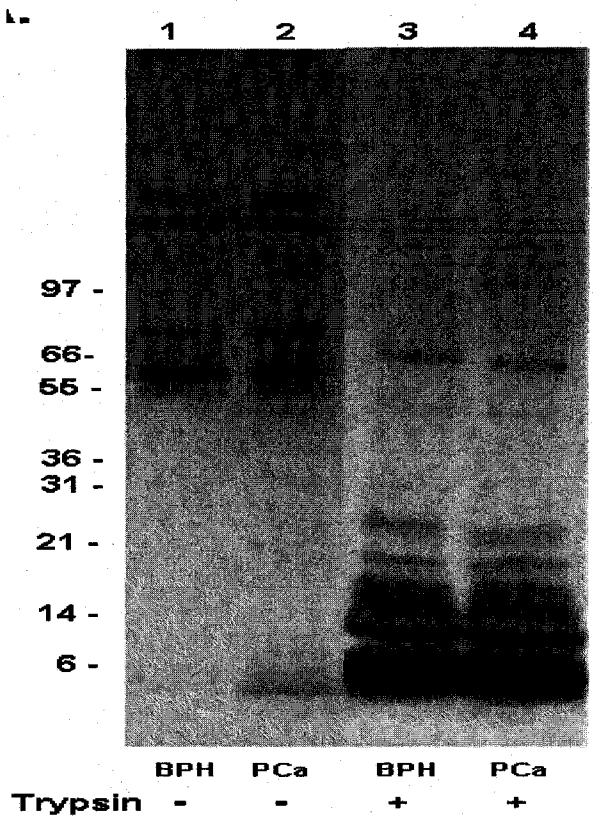


Figure 36. Con A Bound Serum Glycoproteins Digested with Trypsin prior to MALDI-TOF. Pooled serum from BPH and prostate cancer sera were bound to Con A lectin then eluted. A portion of each eluate was separated on 8-16% SDS gel. (A) ether intact (lanes 1, 2) or digested overnight with trypsin (lanes 3, 4).

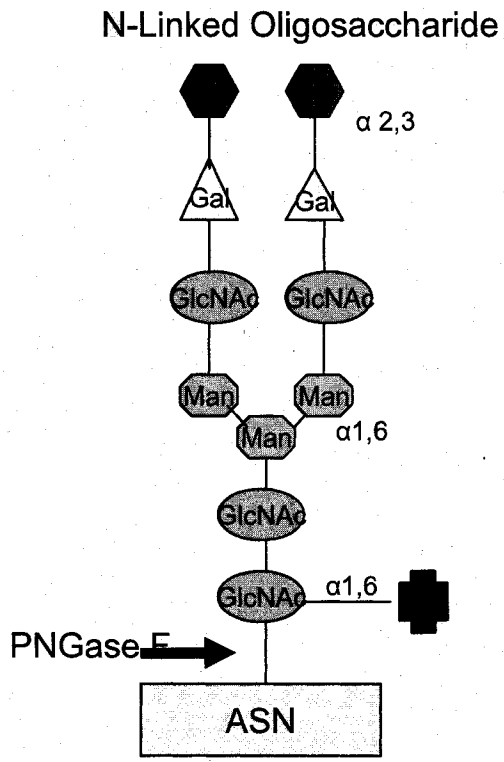


Figure 37. PNGase F Activity.
PNGase F cleaves between the Asn-R group on the protein and the first GlcNAc residue of the oligosaccharide. Following cleavage the Asn residue is converted to Asp.

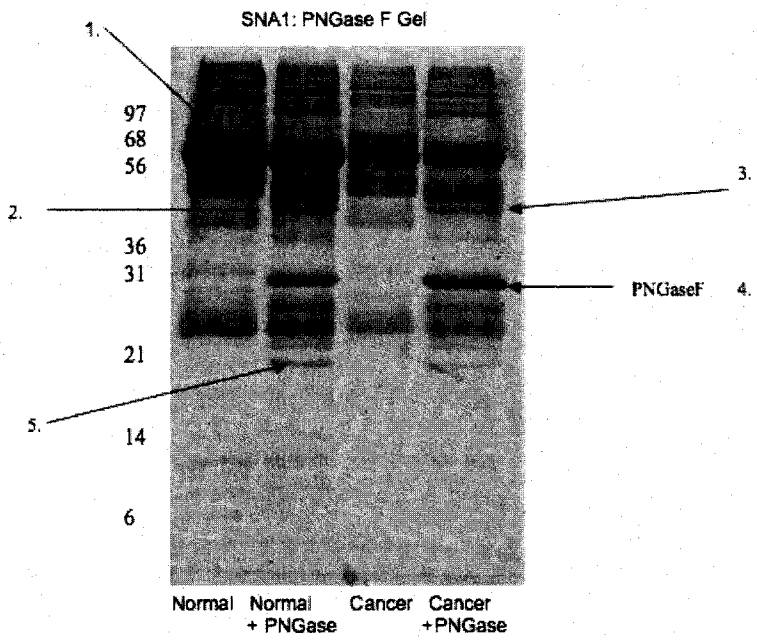


Figure 38. PNGase F Cleavage of N-linked Glycans. Prostate serum pools of normal or cancer were captured by SNA1 lectin and then either incubated with PNGase F or mock for 16 hours. The arrows indicate protein bands with increased mobilities following Pngase F treatment. (1,2,3,5) This is a representative gel view run on an 8-16% Critierion SDS gel visualized with coomassie.

of glycans indicated by arrows (1,2,3,5). For example the band in the undigested sample at molecular weight 68 is no longer there in the digested sample, since it has migrated to a lower position. This again is evident in the band at MW 40 in the undigested sample becomes a doublet in the digested lane running at 38 and 26kd. To determine whether PNGase F cleavage and glycan capture would effectively release the glyans on putative glycoprotein biomarkers, prostatic acid phosphatase (PAP) was utilized in a proof of principle experiment.

PAP is an enzyme secreted by the epithelial cells of the prostate into seminal fluids, and determination of PAP levels has previously been used as a biological tumor marker of prostate cancers (151, 152). Elucidation of the crystal structure of PAP has found 3 potential glycosylation sites occurring at Asn62, Asn 188 and Asn 301. At Asn 62 and Asn 301 high mannose oligosaccharide structures were determined, while at Asn 188, a complex glycan structure was present. (151) In addition, there is evidence that PAP may have differential glycosylation in relation to cancer, specifically having a decrease in the branching of the high mannose chains of Asn 62 and Asn 301 (152). Because PAP is a clinically significant marker of prostate cancer, and it is highly abundant in seminal fluids we decided to use PAP in our initial approach. PAP is a homodimer with a molecular weight of 109,000, and when separated in SDS-polyacrylamide gels runs as a 54,000 dalton monomer. Western blot analysis of human PAP from seminal fluid samples obtained from men with prostate cancer and from healthy normals, found PAP to run at 50 kD as shown in Figure 39. Figure 40 is a 1D PAGE separation of the same seminal plasma samples used in the western blot analysis. To be sure that the band we believed to be PAP was indeed the protein, the 54 kD band in the first two lanes were excised, digested with trypsin and sequence determined using peptide mass fingerprinting on the MALDI-TOF/TOF. The two bands were identified as PAP with mowse scores over 100, which encouraged us to proceed with the planned glycan structural analysis.

From lane three of the same gel (Figure 40), the PAP band was excised, digested with PNGase F, and the glycans isolated as described in the methods section. The isolated glycans were spotted 1:2 with DHB matrix and analyzed on the MALDI-TOF. Figure 41 is the representative spectra, which was acquired from the PAP glycan sequencing experiment. A is

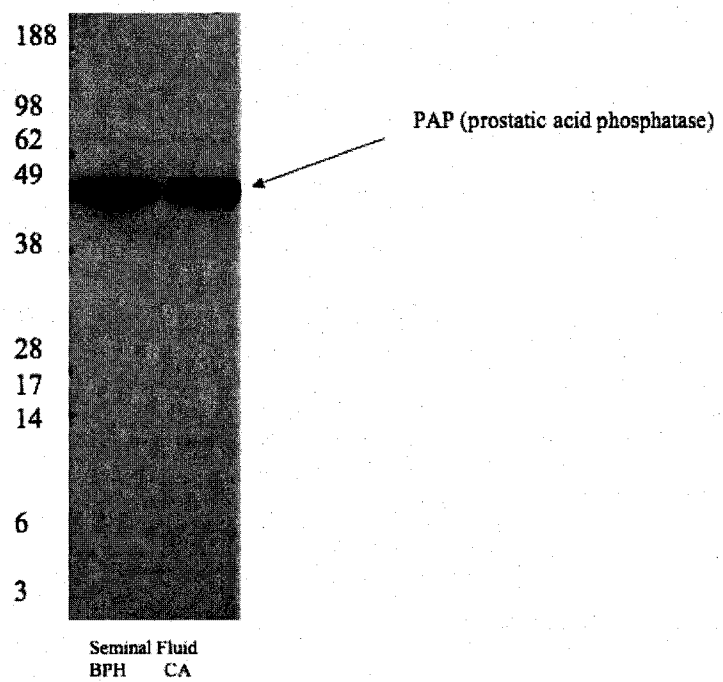


Figure 39. Western Blot Analysis of PAP from Seminal Fluid. Roughly 30 ug of seminal fluids was separated on Novex Nu-Page mini gels and blotted for PAP at a 1:2000 concentration. Lanes 1 and 2 are from healthy men.

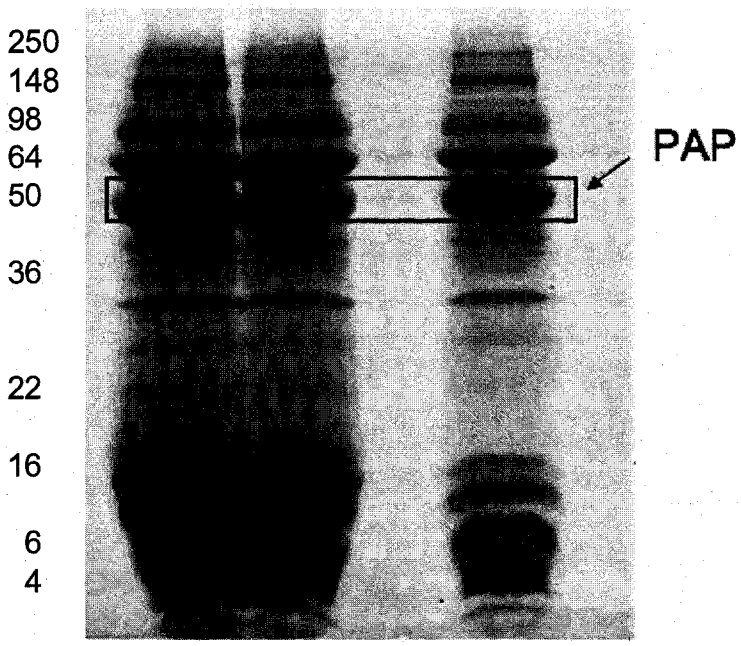


Figure 40. PAP Separation Prior to Glycan Analysis. Roughly 40 ug of seminal fluid from a healthy male was separated using an 8-16% Criterion SDS-PAGE gel and visualized with coomassie stain. Lanes 1 and 2 were verified by PMF on the MALDI and lane 3 was used for glycan analysis.

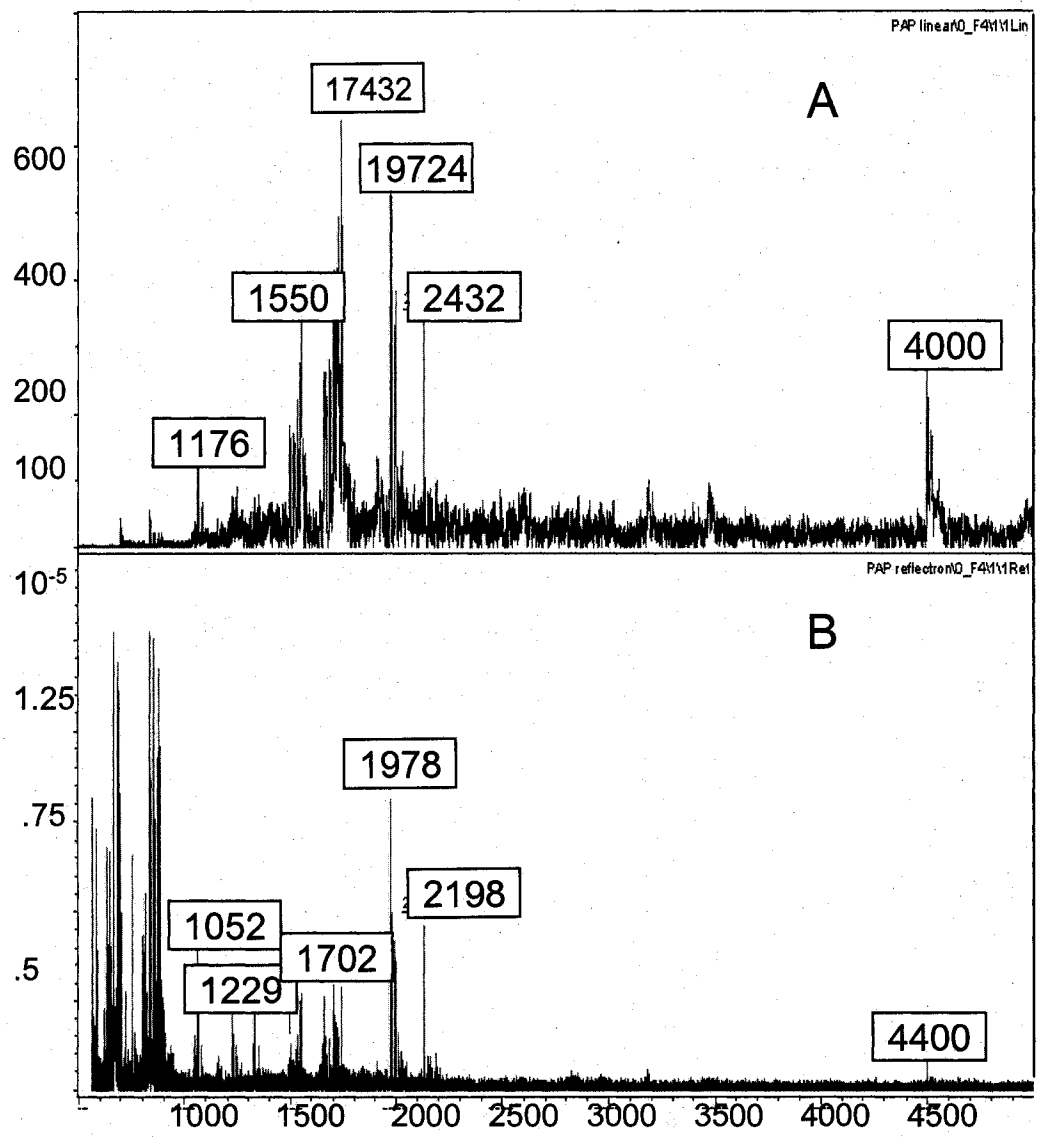


Figure 41. MALDI spectra of PAP Glycans.
Glycans were captured from PAP and spotted 1:2 with saturated DHB matrix and run in both linear (A) and reflectron (B).
The glycans are labeled according to their m/z or mass to charge ratios.

the linear read while B is the reflectron read. Since there are no comprehensive MS database for glycan masses determined on MALDI platforms, it was necessary to manually assign putative glycan structures based on the determined masses, the known masses of constituent sugars, and standard reference glycans. For example, a 2-GlcNAc, 8-mannose glycan structure having a mass of 1743 was provided in a Sigma Chemical product literature (Figure 42), which can also be seen in the PAP glycan spectra shown in Figure 43. This high mannose structure is consistent with those described for the crystal structure of PAP at Asn 62 and Asn 301(141). In addition, in order to determine which carbohydrates are making the given peaks, high energy fragmentation or collision induced dissociation is necessary to fragment the parent ion or in this case the 1743 peak.

IgG glycan analysis

Immunoglobulins represent a reservoir of potential different oligosaccharide biomarkers since they play such a large role in the body's defense to disease. In addition, our collaborators have been looking specifically at the oligosaccharide characterization of immunoglobulin classes in relation to cancer progression (37). Serum IgG is known to have an N linked glycosylation site at Asn 297 of the CH2 domain (66) . The IgG heavy chain runs at around 50 kD. IgG was separated from normal and cancer seminal plasma and run out on 1D PAGE as shown in Figure 44. To follow up on this data, IgG capture and analysis was performed as described for PAP above to determine the identities of the IgG glycans from normal or cancer samples, the glycans captured from IgG are shown in Figure 45.

CID fragmentation of glycan standards

Since we did not have any annotating software available for this experiment, we decided to run perform CID analysis on some glycan standards, which we had purchased from Prozyme. The standard shown in Figure 46 is an M3N2 or three mannose, two GlcNAc glycan. From the initial MS spectrum the intact glycan is represented by the peak at m/z 933. Upon CID of two parent ions initial scan 609 (middle spectrum) and 933 (bottom spectrum) the fragments or subsequent components of the 933 peak become apparent. For example the peak at 226 could correspond to a GlcNAc molecule, which is detaching from the core upon fragmentation. Clearly,

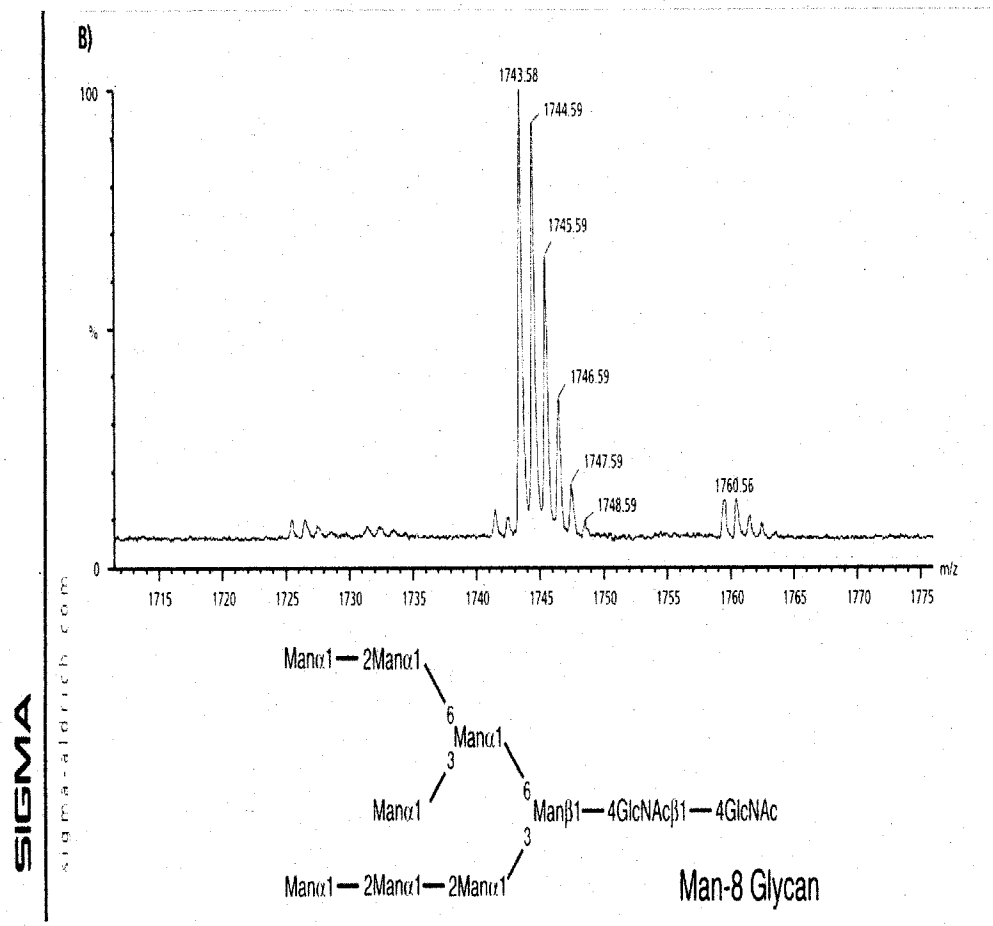


Figure 42. Sigma GlcNAc2 Man 8 glycan. A high mannose 8 glycan (GlcNAc2 Man8) which is similar to the PAP glycan found in figure 43.

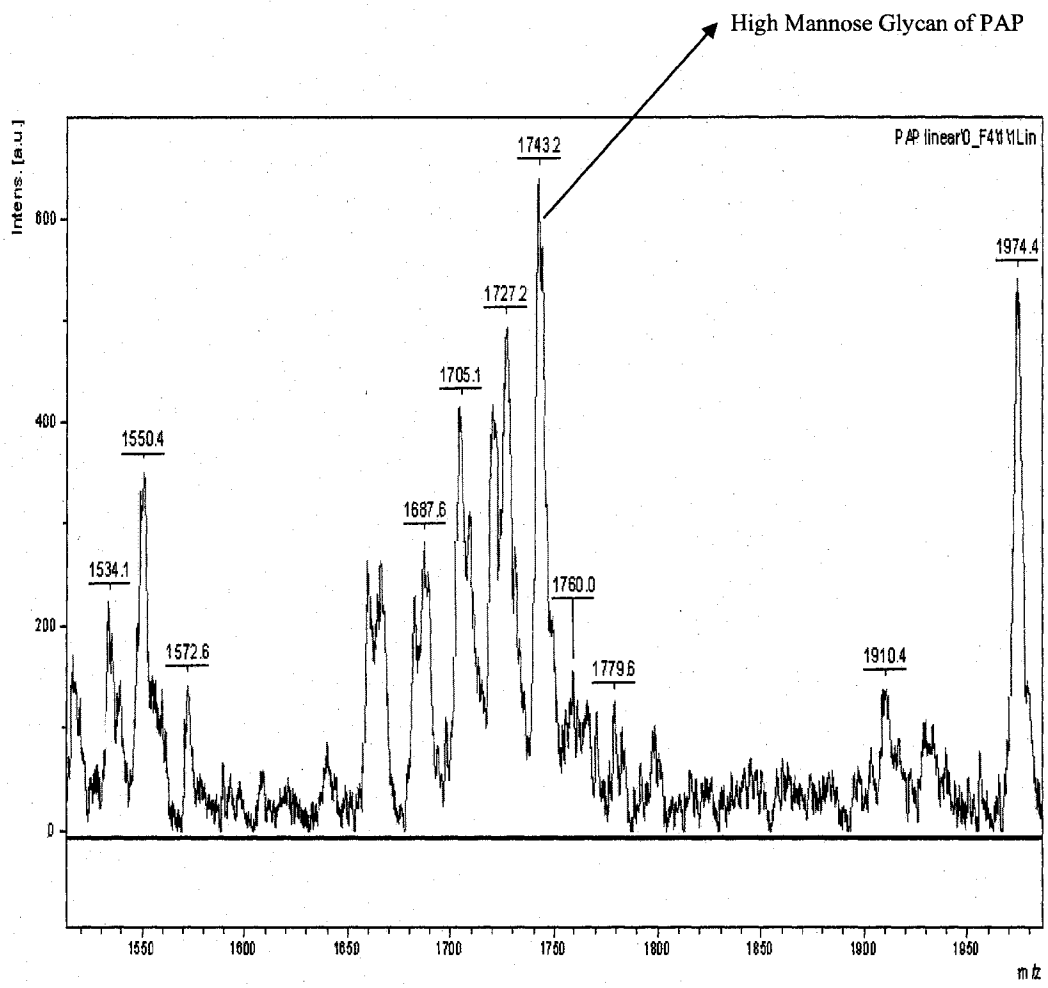


Figure 43. PAP GlcNAc2 Man 8 Glycan. Peak 17432 is representative of the sigma high mannose glycan in figure 42.

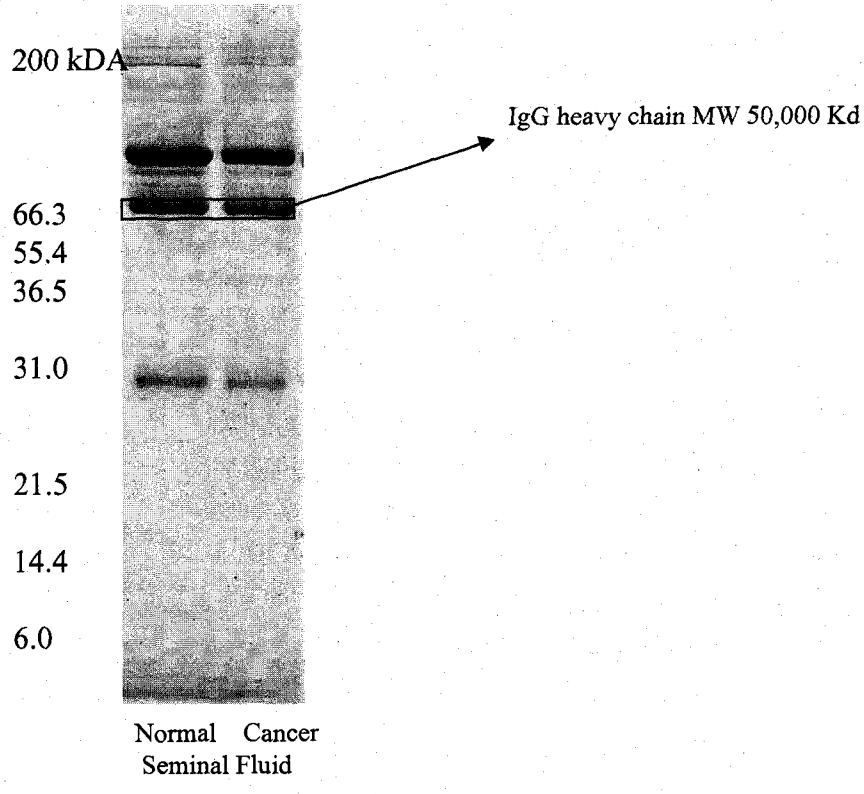


Figure 44. 1D Gel Separation of IgG Heavy Chain.
IgG was isolated from human seminal fluid of normal and cancer patients and separated using Criterion SDS-PAGE 8-16% gels visualized with coomassie stain.

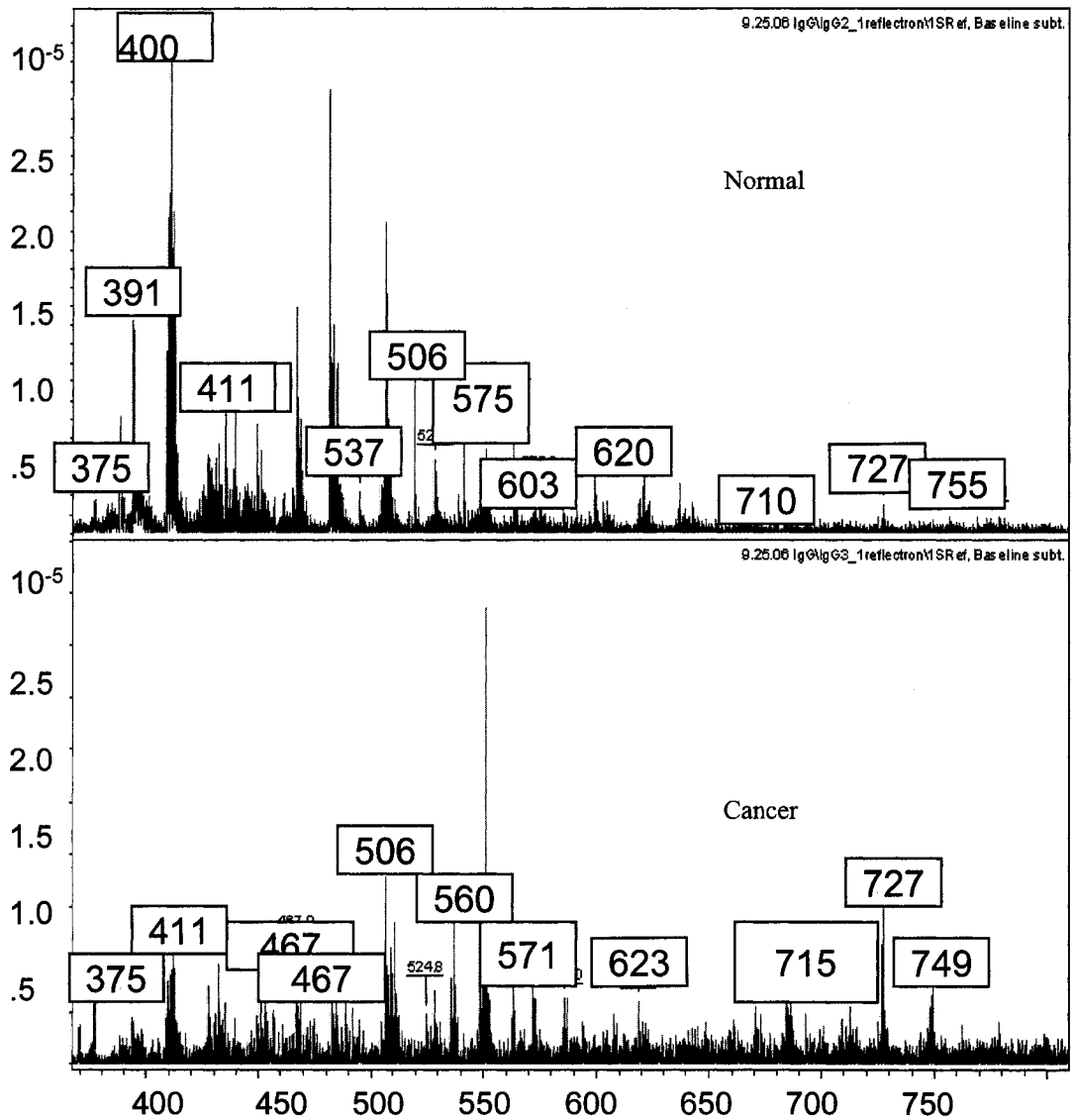


Figure 45. MALDI Spectra of IgG Glycans.
Glycans were captured from IgG and spotted 1:2 with saturated DHB matrix and run from a normal or prostate cancer patient (bottom). The glycans are labeled according to their m/z or mass to charge ratios.

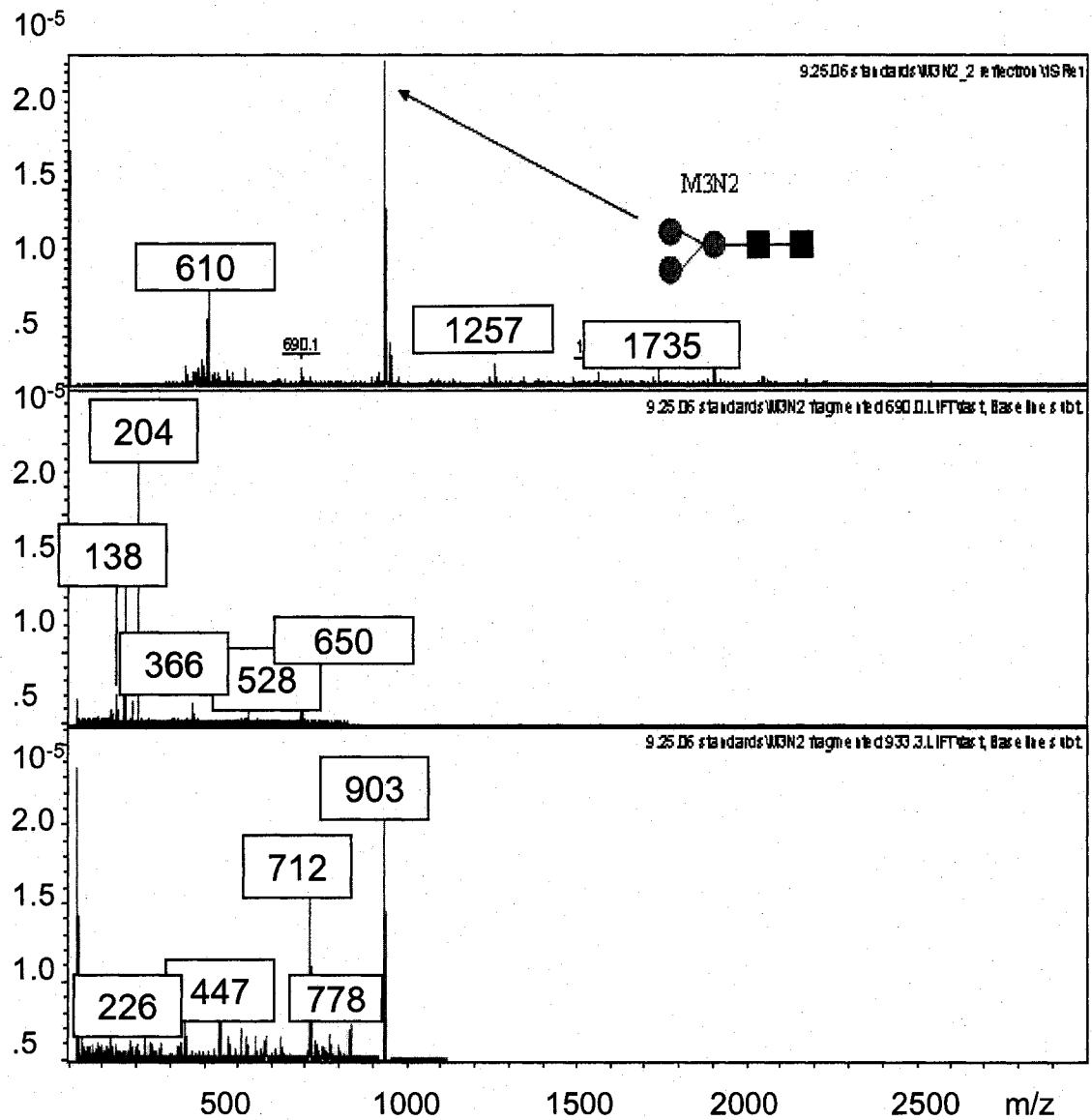


Figure 46. Collision Induced Dissociation of M3N2 Glycan. Using CID or collision induced dissociation the core M3N2 structure can be broken up into its subsequent sugar molecules.

software capable of annotating these structures will be necessary for future experiments, though it is promising that this method is sufficient to allow this type of analysis on the MALDI.

3.4 Discussion

Chapter III highlights one strategy for the characterization of glycoproteins within a clinically defined serum cohort composed of varying stages of prostatic disease. The approach shown here uses an upfront lectin capture of serum glycoproteins coupled to a mass spectrometer for both protein and glycan identification. By incorporating various lectins each with a specific oligosaccharide binding partner, we were able to capture many classes of glycoproteins. The lectin panel utilized here was carefully chosen based on current literature, (75, 79-82, 88, 153-156), and previous research. For example, our collaborators on the HCC SELDI studies have recently found an increase in fucosylation of serum glycoproteins in relation to HCC progression. (30, 37) Therefore, AAA (Anguilla anguilla) and AAL (Alueria aurantia) were chosen. In addition, it is known that sialic acids are important regulators of cellular and molecular interactions since they can mask recognition sites or serve as recognition determinants. Therefore the lectins SNA and WGA were used which recognizes alpha 2,6 linked sialic acids (SNA) and sialic acids in any linkage (WGA) (24). By cataloguing all of the proteins identified we are able to gain an idea about the post-translational modifications occurring on some of the captured proteins. For example, alpha-1-antitrypsin was captured from SNA and WGA which both recognize sialic acid residues. This is very interesting considering a recent publication, which found A1AT to have an isotype change in relation to pancreatic cancer (105). Con A recognizes alpha mannoses, which will be very prevalent in all N-linked oligosaccharides and therefore select a large number of glycoproteins. In addition, HPA which recognizes GlcNAc,, GalNAc and sialic acid residues, has been used to identify the presence of glycans linked to aggressive types of breast cancer, and thus has a direct biological role linked to metastasis, (80). The variation in specificity of the lectin panel can be visualized by gel analysis, which was shown in figure 18. Clearly WGA and Con A capture the most proteins while AAA and AAL capture far fewer proteins see Table 7.

The use of a clinically defined prostate serum cohort of varying disease stages is a novel aspect of this study since most other glycoprotein studies have involved fewer sample numbers

and focused more on method development. In contrast to these technical studies, we wanted to determine whether we could capture glycoproteins indicative of disease stage, and assess whether this strategy could allow for the identification of potential biomarkers for the early detection prostate cancers. In addition, the primary means of separation utilized herein is gel electrophoresis unique to our study design. One initial reason to run the gel step was to visualize patterns between our groups, since we were more interested in changes between groups than total proteins identified. This gel separation step also facilitates rapid identification and cataloguing of all the proteins captured. A complete list of all of the proteins identified by this approach was shown in Figure 34. Of these glycoproteins AFP, haptoglobin and A1AT showed patterns which suggested they were differentially expressed across the prostate serum groups.

To determine whether this was occurring in our sample set, ELISA analysis was performed. The A1AT ELISA results are shown in Figure 35. Interestingly, across the group C prostate serum samples the levels of A1AT levels are higher in the WGA and SNA captured samples, and in Con A eluates as well. These results indicate that A1AT protein levels may be overexpressed in low gleason scored cancers as compared to the other conditions. It is also apparent that the WGA eluate C3 is higher then the SNA captured C3 that is due to the specificity of the lectins. WGA will bind sialic acids linked in every configuration, while the SNA lectin binds only sialic acids linked in a 2,6 configuration. In addition, it is possible that there may be more A1AT with sialic acids in the terminal position then Con A which could account for the difference across the lectins. These results are interesting in regards to prostate cancer, since recently an isoform of A1AT with differential sialic acids was identified to be a marker of pancreatic cancer (105). There have been several recent studies highlighting the potential roles of A1AT in cancer progression which further validate our findings.

For example, in a recent study conducted by Huang et al A1AT was found to be secreted from tumor masses (157). Huang et. al used mass spectrometry to identify the in vivo secretome from a skin tumor mass injected into a mouse model. Using this approach they could identify the secretome from both progressive and regressive tumors (157). Five secreted proteins including A1AT were found from tumor masses at regressive stages. Interestingly, these five proteins three

were protease inhibitors (157). Though this is an animal model and would need to be re-assessed in a human tumor model, it could provide clues to the overexpression of A1AT in group C3 (low gleason score cancer) as opposed to the high gleason score group C4. Low gleason score prostate tumor tissues resemble more normal prostate tissue and are less likely to spread then high grade gleason scored tumors. Additional studies would be needed to determine if A1AT has implications on tumor regression, and to determine the number of tumors expressing A1AT which do regress.

In another study by Petrache et al, A1AT was found to inhibit caspase 3 activity in the lung, thereby preventing lung endothelial cell apoptosis (158). Native A1AT could prevent apoptosis by directly inhibiting caspase 3. This finding provides a direct relationship of an anti-apoptotic mechanism guiding disease progression and is relevant to disease processes characterized by structural cell apoptosis, oxidative stress and inflammation (158). Of particular interest is the ability of A1AT to be a marker of inflammation which is a well documented phenomenon in cancer patients, in prostate cancer patients (158, 159). In a recent prostate cancer study, a direct correlation between serum prostate specific antigen and A1AT was found. In this study the levels of acute phase proteins such as alpha 1 acid glycoprotein, C reactive protein and A1AT were measured in men without prostate cancer, and those with newly diagnosed prostate cancer. A1AT levels correlated with PSA levels and were high in men with prostate cancers as opposed to those without prostate cancer (159). Taken together this could suggest multiple roles of A1AT in prostate cancer progression. It is possible that A1AT could be directly linked to the initial immune responses to cancer, or the inflammation stages. In this way, A1AT could be secreted by the stromal cells surrounding the tumor in the initial stages of tumor growth and act as an inhibitor of apoptosis to allow for continued cancer growth. Yet A1AT could also be differentially glycosylated in the microenvironment of the tumor and regulated by the stromal cells to alter signaling cascades to promote increased growth receptor pathways. The possibilities of the involvement of A1AT in cancer progression are numerous and will therefore need to be evaluated further. One area which could greatly help determine A1AT function here would be the characterization of the expression levels of A1AT in prostate tumors as well as the glycans

attached to the protein. It is possible that characterization of the isoforms of A1AT could also provide information helpful for cancer detection, which has recently been shown to be directly related to disease progression in a handful of cancers. For example, a fucosylated version of AFP is well described in sera associated with liver cancers (95, 97). Identification of an AFP variant in prostate cancer sera is a novel finding, while identification of the haptoglobin and ApoA-1 variants are also consistent with previous expression profiling studies of cancer sera (9, 24, 26, 30).

Though we believe the methodology here is an attractive prostate cancer biomarker discovery approach, other lectin capture approaches for the characterization of the serum glycoproteome have been reported (105, 115, 149). For example Qiu et. al performed total trypsin digestion of the serum first, and then used differential isotope tagging of the disease and normal samples prior to glycopeptide capture with a sialic acid binding lectin. In this way, glycopeptides can be captured and quantitated, yet within this study they use serum purchased from a company, which has minimal clinical relevance. However, many of the proteins identified from this study overlap with our protein list. In another study by Zhou et al, a multi-lectin affinity column which combines the lectins ConA, WGA and Jacalin was used to allow capture of a larger subset of proteins (24).The jacalin lectin recognizes some forms of O-linked glycoproteins, like the mucin family. This approach is valuable for a large scale approach to capture as many proteins from one step. Alternatively for our study, one lectin at a time allowed for an opportunity to differential profile specificities of the lectins, since again we were more interested in the differences between the cancer groups. As mentioned earlier, a different approach would be to utilize a combination of lectins to therefore increase a specific subset of glycoproteins. For example, Zhao et. al have recently described an approach to selectively target all of the sialic acid glycoproteins within a subset of pancreatic cancer serum samples (105). They achieved this by using a panel of lectins that included WGA, SNA and MAL, and over 130 sialylated glycoproteins were identified (105). This could be a valuable strategy, since it also provides information concerning oligosaccharide structures. In addition, within the pancreatic study, they chose to use a top 12 depletion strategy (that removes the 12 most abundant serum proteins) prior to lectin capture and mass

spectrometry analysis, and then compared this directly to lectin capture without the depletion step. Interestingly, the extra depletion step only yielded identification of one unique protein relative to those identified with lectin capture alone without the depletion step.

One characteristic of glycosylation is that any single site has the ability to be occupied by in a diverse heterogeneous population of glycan structures (103). Furthermore, there have been numerous studies, which have shown the prevalence of many different isoforms of PSA as a result of prostate cancer incidence and progression (93, 94). These findings suggests there maybe a great amount of key information obtained within the glycans attached to current markers of disease, have encouraged the development of methodologies that characterize the attached glycans. It is our belief that the future of glcyoproteomics for cancer diagnostics is in the different isoforms of possibly known or existing glycoprotein tumor markers. It is possible that these glycoprotein isoforms may represent new potential biomarkers for detection of disease, monitoring cancer treatment, or surveillance for recurrence post-therapy.

In conclusion, we have demonstrated that an affinity lectin fractionation of serum can reproducibly capture glycoproteins from prostate serum. Alpha-1antitrypsin is one protein identified here which may be a biomarker of prostate cancer as shown here, though it needs to be addressed further. Many of the glycoproteins identified are highly abundant serum proteins, which indicates that the difference may be in the attached glycans. For this reason a strategy for structural elucidation of the attached oligosaccharides was also evaluated.

CHAPTER IV

AIM 3: DETERMINE WHETHER LECTIN CAPTURE OF SERUM IS A REPRODUCIBLE PRE-FRACTIONATION APPROACH FOR PROFILING THE ENTIRE DYNAMIC RANGE OF THE SERUM/PLASMA PROTEOME

4.1 Introduction

As discussed in section 1.2, there has been much progress in the field of mass spectrometry regarding protein analysis since the introduction of soft-ionization technologies in1990 (12). These successes have allowed for the development of higher resolution mass spectrometers with increased throughput capabilities, improved automation for up-front sample fractionation and processing, as well as ever evolving data analysis and user interface software (9, 111, 112, 116). Yet, although these efforts have successfully increased the sample throughput capabilities of protein biomarker mass spectrometry experiments, many aspects of these processes are still not easily amenable to large scale clinical proteomic applications (9, 112, 116). For example, biomarker studies using clinical serum/plasma cohorts with SELDI (surface enhanced laser desorption ionization) and MALDI (matrix assisted laser desorption ionization) time-of-flight (TOF) mass spectrometers have highlighted both the advantages and disadvantages to using these platforms for clinical protein profiling and biomarker discovery (45-49, 52, 160, 161). The main advantages to these methods is their ability to comprehensively analyze all proteins within a given sample, and when combined with automated sample processing, large numbers of samples can be efficiently and reproducibly analyzed as shown in chapter 2, (43, 161, 162). However, in complex clinical fluids like serum or plasma, these profiling methods are ineffective in evaluating the large dynamic concentration range of these proteomes, (111, 116, 161, 163), and are most effective at examining the most abundant proteins. In addition, while these profiling methods were very effective at identifying multiple low-mass peptide/protein biomarker candidates under 15,000 m/z, efficient follow-up methods were lacking for rapid isolation and sequencing determinations of these candidates. Biomarker candidates that have been identified following these profiling studies generally represent fragments or truncations of high abundance proteins, (144, 161) specific disease-associated modifications (26), or potential end-products resulting from the

presence of disease specific proteases, (144), and/or protease activities associated with sample preparation (164). Clearly analysis of the entire dynamic range of the proteome has become a demanding challenge facing the future of clinical biomarker assay development as discussed in chapters II and III of this dissertation (45, 108, 111, 113, 116, 161). Hence, there is a continued need for the development of efficient methods to improve specific mining of complex proteomes in an efficient, cost effective and high throughput manner that facilitates identification of novel disease biomarkers (6, 105, 111).

The need to identify low-abundance biomarker candidates has led to a rejuvenation in the application of chromatographic methods to pre-fractionate complex clinical samples prior to mass spectrometry evaluation (9, 45, 108, 116). One chromatographic strategy gaining widespread use is to specifically deplete the most abundant serum and plasma proteins (like albumin and immunoglobulins) as the first step in the biomarker analysis workflow (108, 116). Most of these approaches use either antibody-based or affinity-gel based resins to deplete the most abundant proteins in serum or plasma, and many are commercially available as kits. There are a variety of options within these serum/plasma depletion kits, including those which target the two most abundant proteins, albumin and immunoglobulin G (IgG), as well as more progressively expensive versions that remove from 5-20 of the most prevalent serum/plasma proteins. A list of the top 20 most serum/plasma proteins depleted by these methods is included in Table 8. Although these antibody-based approaches do achieve the desired depletions, they are not cost efficient, require large starting sample volumes, and cannot be easily automated, significantly limiting their use for larger scale clinical studies that require analysis of hundreds and thousands of samples (108, 111, 116). In contrast, the use of carbohydrate binding lectins as a pre-fractionation approach has emerged as a more cost effective and fully automated alternative to other depletion approaches. Lectins have historically been utilized for elucidation of glycoprotein structure and function, (101, 165-167). Yet, as shown in chapter 3, lectin affinity approaches have recently been utilized to selectively enrich and catalogue glycoproteins associated with cancer progression (81). As applied to proteomic mass spectrometry studies, lectins have been used to

Table 8. List of Proteins Commonly Depleted by Top 20 Depletion Kits.

PROTEIN	MOLECULAR WEIGHT
Albumin	66 kDa
Apolipoprotein A1	28 kDa
Apolipoprotein A2	11 kDa
Apolipoprotein B	550 kDa
Immunoglobulin G	150 kDa
Transferrin	77 kDa
Fibrinogen	340 kDa
Acid-1-Glycoprotein	43 kDa
Immunoglobulin A	150-300 kDa
Ceruloplasmin	130 kDa
Alpha-2-macroglobulin	700 kDa
Complement C4	200 kDa
Immunoglobulin M	950 kDa
Complement C1q	26 kDa
Alpha-1-Antitrypsin	52 kDa
Immunoglobulin D	180 kDa
Complement C3	187 kDa
Transthyretin	55 kDa
Haptoglobin	400 kDa
Plasminogen	81 kDa

specifically isolate glycopeptides in complex peptide mixtures (37, 104, 105, 113). These types of studies have re-introduced lectins as a valid pre-fractionation approach for glycoprotein rich clinical fluids like serum and plasma (104, 113).

When we first began using lectins as a glycoprotein enrichment step in chapter 3, we noticed that the material which bound to the lectins, was more simplified than our starting serum samples. For example, lectin pre-fractionation decreased the amount of serum albumin then samples, which were not pre-fractionated with the lectin. This finding led us to question whether lectin fractionation could be useful as a pre-fractionation approach for proteomic profiling studies of serum. The study herein shows the potential of using lectin affinity capture as a pre-fractionation step in proteomic studies to reduce the sample complexity of serum and/or plasma samples prior to proteomic analysis. We have mainly assessed this approach by gel fractionation, though recently we have begun assessing the incorporation of a quantitative isotope labeling method (iTRAQ™) coupled directly to sequencing with an LTQ mass spectrometer. We have greatly increased the number of proteins identified by using this strategy. We believe this approach provides an alternative, low-cost option to decreasing albumin levels within plasma and serum while providing better sample throughput options important for larger-scale clinical proteomic studies.

4.2 Materials and Methods

Serum and Plasma Samples

The Virginia Prostate Center Tissue and Body Fluid Bank obtained all serum and plasma samples. Blood samples were collected under the same protocol from properly consented patients and obtained from the Department of Urology, Eastern Virginia Medical School. Prostate seminal plasma samples were also obtained from the Virginia Prostate Center Tissue and Body Fluid Bank from patients diagnosed as having either prostate cancer or benign prostatic hyperplasia.

Removal of Albumin and IgG

Two commercially available kits were used to evaluate depletion of major blood proteins prior to lectin incubations including an antibody-based albumin and immunoglobulin G depletion kit (ProteoPrep IA, Sigma-Aldrich), and a blue-dye affinity resin Mcntage albumin depletion column (Millipore). Sigma depletion column – the sigma spin column was incubated with 400 ul of sigma equilibration buffer and allowed to spin at 8,000-x g for 30 seconds. The flow through was collected and the column was equilibrated 2 more times in the same way. Following equilibration, 70 µl of serum or plasma + 30 ul of equilibration buffer was applied to the column and allowed to incubate for 10 minutes at RT. After 10 minutes the column was spun for 1 minute at 8,000-x g and the flow through collected and re-applied to the column, which again was allowed to incubate for 10 minutes. The flow through was collected and combined with 100 ul of equilibration buffer, which was applied to collect all the remaining material from the column. To strip the albumin and IgG from the column, 100 ul of stripping buffer was applied to the column and incubated for 10 minutes, the flow through was collected.

Montage depletion –to prepare the column for depletion, 400 ul of equilibration buffer was applied to the column and spun at 5,000-x g, and the flow through discarded. The column was processed this way 2 times. 70 ul + 30 ul of equilibration buffer was added to the newly equilibrated column and allowed to spun at 5,000 x g. For comparison purposes to the Sigma column, the newly depleted serum was re-applied to the column subjected to a spin and collected. This was combined with 100 ul of equilibration wash buffer.

Lectin Capture with agarose beads

Lectins with known oligosaccharide binding properties which were bound to agarose beads were purchased from Vector or E-Y laboratories. For undepleted serum and plasma samples the lectin incubations were carried out as follows: 50 µl of patient sera/plasma was incubated with 150 µl of lectin binding buffer and added to 150 µl of lectin beads. The bead/serum mixture was allowed to shake and incubate at 4° for 16 hours. Following incubation all unbound material was collected; the beads were washed 3x with appropriate binding buffer (Table 6). The bound glycoproteins were eluted by competitive inhibition of the lectin's appropriate sugar 200 ul (Table 6). Con A

elutes with mannose, and SNA with lactose. Following elution, protein concentrations were determined by Bradford assay to assure equal loading of gels. For depleted serum samples prior to lectin capture, 150 µl of depleted serum was mixed with 50 µl binding buffer and allowed to incubate with 150 µl of the appropriate lectin agarose bead slurry shaking at 4° for 16 hours. Plasma samples were utilized for the lectin capture followed by depletion analysis. A combination of lectins were utilized which included ConA and WGA beads (150ul each). 200 ul depleted plasma was incubated with the bead slurry and allowed to incubate at 4° for 16 hours. Bead elution was carried out sequentially, with addition of 100 ul (200mM mannose), which was allowed to incubate for 10 minutes, and collected followed by 100 ul (200 mM GlcNac). The eluates were acetone precipitated prior to iTRAQ ™ labeling.

Lectin Capture with Bruker Magnetic Beads

Capture with Bruker Magnetic beads (ConA,WGA) was carried out according to the manufacturer protocols and as shown in Sparbier et al (78). The magnetic beads were inverted repeatedly to shake the bead solution (20 times). WGA bead capture: 20 µl of resuspended magnetic beads were added to a standard thin wall PCR-tube and 100 µl of WGA-Wash Buffer 1 was immediately added. The tube was placed in a magnetic bead separator (MBS) and moved back and forth 20 times to allow mixing. The supernatant was collected. The beads were resuspended in 100 µl Wash Buffer 1 and the wash was repeated two more times. Next the beads were resuspended in 20 µl WGA-WB1 to which 20 µl of serum was added. The beads were incubated for 1 hour at room temperature, swirling the tube from time to time to mix the beads. Following incubation, the tube was placed in the magnetic bead separator and the supernatant (unbound) material was collected. Next 100 µl of WGA-Wash Buffer 1 was added and placed on the bead separator and moved back and forth 20 times to wash the beads, and the supernatant collected. 100 µl WGA-WB2 was added and the tubes in the bead separator same as above, this was repeated 2 times. For the elution of immobilized peptides and proteins 20 µl of WGA-elution solution was added and allowed to incubate for 15-20 minutes. Next the tube was placed in the MBS and the eluate was collected. Con A capture . The beads were inverted to mix same as for WGA. 20 µl of resuspended magnetic beads were transferred to a standard thin wall PCR tube and 100 µl

ConA-BS added immediately. The tube was placed in the MBS and moved back and forth between adjacent wells 20 times. The supernatant was removed. The beads were next resuspended in 100 µl ConA-BS and washed 2 additional times as above. For sample addition, 20 µl Con A BS was added first, followed by 20 µl serum and allowed to incubate for 1 hour at room temperature, swirling the tube from time to time to mix. Following incubation the tube was placed on the MBS and unbound removed. 100 µl ConA WS 1 was added, mixed and removed, followed by two washes with ConA WS 2, which was removed. For the elution, 20 µl Con A elution solution was added and allowed to incubate for 10 minutes mixing time to time. For collection of the eluate the tube was placed in the MBS and the eluate bound material collected.

1D-SDS PAGE Analysis

Either NuPAGE 12% Bis-Tris gels (Invitrogen, Carlsbad, CA) or Criterion SDS-PAGE 8-16% gels (BioRad) were utilized. The NuPAGE system uses a NuPAGE MES SDS running buffer in a Novex Mini-Cell system (Invitrogen) and is run using constant voltage of 200 for 45 minutes.

Trypsin Digest

All gel bands were excised and placed into Axygen Maximum Recovery tubes 1.5 mL and washed with 100 mM NH_4HCO_3 to ensure correct pH. To destain silver stain gel pieces they were incubated with a 1:1 mix of destain A and B reagents (Invitrogen) and allowed to shake for 20 minutes followed by two 5 minute washes with 100 mM NH_4HCO_3. To coomassie stained gel pieces they were destained by incubation with a 50% ACN, 50% 100 mM NH_4HCO_3 solution for 10 minutes. Following destaining the gel bands were dehydrated with 50 ul ACN (HPLC grade Fisher) for ten minutes or until the gel slice appears thin and grainy. Next the gel bands were reduced with 60 ul of 20 mM DTT (BioRad) for 30 minutes at RT. Following reduction; the proteins were alkylated with 60 ul of 50 mM iodoacetamide at RT in the dark. (BioRad) Following alkylation the gel pieces were dehydrated three times with ACN, and rehydrated once with 10 mM NH_4HCO_3 for ten minutes. Next the gel pieces were dehydrated three more times with ACN and dried in the speed vacuum 25 ug roche trypsin was mixed 100 mM NH_4HCO_3 and added to the gel pieces. Promega trypsin was mixed to a final concentration of 20 ug with 100 mM NH_4HCO_3 and added to the gel pieces. The digest was carried out at 37° for no longer then 16 hours.

Following digestion the MALDI-bound samples were spotted directly, 1:2 with 10mg/ml CHCA matrix.

iTRAQ™ Labeling of lectin eluates

Following lectin capture and sigma albumin and IgG depletion, the eluates acetone precipitated and labeled using the iTRAQ™ reagents as follows: 20 µL of dissolution buffer was added to the pellet, the samples was reduced for 1 hour at 60 º C for 1 hour. Following incubation, 1 µL cysteine blocking reagent was added and mixed. This was allowed to incubate for 10 minutes at RT. Next the proteins were digested with trypsin. The trypsin was included in the kit and reconstituted with 25 µL Milli-Q® water. 10 µL of the trypsin solution was added to each sample tube and allowed to incubate at 37 °C overnight 12 to 16 hours. Following digestion the samples were labeled with iTRAQ™ reagents. 70 µL of ethanol was added to the each reagent vial (114,115,116,117) and added to the samples. The samples were incubated for 1 hour. After incubation all of the tubes were combined, and purified through an SCX (Applied Biosciences) column to reduce the concentration of buffer salts and organics. The sample is then diluted 1:10 with Buffer A (formic acid, HFBA and acetonitrile) and run on a ThermoElectron LTQ.

LTQ Analysis

The iTRAQ™ labeled samples can be quantitated using the reporter tags which during fragmentation produce ions at m/z 114, 115, 116 or 117. Protein identifications were verified by manual inspection of all peptide ion series, ensuring at least 80% of the combined b and y ion series are represented by ions exceeding 3X background noise. In addition, at least 2 unique peptides are required for identification of any single protein. High quality MS/MS spectra , which resulted in, sequence data but no protein identities were submitted for BLAST search to determine if homologues exist in any of the established NCBI databases. In addition where necessary, de novo sequencing software was used to make high quality sequence assignments, which are searched against the species specific EST databases.

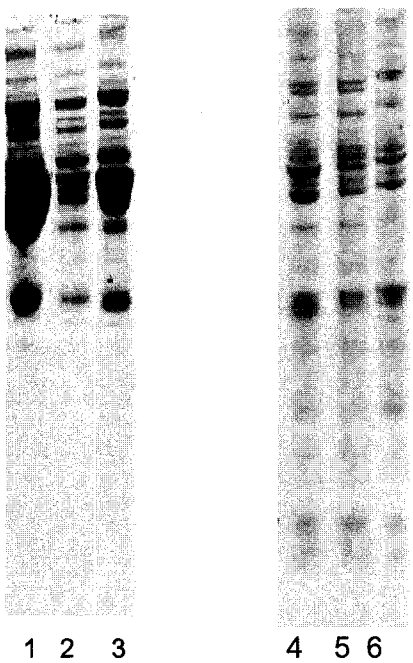
MALDI-TOF MS Analysis

Tryptic eluates were prepared with α-cyano-4-hydroxycinaminic acid (10mg/ml) and analyzed using an Ultraflex MALDI-TOF mass spectrometer (Bruker Daltonics) read in reflectron mode. Eluates from the Bruker Con A magnetic bead experiment were manually spotted 1:2 with 10mg/mL CHCA on an anchor plate and read in linear mode.

4.3 Results

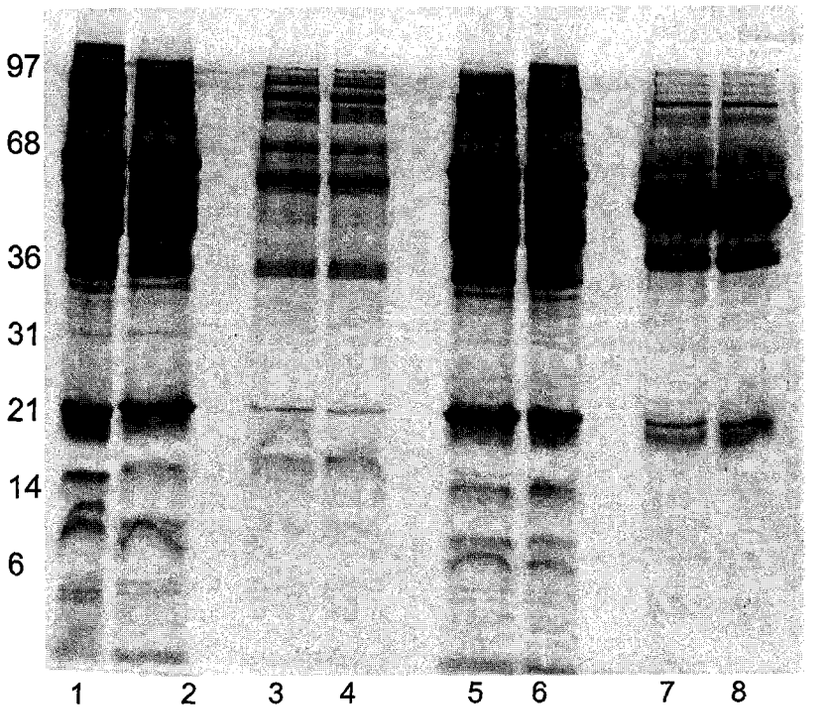
Comparison of albumin depletion and lectin capture

Lectins selectively bind glycosylated proteins; therefore, since many abundant serum proteins are not glycosylated therefore, lectins could decrease levels of albumin in serum or plasma, and provide an alternative strategy to depletion. To determine whether lectin capture could potentially be an alternative to albumin depletion, we first assessed whether lectin capture is comparable to the albumin depletion columns by one-dimensional analysis. Two-serum/plasma depletion strategies were utilized including an antibody based Sigma albumin and immunoglobulin depletion kit and a Montage albumin depletion kit (Millipore) which uses a blue-dye affinity resin for albumin depletion. Figure 47a is a representative gel analysis of raw serum in relation to serum that has been depleted using either the Sigma depletion kit or Montage Albumin depletion kit and visualized by silver stain. As mentioned, the Sigma kit will deplete both albumin and IgG while the Montage kit depletes only albumin. Clearly these kits are efficient in depleting albumin from the serum sample when compared to the raw undepleted serum. However, the kit from sigma appears to have lower levels of albumin remaining following depletion. To compare the lectin pre-fractionation of serum with the depletion approach for albumin removal, serum was incubated with the Con A lectin and then quantitatively compared with the depleted samples as shown in figure 47b. Interestingly, by one-dimensional gel analysis albumin levels of the lectin captured serum are similar to the serum that was first depleted and then lectin fractionated. Comparison of the bound protein profiles from depleted fractions with that of lectin alone indicate a similar pattern of bound proteins independent of depletions. Similar results were obtained with the WGA shown in Figure 48.



1. C4 raw sera
2. C4 sigma depleted
3. C4 montage depleted
4. C4 con A captured
5. C4 sigma depleted Con A
6. C4 montage depleted Con A

Figure 47. Gel View Analysis of Albumin Depletion vs. Lectin Fractionation. 1D gel analysis using 12% Bis-Tris Nu-PAGE gels. Comparison of A) raw serum 2) serum depleted of albumin and IgG (Sigma) and 3) serum depleted with a Montage column (Millipore) Roughly 15 ug was loaded and visualized with coomassie stain. B) Comparison of 1) serum captured with Con a lectin 2) serum depleted with sigma prior to Con A capture or 3) serum depleted with Montage prior to capture with Con A. Roughly 50 ug of protein was loaded and visualized by silver stain.



Lane 1: normal serum depleted of albumin Sigma
Lane 2: normal serum depleted of albumin UB WGA
Lane 3: normal serum depleted of albumin B WGA
Lane 4: cancer serum depleted of albumin B WGA
Lane 5: ca serum depleted of albumin UBWGA
Lane 6: cancer serum depleted of albumin
Lane 7: albumin & IgG eluted off sigma column
Lane 8: albumin & IgG eluted off sigma column

Figure 48. Comparison of Serum Protein Depletion and Lectin Capture Strategies. Pooled prostate cancer serum sample 50 ul was incubated with WGA alone or prefractionated with montage or sigma albumin depletion column prior to incubation with the WGA lectin. Eluates were separated on an 8-16% gel and stained with coomassie.

Reproducibility and Automation of Lectin Capture Approach

A significant advantage of using lectin capture instead of depletion is in its ease of adaptability to automation, since many companies are introducing lectins attached to bead surfaces. Recently, Bruker Daltonics has made available the ConA and WGA lectins attached to magnetic beads, which is therefore compatible to a liquid sampling ClinProt robot. The availability of these beads offers an attractive way to increase both the throughput and reproducibility of this approach as applied to high scale protein profiling experiments. To determine whether the magnetic beads offer the range of proteins bound by the agarose beads we had been using, an initial comparison of both beads in serum and plasma was performed. Figure 49 is a one dimensional gel view of the Con A agarose beads (Vector labs) in comparison with the new Con A Bruker magnetic beads. Both sets of beads appear to bind a wide range of proteins as shown by silver stain, which is apparent for both serum and plasma samples as indicated. To determine whether the magnetic beads can perform selective capture in a highly reproducible manner, next the sample was run in five different bead capture experiments. Figure 50 shows the results for both serum and plasma as separated by 1D gel analysis and visualized with silver stain. A protein assay was not done in this case to further illustrate the reproducibility of the bead capture using the magnetic beads, instead 20 ul was loaded of each sample. By gel view the bead capture looks to be very quantitative and reproducible. Although experiment was not automated since at the lectin beads were not commercially available at the time of this experiment, clearly when processed manually, reproducibility of using these magnetic beads is great. This feature could improve when the entire procedure is fully automated.

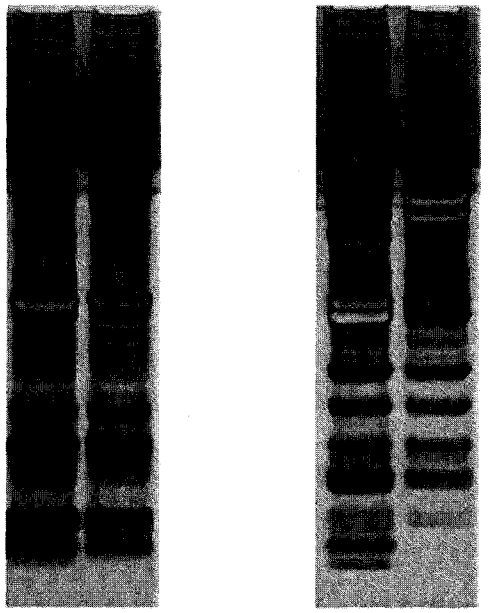
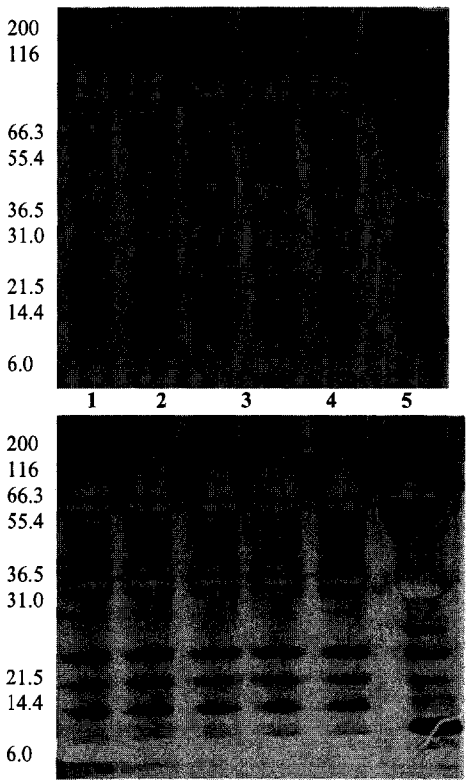


Figure 49. Comparison of Lectin Agarose Beads with Bruker Magnetic Beads. NuPAGE Bis-Tris 12% gel view of serum and plasma samples incubated with either agarose Con A or Bruker Con A magnetic beads. 15 ul of lectin eluate was loaded And visualized by silver stain.



A) Con A serum (MB)

B) Con A Plasma (MB)

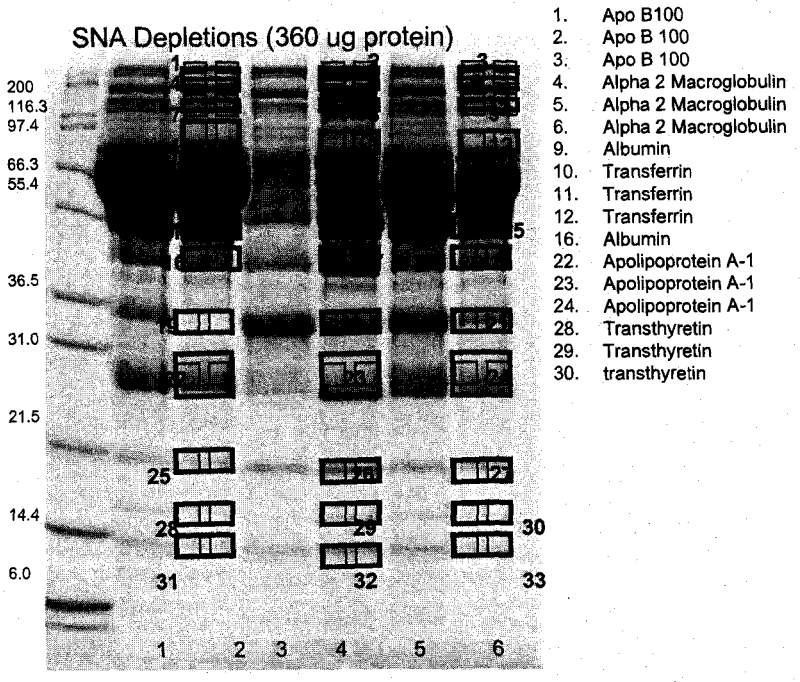
Figure 50. Assessment of Bruker Magnetic Beads Reproducibility in Serum and Plasma. Gel view analysis of the 1 serum sample and 1 plasma sample, which was run in five separate bead fractionation experiments to assess reproducibility of the magnetic bead cature. Eluates were run on 12% NuPAGE Bis-Tris gel and visualized with silver stain.

Comparison of protein identifications of lectin capture vs. depleted lectin capture

To determine if lectin capture without depletion could capture the same sets of proteins when serum is depleted, or see the dynamic range we were achieving, serum samples treated with the lectin alone, or depleted and then treated with the lectin were loaded quantitatively and subject to trypsin digest and peptide mass fingerprinting using a Bruker MALDI-TOF/TOF. As shown in figure 51, the proteins, which are identified by the lectin, are similar to the proteins identified following the depletion strategies, further demonstrating the comparability of using lectin fractionation as opposed to a costly depletion approach. Though this strategy does allow the starting serum sample to be fractionated thereby decreasing albumin levels, and this allows subsequent identification as shown here, the strategy is limited by only being able to identify protein bands as visualized on a gel. In order to increase the amount of proteins identified and see more of the dynamic range of the proteome, it will be necessary to perform a total trypsin digest of lectin eluate and identification without the gel step. Since there is no gel fractionation in this strategy, it may be necessary to get rid of the albumin, yet even following albumin depletion there are still trace amounts of albumin, which can interfere with analysis. This prompted us to determine whether using a depletion column following lectiin fractionation could yield a sample with less total albumin, and thereby increase the dynamic range of proteins seen within a serum or plasma sample.

Lectin Capture followed by Depletion

To assess whether albumin depletion following lectin capture could further decrease the sample complexity of our beginning serum sample and result in more protein identities, lectin capture was carried according to protocol. To increase the coverage of proteins seen from the lectin capture, a combination of lectins was utilized which included 150 ul of ConA and 150 ul of WGA. Roughly 50 ul of plasma sample with the lectin bead slurry and allowed to incubate 16 hours. The eluate was acetone precipitated and depleted using the sigma column and separated by 1D PAGE and compared with serum which had first been depleted and then lectin captured similar to what has been shown previously. This is shown in Figure 52 by gel analysis, in which the level of albumin is now is very minimal compared to the serum which was depleted first and then fractionated with



1. Benign SNA
2. Cancer SNA
3. Benign sigma depleted SNA
4. Cancer sigma depleted SNA
5. Benign montage depleted SNA
6. Cancer Montage depleted SNA

Figure 51. Protein Identification of Different Depletion Approaches. SDS-PAGE separation of serum samples (360 ug) of lectin fractionated or depleted then fractionated samples. Gels were run on 8-16% SDS gels and visualized with coomassie. Red boxes indicate proteins which were excised for identification.

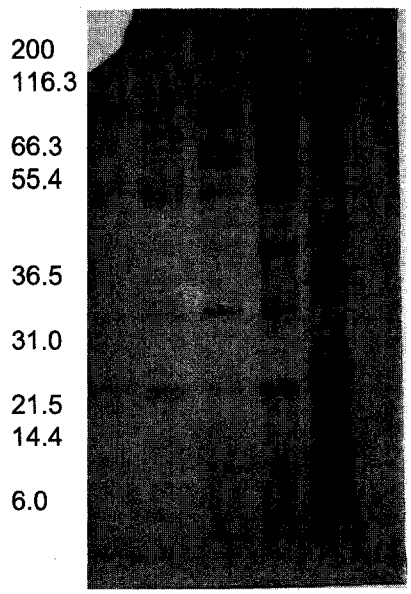


Figure 52. Lectin Fractionation Followed by Albumin Depletion. 12ul of prostate plasma was either 1 and 2 lectin fractionated and then albumin depleted or 2 nd 3, depleted before lectin fractionation.

lectin. To determine whether this approach could allow analysis of more of the dynamic range of the proteome, we chose to perform an experiment in which a quantitative label could be used in combination with total trypsin digest of the eluates and identification with an LTQ instrument. The tentative list of proteins identified by this method is listed in Table 9. By addition of the depletion step following lectin capture it allows for analysis of proteins which are found at less then 50 picograms per mL in human plasma, therefore allowing more of the dynamic range to be analyzed.

4.4 Discussion

Proteomic profiling of clinical fluids has had many recent successes in the discovery of cancer biomarkers which can be attributed in part to some of the technological advancements made within the field of proteomics in the past 10 years. (46-51, 111, 168) However, one issue that continues to plague the future of biomarker research is the extent to which these methodologies can analyze the entire dynamic range of the proteome. (9, 105, 113, 169-171) As a result of these concerns, there has been a concerted effort to develop better pre-fractionation strategies capable of reducing the complexity of serum/plasma enabling the analysis of the entire dynamic range of the proteome (108, 111, 116). Many of these strategies utilize a depletion approach, in which the serum/plasma sample to be analyzed is first depleted of its most abundant proteins. This is done in order to decrease sample complexity, and allow analysis of the low abundant proteins that are otherwise unseen. This manuscript presents lectin fractionation as an attractive alternative approach to decrease albumin levels in clinical fluids prior to mass spectrometry analysis. Also, in addition to decreasing the complexity of serum and plasma comparable to other methods, lectin capture has many advantages over other current strategies.

True, depletion columns will effectively decrease the amount of albumin in serum and plasma, (111, 114), as well as deplete the most abundant proteins in the samples if necessary, yet whether this depletion is also getting rid of other important features is likely (161). For example, controversy exists concerning the specificity of the depletion approaches, and whether potentially important carrier proteins will also be lost (116, 117, 172). This hypothesis has been evaluated by analysis of the bound fraction of albumin depleted from human plasma. In this study

Table 9. List of Proteins Identified

•Alpha 2 macroglobulin	9
•Hemoglobin subunit beta	8
•Ig mu chain C region	6
•Ig mu heavy chain protein	6
•Serotransferrin precursor	7
•Haptoglobin precursor	6
•Alpha-1-antitrypsin	8
•Ig alpha 1 chain C region	4
•Hemoglobin subunit delta	5
•Hemopexin precursor	5
•Fibrinogen	4
•Alpha-1-antichymotrypsin	3
•Complement C4	4
•Ig lambda chain C	4
•Complement C3	4
•ADAMTS precursor	4
•Fibrinogen	2
•Metallothionein-1B	2
•Apolipoprotein C-III	2
•FRAS1-related ECM protein	2
•Cingulin	2
•Vitamin D binding protein	2
•Scavenger receptor protein	2
•Complement C4B	2
•Amyloid like protein	2
•Urokinase type activator	2
•Gap junction alpha 1 protein	2
•ADAM 12 precursor	2
•CD5 antigen like precursor	2
•Interluekin 25 precursor	**2**

57 unique proteins were identified in the bound fraction, complexed with albumin. These proteins which would have otherwise been discarded, could be of interest, since albumin functions as a transporter of many substances in the body including drugs, hormones and fatty acids (108, 116-118, 172). Lectins on the other hand will offer a different fraction of the starting sample material in that they will only bind the glycosylated proteins within the sample. In this way, only proteins with glycosylation events will be in the fraction, which will by default yield a starting material with less albumin which is comparable to front end depletion kits as shown in Figure 47b. Furthermore, when a depletion step is added following lectin fractionation, it allows analysis of proteins within (order of dynamic range).

In addition, lectin strategies offer many practical advantages over most depletion strategies. For example, lectins require small sample starting materials, and are very amenable to automation, as demonstrated by the Bruker magnetic bead experiment in Figures 50-53. Furthermore, lectins are an inexpensive fractionation approach, especially compared to some of the current top 20 abundant protein depletion kits which are very pricey, in their depletion of: apolipoproteins A and B, alpha-1-antitrypsin, alpha-1-acid glycoprotein, haptoglobin and transthyretin to name a few (108). Besides the high cost of these methods, there is also concern about whether these proteins should be removed, since many of these proteins have recently been found by biomarker discovery groups to be differentially expressed in cancers. For example, alpha 1 acid glycoprotein, a known acute phase protein is thought to be down regulated in breast cancers, while levels of transthyretin, a thyroid binding protein is decreased in ovarian cancers. Also, haptoglobin has been shown to be over expressed in late stage ovarian cancers as well as cancers of the pancreas and lung (29, 38, 105, 119, 120). Furthermore, it is also possible that these high abundant proteins have important post-translational events, which could be indicators of disease. For example, at the level of their structural glycans as opposed to only protein expression levels. Until the specific differences in glycosylation constituents are determined and/or other post-translational modification characterized, excluding these types of median abundance serum proteins in biomarker analysis may be premature.

In support of this, for example, Communale et. al recently identified many abundant serum proteins having increased fucosylation in hepatocellular carcinomas. In addition, another recent study looking for biomarkers of pancreatic cancer found altered glycosylation sites of alpha-1-antityrpsin in cancer versus normal serum samples (105). Taken together this data highlights some potential considerations associated with large-scale depletion, and further demonstrates the feasibility of lectin pre-fractionation for serum/plasma profiling. Finally, in order to bring clinical significance to the biomarker discovery process, proteomic profiling studies require reproducibility as well as the need for high throughput and automation. A drawback to the depletion approaches is in their lack of high throughput capabilities. This could allow for the entire process to be run in an automated manner, producing reproducibly lower levels of albumin in each beginning serum and plasma sample as a pre-fractionation approach.

In conclusion, we believe lectin fractionation of serum or plasma prior to proteomic analysis is a legitimate alternative to other approaches, which mainly focus on albumin and abundant protein depletion. Lectins are available in many forms, with numerous specificities at a low cost. They are easily adaptable to automation, and are reproducible, and deplete the albumin levels of serum/plasma to levels comparable to high-end depletion kits. In this way, lectin fractionation offers an affordable and practical approach to possibly analyzing more of the dynamic range of the proteome. It is our belief that this strategy could be very useful for large-scale protein profiling experiments.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Aim I: SELDI-TOF MS profiling of serum from patients with different stages of liver disease.

1. SELDI-TOF MS profiling can be used as a surveillance tool to follow disease progression, offering the flexibility of incorporating known serum markers or other clinical test data into the SELDI data analysis.

2. 38 peaks were identified across the four different clinically stratified liver cancer groups. Of these, four peaks differentially expressed with the ability to discriminate between the different serum sample groups: including a peak at m/z 5,808 found to be increased in HCC serum, a peak at m/z 8939 increased in HCV cirrhosis samples, m/z 9,501 increased in normal serum samples and m/z 11,735 which was increased in HCC serum samples.

3. Using the 38 peaks identified using the SELDI analysis; two- way decision classification trees were made in order to determine the sensitivity and specificity of the peaks for detecting disease. From these results it was clear that SELDI peaks are more effective at distinguishing normal from non-cirrhosis, cirrhosis and HCC conditions as disease severity increases.

4. Decision classification trees constructed from the SELDI-TOF analysis can be used to screen a blinded dataset. Using a blind dataset, the trees could correctly identify 91% of HCC cases (51/56) and 76% of healthy cases (33/42).

5. Incorporation of known clinical tumor markers such as AFP, DCP and GP73 increased the correct classification of chronic HCV vs. HCV-HCC from 71%/64% to 79%/86% and HCV disease vs. HCV-HCC from 61%/76% to 75%/92% sensitivity and specificity.

5.2 Future Directions

One of the largest disadvantages of the SELDI-TOF MS approach utilized in this aim is its inability to identify the peaks/proteins, which classify disease. Various theories exist as to the likely identities of the proteins. Within our study, differential peaks were found across each of the different stages of disease which indicate that the discriminatory peaks are not acute phase reactants as previously suggested, but rather, specific biomarkers of disease. Another possibility

is that these protein peaks are different glycoforms of proteins specific to disease stage or glycoforms of existing tumor markers for HCC, such as AFP, which has recently been shown to have increased fucosylation with HCC progression. (147) Therefore to assess whether glycosylation is responsible for the protein peaks visualized in the data set, the lectin affinity methods from chapters 3 and 4 will be applied to the disease stratified serum sample groups which are currently being collected by the Early Detection Research Network. In this way, glycoproteins can be identified which may be specific to disease stage, as well as assessed quantitatively by the incorporation of iTRAQ™ which could be used to individually label samples from each of the four different sample groups. All potential glycoproteins that are differentially expressed, or determined to be of interest, can then be analyzed similar to that described for the glycan analyses of PAP and IgG in the Aim #2 experiments.

5.3 Aim II: To use a targeted multi-lectin affinity approach to identify glycoproteins present in serum, which can distinguish prostate cancer from benign disease.

1. Upfront differential lectin affinity capture is a reproducible fractionation approach for capturing serum glycoproteins
2. Each lectin utilized produced distinct glycoprotein patterns visualized by SDS-polyacrylamide gel analysis. Coupling lectin fractionation with mass spectrometry analysis allowed for the identification of proteins specifically captured by each lectin.
3. Application of the lectin fractionation and mass spectrometry approach on a set of clinically diagnosed prostate serum samples allowed for the identification of serum glycoprotein isoforms specific to disease stage.
4. Treatment of serum with the enzyme PNGase F allows the release of N-linked oligosaccharides as shown by 1D gel analysis.
5. PNGase F treatment of human seminal plasma prostatic acid phosphatase allows release of the N-glycans attached to the protein. Coupling enzyme digestion with mass spectrometry allows visualization of the PAP glycans. This was also performed for serum Immunoglobulin G (IgG).

5.4 Future Directions

The results of Aim II have established that lectin fractionation is a reproducible approach for the specific capture and characterization of serum glycoproteins. By coupling lectin capture with gel fractionation and mass spectrometry analysis various serum glycoproteins with the potential to detect prostate cancers have been found, although it is unclear whether the glycoproteins are being captured due to over expression at the protein level, or as glycovariants at the carbohydrate level. Therefore, this will need to be assessed by ELISA analysis of lectin captured glycoproteins. Some of the proteins that need to be assessed include: haptoglobin, alpha-1-antitrypsin and alpha-fetoprotein. If these proteins are not differentially expressed at the protein level it will be necessary to identify their glycans using the experimental procedure shown in chapter 3 for prostatic acid phosphatase. However, it will be necessary to incorporate the appropriate software and access to databases, which can annotate sugar structures such as GlycoMod or GlycoSuite DB.

In addition, since PSA is the current diagnostic standard for prostate cancer detection, and it is also a glycoprotein, it will be interesting and necessary to characterize all of the glycoforms of PSA in relation to cancer progression. There have been a handful of studies assessing the different glycoforms of PSA in relation to prostate cancer progression (92-94). From these studies, it has been shown that PSA glycans change with disease progression, although these conclusions were based on analysis of very few samples, highlighting the need for confirmation in a larger serum cohort. Therefore to comprehensively assess all of the PSA glycoforms in relation to cancer stage, we propose to develop an HPLC methodology incorporating a PSA purification step, initially from archived seminal plasma from healthy normals and cancers where the PSA concentrations can be as high as 0.5 mg/ml. The PSA glycoforms will then be fractionated by gel separation followed by glycan analysis. In this way the relationship between PSA glycoforms and cancer progression can be further elucidated, potentially identifying a variant of PSA to aid in early detection diagnostics.

5.5 Aim III: Determine whether lectin capture of serum is a reproducible pre-fractionation approach for profiling the entire dynamic range of the serum/plasma proteome.

1. Upfront lectin fractionation can effectively deplete the amount of albumin within a starting serum or plasma sample, as well as enrich for specific glycoprotein subsets, as visualized by 1D polyacrylamide gel electrophoresis.
2. The amount of depletion of albumin by lectins is comparable to other albumin depletion kits as visualized by gel analysis. This was confirmed by comparing the proteins captured by each method as determined by peptide mass fingerprinting.
3. Bruker Con A magnetic beads can reproducibly fractionate both serum and plasma samples as demonstrated by both gel and mass spectrometry analysis. Therefore, fractionation with lectins offers an affordable and high throughput alternative to high end albumin depletion kits.
4. Incorporation of an albumin depletion step following lectin fractionation decreases residual albumin levels and allows for the identification of proteins within the entire dynamic range of the proteome, as elucidated by LC-MS/MS LTQ analysis.

5.6 Future Directions

From Aim III, it is shown that upfront lectin fractionation is a reproducible alternative approach for albumin depletion prior to proteomic analysis. Lectin fractionation is advantageous over other approaches since it has the capability of being automated, as demonstrated with the Bruker magnetic bead experiment in chapter IV. In addition, when lectin fractionation is followed by albumin depletion, proteins otherwise masked by the most abundant proteins in plasma are identifiable, allowing a greater depth of analysis to the lower ranges of the serum proteome. Combined with the lectin enrichment, this finding could allow identification of biomarkers not previously detected by mass spectrometry methods in relation to disease progression. In addition, this approach has an inherent automation capability that can be applied to a large scale clinical sample set. Therefore, a future experiment would be to fully automate the process on the ClinProt robot with the Bruker beads followed by immunoaffinity beads with an antibody to

albumin. The depleted, glycoprotein enriched samples can then be used for direct analysis by MALDI-TOF/TOF, or further fractionation using trypsin digestion and iTRAQ type approaches.

5.7 Concluding Remarks

In conclusion, proteomic technologies as applied to biomarker research have evolved from global profiling based approaches typified by SELDI-TOF MS to more targeted approaches focusing on capturing specific classes of proteins with a distinct relationship to cancer development. Many of these approaches focus specifically on characterizing the post-translational modifications occurring on proteins such as glycosylation, which has been known to be directly related to cancer progression. The studies described herein illustrate the technological evolution of proteomic expression profiling strategies, ranging from SELDI TOF/MS to affinity capture approaches. The overexpression of Alpha-1-Antitrypsin in the prostate cancer sample group C3 as shown in chapter III is one example of a potentially useful marker found using this the lectin capture methodology. There are numeous roles A1AT may play in cancer progression including inhibiting apopotosis or aiding in the beginning stages of an immune response to cancer. Additionally, A1AT may have different isoforms which could provide additional critical information concerning prostate cancer progression. It is possible that characterization of the isoforms of A1AT or other known existing markers such as PSA, could provide the information to assist known clinical assays to improve the early detection of cancers. Much work remains to be done to optimize and adapt these approaches to large-scale clinical analyses and attaining the goal of a high-throughput clinical assay, yet the strategies described herein can serve as a roadmap to achieve this goal.

REFERENCES

1. Lander E.S. LLM, Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., et. al. Initial Sequencing and analysis of the human genome. Nature 2001;409:860-921.

2. Simpson RJ. Proteins and Proteomics. New York: Cold spring harbor laboratory press; 2002.

3. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science 2001;291(5507):1304-51.

4. Duncan MW, Hunsucker SW. Proteomics as a tool for clinically relevant biomarker discovery and validation. Exp Biol Med (Maywood) 2005;230(11):808-17.

5. Chatterjee SK, Zetter BR. Cancer biomarkers: knowing the present and predicting the future. Future Oncol 2005;1(1):37-50.

6. Drake RR, Cazare LH, Semmes OJ, Wadsworth JT. Serum, salivary and tissue proteomics for discovery of biomarkers for head and neck cancers. Expert Rev Mol Diagn 2005;5(1):93-100.

7. Semmes OJ. Defining the role of mass spectrometry in cancer diagnostics. Cancer Epidemiol Biomarkers Prev 2004;13(10):1555-7.

8. Semmes OJ, Cazares LH, Ward MD, et al. Discrete serum protein signatures discriminate between human retrovirus-associated hematologic and neurologic disease. Leukemia 2005;19(7):1229-38.

9. Semmes OJ, Malik G, Ward M. Application of mass spectrometry to the discovery of biomarkers for detection of prostate cancer. J Cell Biochem 2006;98(3):496-503.

10. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. Nat Genet 2003;33 Suppl:311-23.

11. Sanger F. The free amino groups of insulin. Biochem J 1945;39(5):507-15.

12. Kinter MaNES. Protein Sequencing and Identification Using Tandem Mass Spectrometry. New york: John Wiley and Sons Inc.; 2000.

13. Edman P. Method for determination of the amino acid sequence in peptides. Acta Chem Scand 1950;4:283-93.

14. Edman P. On the mechanism of the phenyl isothiocyanate degredation of peptides. Acta Chem Scand 1956;10:761-8.

15. Edman P. Phenylthiohydantoins in protein analysis. Ann NY Acad Sci 1960;88:602-10.

16. Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 1975;26(3):231-43.

17. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem 1975;250(10):4007-21.

18. Gorg A. Two-dimensional electrophoresis with immobilized pH gradients: current state. Biochem Soc Trans 1993;21(1):130-2.

19. Gorg A, Obermaier C, Boguth G, et al. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 2000;21(6):1037-53.

20. Gorg A, Postel W, Gunther S. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 1988;9(9):531-46.

21. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science 1989;246(4926):64-71.

22. Tomer KB, Crow FW, Gross ML, Kopple KD. Fast atom bombardment combined with tandem mass spectrometry for the determination of cyclic peptides. Anal Chem 1984;56(6):880-6.

23. Chang DY, Lee CC, Shiea J. Detecting large biomolecules from high-salt solutions by fused-droplet electrospray ionization mass spectrometry. Anal Chem 2002;74(11):2465-9.

24. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol 2003;21(6):660-6.

25. Wright Jr GL, Semmes OJ. Proteomics in Health and Disease. J Biomed Biotechnol 2003;2003(4):215-6.

26. Malik G, Ward MD, Gupta SK, et al. Serum levels of an isoform of apolipoprotein A-II as a potential marker for prostate cancer. Clin Cancer Res 2005;11(3):1073-85.

27. Semmes OJ. The "omics" haystack: defining sources of sample bias in expression profiling. Clin Chem 2005;51(9):1571-2.

28. Semmes OJ, Feng Z, Adam BL, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of p prostate cancer: I. Assessment of platform reproducibility. Clin Chem 2005;51(1):102-12.

29. Alaoui-Jamali MA, Xu YJ. Proteomic technology for biomarker profiling in cancer: an update. J Zhejiang Univ Sci B 2006;7(6):411-20.

30. Block TM, Comunale MA, Lowman M, et al. Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. Proc Natl Acad Sci U S A 2005;102(3):779-84.

31. Marrero JA. Screening tests for hepatocellular carcinoma. Clin Liver Dis 2005;9(2):235-51, vi.

32. Marrero JA. Hepatocellular carcinoma. Curr Opin Gastroenterol 2005;21(3):308-12.

33. Marrero JA. Hepatocellular carcinoma. Curr Opin Gastroenterol 2006;22(3):248-53.

34. Marrero JA. Staging systems for hepatocellular carcinoma: should we all use the BCLC system? J Hepatol 2006;44(4):630-2.

35. Marrero JA, Fontana RJ, Barrat A, et al. Prognosis of hepatocellular carcinoma: comparison of 7 staging systems in an American cohort. Hepatology 2005;41(4):707-16.

36. Chen SL, Morgan TR. The natural history of hepatitis C virus (HCV) infection. Int J Med Sci 2006;3(2):47-52.

37. Comunale MA, Lowman M, Long RE, et al. Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. J Proteome Res 2006;5(2):308-15.

38. Huang RH, Hu KQ. A practical approach to managing patients with HCV infection. Int J Med Sci 2006;3(2):63-8.

39. Favre D, Muellhaupt B. Potential cellular receptors involved in hepatitis C virus entry into cells. Lipids Health Dis 2005;4(1):9.

40. Bartenschlager R, Lohmann V. Replication of the hepatitis C virus. Baillieres Best Pract Res Clin Gastroenterol 2000;14(2):241-54.

41. Micallef JM, Kaldor JM, Dore GJ. Spontaneous viral clearance following acute hepatitis C infection: a systematic review of longitudinal studies. J Viral Hepat 2006;13(1):34-41.

42. Comunale MA, Mattu TS, Lowman MA, et al. Comparative proteomic analysis of de-N-glycosylated serum from hepatitis B carriers reveals polypeptides that correlate with disease status. Proteomics 2004;4(3):826-38.

43. Schwegler EE, Cazares L, Steel LF, et al. SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. Hepatology 2005;41(3):634-42.

44. El-Houseini ME, Mohammed MS, Elshemey WM, Hussein TD, Desouky OS, Elsayed AA. Enhanced detection of hepatocellular carcinoma. Cancer Control 2005;12(4):248-53.

45. Zhou L, Liu J, Luo F. Serum tumor markers for detection of hepatocellular carcinoma. World J Gastroenterol 2006;12(8):1175-81.

46. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62(13):3609-14.

47. Adam BL, Vlahou A, Semmes OJ, Wright GL, Jr. Proteomic approaches to biomarker discovery in prostate and bladder cancers. Proteomics 2001;1(10):1264-70.

48. Cazares LH, Adam BL, Ward MD, et al. Normal, benign, preneoplastic, and malignant prostate cells have distinct protein expression profiles resolved by surface enhanced laser desorption/ionization mass spectrometry. Clin Cancer Res 2002;8(8):2541-52.

49. Poon TC, Yip TT, Chan AT, et al. Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. Clin Chem 2003;49(5):752-60.

50. Vlahou A, Giannopoulos A, Gregory BW, et al. Protein profiling in urine for the diagnosis of bladder cancer. Clin Chem 2004;50(8):1438-41.

51. Vlahou A, Laronga C, Wilson L, et al. A novel approach toward development of a rapid blood test for breast cancer. Clin Breast Cancer 2003;4(3):203-9.

52. Wadsworth JT, Somers KD, Cazares LH, et al. Serum protein profiles to identify head and neck cancer. Clin Cancer Res 2004;10(5):1625-32.

53. Blum HE. Hepatocellular carcinoma: therapy and prevention. World J Gastroenterol 2005;11(47):7391-400.

54. Mimeault M, Batra SK. Recent advances on multiple tumorigenic cascades involved in prostatic cancer progression and targeting therapies. Carcinogenesis 2006;27(1):1-22.

55. Aumuller G, Goebel HW, Bacher M, Eicheler W, Rausch U. [Current aspects on morphology and functions of the prostate]. Verh Dtsch Ges Pathol 1993;77:1-18.

56. Buhmeida A, Pyrhonen S, Laato M, Collan Y. Prognostic factors in prostate cancer. Diagn Pathol 2006;1:4.

57. Kasper S, Cookson MS. Mechanisms leading to the development of hormone-resistant prostate cancer. Urol Clin North Am 2006;33(2):201-10, vii.

58. Hendriksen PJ, Dits NF, Kokame K, et al. Evolution of the androgen receptor pathway during progression of prostate cancer. Cancer Res 2006;66(10):5012-20.

59. Isono T, Tanaka T, Kageyama S, Yoshiki T. Structural diversity of cancer-related and non-cancer-related prostate-specific antigen. Clin Chem 2002;48(12):2187-94.

60. Tricoli JV, Schoenfeldt M, Conley BA. Detection of prostate cancer and predicting progression: current and future diagnostic markers. Clin Cancer Res 2004;10(12 Pt 1):3943-53.

61. Bindukumar B, Kawinski E, Cherrin C, et al. Two step procedure for purification of enzymatically active prostate-specific antigen from seminal plasma. J Chromatogr B Analyt Technol Biomed Life Sci 2004;813(1-2):113-20.

62. Diamandis E. Prostate-specific antigen: its usefulness in clinical medicine. Trends Endocrinol Metab 1998;9:310-6.

63. Catalona WJ, Loeb S, Han M. Viewpoint: expanding prostate cancer screening. Ann Intern Med 2006;144(6):441-3.

64. Khauli RB. Prostate cancer: diagnostic and therapeutic strategies with emphasis on the role of PSA. J Med Liban 2005;53(2):95-102.

65. De Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. Nat Rev Cancer 2006;6(1):24-37.

66. Janeway Charles PT, Mark Walport, Mark Shlomchik. Immunobiology. 5 ed. New York City: Garland Publishing; 2001.

67. Dunn GP, Koebel CM, Schreiber RD. Interferons, immunity and cancer immunoediting. Nat Rev Immunol 2006;6(11):836-48.

68. Zou W. Immunosupressive networks in the tumour environment and their therapeutic relevance. Nature Reviews 2005;5:263-74.

69. Gilboa E. The promise of cancer vaccines. Nat Rev Cancer 2004;4(5):401-11.

70. Lollini PL, Cavallo F, Nanni P, Forni G. Vaccines for tumour prevention. Nat Rev Cancer 2006;6(3):204-16.

71. Chung LW, Baseman A, Assikis V, Zhau HE. Molecular insights into prostate cancer progression: the missing link of tumor microenvironment. J Urol 2005;173(1):10-20.

72. Ishigaki H, Yamamoto Y, Ishida H, et al. Preparation and functional analysis of tumor-infiltrating stroma cells using bone marrow chimera mice. Microbiol Immunol 2006;50(8):655-62.

73. Fassnacht M, Lee J, Milazzo C, et al. Induction of CD4(+) and CD8(+) T-cell responses to the human stromal antigen, fibroblast activation protein: implication for cancer immunotherapy. Clin Cancer Res 2005;11(15):5566-71.

74. Mosca PJ, Lyerly HK, Ching CD, Hobeika AC, Clay TM, Morse MA. Proteomics for monitoring immune responses to cancer vaccines. Curr Opin Mol Ther 2003;5(1):39-43.

75. Butler M, Quelhas D, Critchley AJ, et al. Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan processing step and provides an insight into pathogenesis. Glycobiology 2003;13(9):601-22.

76. Hakomori S. Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. Cancer Res 1996;56(23):5309-18.

77. Sharon N, Lis H. History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology 2004;14(11):53R-62R.

78. Sparbier K, Koch S, Kessler I, Wenzel T, Kostrzewa M. Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. J Biomol Tech 2005;16(4):407-13.

79. Dwek MV, Brooks SA. Harnessing changes in cellular glycosylation in new cancer treatment strategies. Curr Cancer Drug Targets 2004;4(5):425-42.

80. Dwek MV, Lacey HA, Leathem AJ. Breast cancer progression is associated with a reduction in the diversity of sialylated and neutral oligosaccharides. Clin Chim Acta 1998;271(2):191-202.

81. Dwek MV, Ross HA, Streets AJ, et al. Helix pomatia agglutinin lectin-binding o oligosaccharides of aggressive breast cancer. Int J Cancer 2001;95(2):79-85.

82. Dwek MV, Ross HA, Leathem AJ. Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. Proteomics 2001;1(6):756-62.

83. Suriano R, Ghosh SK, Ashok BT, et al. Differences in glycosylation patterns of heat shock protein, gp96: implications for prostate cancer prevention. Cancer Res 2005;65(14):6466-75.

84. Dennis JW, Granovsky M, Warren CE. Protein glycosylation in development and disease. Bioessays 1999;21(5):412-21.

85. Kim YJ, Varki A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. Glycoconj J 1997;14(5):569-76.

86. Orntoft TF, Vestergaard EM. Clinical aspects of altered glycosylation of glycoproteins in cancer. Electrophoresis 1999;20(2):362-71.

87. Fernandes B, Sagman U, Auger M, Demetrio M, Dennis JW. Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. Cancer Res 1991;51(2):718-23.

88. Handerson T, Pawelek JM. Beta1,6-branched oligosaccharides and coarse vesicles: a common, pervasive phenotype in melanoma and other human cancers. Cancer Res 2003;63(17):5363-9.

89. Siddiqui SF, Pawelek J, Handerson T, et al. Coexpression of beta1,6-N-acetylglucosaminyltransferase V glycoprotein substrates defines aggressive breast cancers with poor outcome. Cancer Epidemiol Biomarkers Prev 2005;14(11 Pt 1):2517-23.

90. Yanagi M, Aoyagi Y, Suda T, Mita Y, Asakura H. N-Acetylglucosaminyltransferase V as a possible aid for the evaluation of tumor invasiveness in patients with hepatocellular carcinoma. J Gastroenterol Hepatol 2001;16(11):1282-9.

91. Ohyama C, Hosono M, Nitta K, et al. Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy. Glycobiology 2004;14(8):671-9.

92. Peracaula R, Tabares G, Royle L, et al. Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. Glycobiology 2003;13(6):457-70.

93. Prakash S, Robbins PW. Glycotyping of prostate specific antigen. Glycobiology 2000;10(2):173-6.

94. Tabares G, Radcliffe CM, Barrabes S, et al. Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA. Glycobiology 2006;16(2):132-45.

95. Breborowicz J, Mackiewicz A, Breborowicz D. Microheterogeneity of alpha-fetoprotein in patient serum as demonstrated by lectin affino-electrophoresis. Scand J Immunol 1981;14(1):15-20.

96. Johnson PJ, Poon TC, Hjelm NM, Ho CS, Blake C, Ho SK. Structures of disease-specific serum alpha-fetoprotein isoforms. Br J Cancer 2000;83(10):1330-7.

97. Miyazaki J, Endo Y., Oda T. Lectin affinities of alpha-fetoprotein in liver cirrhosis, hepatocellular carcinoma and metastatic liver tumor. Acta hepatol Jpn 1981;22:1559-68.

98. Hall MP, Schneider LV. Isotope-differentiated binding energy shift tags (IDBEST) for improved targeted biomarker discovery and validation. Expert Rev Proteomics 2004;1(4):421-31.

99. Xiong L, Andrews D, Regnier F. Comparative proteomics of glycoproteins based on lectin selection and isotope coding. J Proteome Res 2003;2(6):618-25.

100. Phillips PG, Horowitz P, Macnab J. Molecular heterogeneity of the concanavalin A tetramer: effects on binding to human red blood cells. Biochem Int 1983;7(5):549-55.

101. Dam TK, Roy R, Das SK, Oscarson S, Brewer CF. Binding of multivalent carbohydrates to concanavalin A and Dioclea grandiflora lectin. Thermodynamic analysis of the "multivalency effect". J Biol Chem 2000;275(19):14223-30.

102. Mandal DK, Kishore N, Brewer CF. Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. Biochemistry 1994;33(5):1149-56.

103. Kuster B, Wheeler SF, Hunter AP, Dwek RA, Harvey DJ. Sequencing of N-linked oligosaccharides directly from protein gels: in-gel deglycosylation followed by matrix-assisted laser desorption/ionization mass spectrometry and normal-phase high-performance liquid chromatography. Anal Biochem 1997;250(1):82-101.

104. Mehta AI, Ross S, Lowenthal MS, et al. Biomarker amplification by serum carrier protein binding. Dis Markers 2003;19(1):1-10.

105. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. J Proteome Res 2006;5(7):1792-802.

106. Alvarez-Manilla G, Atwood J, 3rd, Guo Y, Warren NL, Orlando R, Pierce M. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. J Proteome Res 2006;5(3):701-8.

107. Itoh S, Kawasaki N, Hashii N, et al. N-linked oligosaccharide analysis of rat brain Thy-1 by liquid chromatography with graphitized carbon column/ion trap-Fourier transform ion cyclotron resonance mass spectrometry in positive and negative ion modes. J Chromatogr A 2006;1103(2):296-306.

108. Liu T, Qian WJ, Gritsenko MA, et al. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. J Proteome Res 2005;4(6):2070-80.

109. Wang Y, Wu SL, Hancock WS. Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap--Fourier transform mass spectrometry. Glycobiology 2006;16(6):514-23.

110. Anderson LaNGA. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. Molecular & Cellular Proteomics 2002;1:845-67.

111. Anderson MPaL. A List of Candidate Cancer Biomarkers for Targeted Proteomics. Biomarker Insights 2006;2:1-48.

112. Qian WJ, Jacobs JM, Liu T, Camp DG, 2nd, Smith RD. Advances and challenges in liquid chromatography-mass spectrometry based proteomic profiling for clinical applications. Mol Cell Proteomics 2006.

113. Drake RR, Schwegler EE, Malik G, et al. Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. Mol Cell Proteomics 2006.

114. Fu Q, Garnham CP, Elliott ST, Bovenkamp DE, Van Eyk JE. A robust, streamlined, and reproducible method for proteomic analysis of serum by delipidation, albumin and IgG depletion, and two-dimensional gel electrophoresis. Proteomics 2005;5(10):2656-64.

115. Qiu R, Regnier FE. Comparative glycoproteomics of N-linked complex-type glycoforms containing sialic acid in human serum. Anal Chem 2005;77(22):7225-31.

116. Gong Y, Li X, Yang B, et al. Different immunoaffinity fractionation strategies to characterize the human plasma proteome. J Proteome Res 2006;5(6):1379-87.

117. Yocum AK, Yu K, Oe T, Blair IA. Effect of immunoaffinity depletion of human serum during proteomic investigations. J Proteome Res 2005;4(5):1722-31.

118. Simard JR, Zunszain PA, Ha CE, et al. Locating high-affinity fatty acid-binding sites on albumin by x-ray crystallography and NMR spectroscopy. Proc Natl Acad Sci U S A 2005;102(50):17958-63.

119. Ahmed N, Oliva KT, Barker G, et al. Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer. Proteomics 2005;5(17):4625-36.

120. Goufman EI, Moshkovskii SA, Tikhonova OV, et al. Two-dimensional electrophoretic proteome study of serum thermostable fraction from patients with various tumor conditions. Biochemistry (Mosc) 2006;71(4):354-60.

121. Wright ME, Han DK, Aebersold R. Mass spectrometry-based expression profiling of clinical prostate cancer. Mol Cell Proteomics 2005;4(4):545-54.

122. Kozak KR, Amneus MW, Pusey SM, et al. Identification of biomarkers for ovarian cancer using strong anion-exchange ProteinChips: potential use in diagnosis and prognosis. Proc Natl Acad Sci U S A 2003;100(21):12343-8.

123. Lehrer S, Roboz J, Ding H, et al. Putative protein markers in the sera of men with prostatic neoplasms. BJU Int 2003;92(3):223-5.

124. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem 2002;48(8):1296-304.

125. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359(9306):572-7.

126. Petricoin EF, 3rd, Ornstein DK, Paweletz CP, et al. Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst 2002;94(20):1576-8.

127. Shiwa M, Nishimura Y, Wakatabe R, et al. Rapid discovery and identification of a tissue-specific tumor biomarker from 39 human cancer cell lines using the SELDI ProteinChip platform. Biochem Biophys Res Commun 2003;309(1):18-25.

128. Zhukov TA, Johanson RA, Cantor AB, Clark RA, Tockman MS. Discovery of distinct profiles specific for lung tumors and pre-malignant lung lesions by SELDI mass spectrometry. Lung Cancer 2003;40(3):267-79.

129. Bruix J, Llovet JM. Prognostic prediction and treatment strategy in hepatocellular carcinoma. Hepatology 2002;35(3):519-24.

130. I-Neng Lee C-HC, Jin-Chuan Sheu, Hsuan-Shu Lee, Guan-Tarn Huang, Chen-Yin Yu, fung-Jou Lu, and Lu-Ping Chow. Identification of Human Hepatocellular Carcinoma-Related Biomarkers by Two-Dimensional Difference Gel Electrophoresis and Mass Spectrometry. Journal of Proteome Research 2005(4):2062-9.

131. Chignard N, Shang S, Wang H, et al. Cleavage of endoplasmic reticulum proteins in hepatocellular carcinoma: Detection of generated fragments in patient sera. Gastroenterology 2006;130(7):2010-22.

132. Ajisaka H, Shimizu K, Miwa K. Immunohistochemical study of protein induced by vitamin K absence or antagonist II in hepatocellular carcinoma. J Surg Oncol 2003;84(2):89-93.

133. Nagaoka S, Yatsuhashi H, Hamada H, et al. The des-gamma-carboxy prothrombin index is a new prognostic indicator for hepatocellular carcinoma. Cancer 2003;98(12):2671-7.

134. Marrero JA, Romano PR, Nikolaeva O, et al. GP73, a resident Golgi glycoprotein, is a novel serum marker for hepatocellular carcinoma. J Hepatol 2005;43(6):1007-12.

135. Marrero JA, Su GL, Wei W, et al. Des-gamma carboxyprothrombin can differentiate hepatocellular carcinoma from nonmalignant chronic liver disease in american patients. Hepatology 2003;37(5):1114-21.

136. Kladney RD, Bulla GA, Guo L, et al. GP73, a novel Golgi-localized protein upregulated by viral infection. Gene 2000;249(1-2):53-65.

137. Kladney RD, Cui X, Bulla GA, Brunt EM, Fimmel CJ. Expression of GP73, a resident Golgi membrane protein, in viral and nonviral liver disease. Hepatology 2002;35(6):1431-40.

138. Bertone P, Kluger Y, Lan N, et al. SPINE: an integrated tracking database and data mining approach for identifying feasible targets in high-throughput structural proteomics. Nucleic Acids Res 2001;29(13):2884-98.

139. Kosuda S, Ichihara K, Watanabe M, Kobayashi H, Kusano S. Decision-tree sensitivity analysis for cost-effectiveness of chest 2-fluoro-2-D-[(18)F]fluorodeoxyglucose positron emission tomography in patients with pulmonary nodules (non-small cell lung carcinoma) in Japan. Chest 2000;117(2):346-53.

140. Wadsworth JT, Somers KD, Stack BC, Jr., et al. Identification of patients with head and neck cancer using serum protein profiles. Arch Otolaryngol Head Neck Surg 2004;130(1):98-104.

141. Kim J, Kim SH, Lee SU, et al. Proteome analysis of human liver tumor tissue by two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-mass spectrometry for identification of disease-related proteins. Electrophoresis 2002;23(24):4142-56.

142. Steel LF, Shumpert D, Trotter M, et al. A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. Proteomics 2003;3(5):601-9.

143. Koopmann J, Zhang Z, White N, et al. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. Clin Cancer Res 2004;10(3):860-8.

144. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 2006;116(1):271-84.

145. Diamandis EP. Validation of breast cancer biomarkers identified by mass spectrometry. Clin Chem 2006;52(4):771-2; author reply 2.

146. Ward DG, Suggett N, Cheng Y, et al. Identification of serum biomarkers for colon cancer by proteomic analysis. Br J Cancer 2006;94(12):1898-905.

147. Nakagawa T, Uozumi N, Nakano M, et al. Fucosylation of N-glycans regulates the secretion of hepatic glycoproteins into bile ducts. J Biol Chem 2006.

148. Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. J Chromatogr A 2004;1053(1-2):79-88.

149. Yang Z, Hancock WS, Chew TR, Bonilla L. A study of glycoproteins in human serum and plasma reference standards (HUPO) using multilectin affinity chromatography coupled with RPLC-MS/MS. Proteomics 2005;5(13):3353-66.

150. He Z, Aristoteli LP, Kritharides L, Garner B. HPLC analysis of discrete haptoglobin isoform N-linked oligosaccharides following 2D-PAGE isolation. Biochem Biophys Res Commun 2006;343(2):496-503.

151. Jakob CG, Lewinski K, Kuciel R, Ostrowski W, Lebioda L. Crystal structure of human prostatic acid phosphatase. Prostate 2000;42(3):211-8.

152. Yoshida KI, Honda M, Arai K, et al. Serial lectin affinity chromatography with concavalin A and wheat germ agglutinin demonstrates altered asparagine-linked sugar-chain structures of prostatic acid phosphatase in human prostate carcinoma. J Chromatogr B Biomed Sci Appl 1997;695(2):439-43.

153. Dwek MV, Alaiya AA. Proteome analysis enables separate clustering of normal breast, benign breast and breast cancer tissues. Br J Cancer 2003;89(2):305-7.

154. Dwek MV, Brooks SA, Streets AJ, Harvey DJ, Leathem AJ. Oligosaccharide release from frozen and paraffin-wax-embedded archival tissues. Anal Biochem 1996;242(1):8-14.

155. Dwek MV, Rawlings SL. Current perspectives in cancer proteomics. Mol Biotechnol 2002;22(2):139-52.

156. Dwek MV, Rawlings SL. Breast cancer proteomics using two-dimensional electrophoresis: studying the breast cancer proteome. Methods Mol Med 2006;120:231-43.

157. Huang CM, Ananthaswamy HN, Barnes S, Ma Y, Kawai M, Elmets CA. Mass spectrometric proteomics profiles of in vivo tumor secretomes: Capillary ultrafiltration sampling of regressive tumor masses. Proteomics 2006.

158. Petrache I, Fijalkowska I, Medler TR, et al. alpha-1 antitrypsin inhibits caspase-3 activity, preventing lung endothelial cell apoptosis. Am J Pathol 2006;169(4):1155-66.

159. Kuvibidila S, Rayford W. Correlation between serum prostate-specific antigen and alpha-1-antitrypsin in men without and with prostate cancer. J Lab Clin Med 2006;147(4):174-81.

160. Cheng AJ, Chen LC, Chien KY, et al. Oral cancer plasma tumor marker identified with bead-based affinity-fractionated proteomic technology. Clin Chem 2005;51(12):2236-44.

161. Conrads TP, Hood BL, Petricoin EF, 3rd, Liotta LA, Veenstra TD. Cancer proteomics: many technologies, one goal. Expert Rev Proteomics 2005;2(5):693-703.

162. Shand RL, Gelmann EP. Molecular biology of prostate-cancer pathogenesis. Curr Opin Urol 2006;16(3):123-31.

163. Zhou M, Lucas DA, Chan KC, et al. An investigation into the human serum "interactome". Electrophoresis 2004;25(9):1289-98.

164. Liotta LA, Petricoin EF. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. J Clin Invest 2006;116(1):26-30.

165. Goldstein IJ. Lectin structure-activity: the story is never over. J Agric Food Chem 2002;50(22):6583-5.

166. Hirabayashi J. Lectin-based structural glycomics: glycoproteomics and glycan profiling. Glycoconj J 2004;21:35-40.

167. Lis H, Sharon N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. Chem Rev 1998;98(2):637-74.

168. Conrads TP, Fusaro VA, Ross S, et al. High-resolution serum proteomic features for ovarian cancer detection. Endocr Relat Cancer 2004;11(2):163-78.

169. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004;20(5):777-85.

170. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 2004;4(4):309-14.

171. Sorace JM, Zhan M. A data review and re-assessment of ovarian cancer serum proteomic profiling. BMC Bioinformatics 2003;4:24.

172. Bjorhall K, Miliotis T, Davidsson P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics 2005;5(1):307-17.

APPENDIX A

ELSEVIER COPYRIGHT PERMISSION

Tables 1-5 and Figures 6-10, 12, and 13 have been reprinted from

Schwegler EE, Cazares L, Steel LF, et. al. SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. Hepatology 2005;41(3): 634-42.

The copyright permission policy listed on the websites from the Hepatology Journal was followed.

http://www.elsevier.com/wps/find/supportfaq.cws_home/rightsasanauthor

What rights do I retain as a journal author*?

As a journal author, you retain rights for large number of author uses, including use by your employing institute or company. These rights are retained and permitted without the need to obtain specific permission from Elsevier. These include:

- the right to make copies of the journal article for their own personal use, including for their own classroom teaching use;
- the right to make copies and distribute copies (including through e-mail) of the journal article to research colleagues, for the personal use by such colleagues (but not commercially or systematically, e.g. via an e-mail list or list serve);
- the right to post a pre-print version of the journal article on Internet web sites including electronic pre-print servers, and to retain indefinitely such version on such servers or sites (see also our information on electronic preprints for a more detailed discussion on these points.);
- the right to post a revised personal version of the text of the final journal article (to reflect changes made in the peer review and editing process) on the author's personal or institutional web site or server, with a link to the journal home page (on elsevier.com);
- the right to present the journal article at a meeting or conference and to distribute copies of such paper or article to the delegates attending the meeting;
- for the author's employer, if the journal article is a 'work for hire', made within the scope of the author's employment, the right to use all or part of the information in (any version of) the journal article for other intra-company use (e.g. training);
- patent and trademark rights and rights to any process or procedure described in the journal article;
- the right to include the journal article in full or in part in a thesis or dissertation (provided that this is not to be published commercially);
- the right to use the journal article or any part thereof in a printed compilation of works of the author, such as collected writings or lecture notes (subsequent to publication of the article in the journal); and
- the right to prepare other derivative works, to extend the journal article into book-length form, or to otherwise re-use portions or excerpts in other works, with full acknowledgement of its original publication in the journal.

***Please Note:** *The rights listed above apply to journal authors only. For information regarding book author rights, please contact the* **Global Rights Department**.*

Elsevier Global Rights Department
phone (+44) 1865 843 830
fax (+44) 1865 833 333
email: permissions@elsevier.co.uk

APPENDIX B

MCP COPYRIGHT PERMISSION

Figure 3 and 96 have been reprinted from Drake RR, Schwegler EE, Malik G, et. al Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. Mol Cell Proteomics 2006.

The copyright procedures listed on the Molecular and Cellular Proteomics Journal website was followed.

http://www.mcponline.org/misc/Copyright_Permission.shtml

Copyright Permission Policy

ASBMB Journals

Journal of Biological Chemistry
Molecular and Cellular Proteomics
Journal of Lipid Research
Biochemistry and Molecular Biology Education
ASBMB Today

ASBMB does not charge for and grants use without requiring your copyright permission request for:

- Original authors wanting to reproduce portions of their own work; or to republish their material in not-for-profit formats or venues.
- Students wanting to reproduce or republish their work for educational purposes.
- Students using other authors' material for their theses.
- Reproduction or republication of abstracts only.
- Photocopying up to 5 copies for personal use.
- Non-profit educational institutions making multiple photocopies of articles for classroom use; all such reproduction must utilize institutionally owned equipment for this purpose.

Use of copyrighted material requires proper citation.

VITA

ELIZABETH ELLEN SCHWEGLER

Eastern Virginia Medical School

Department of Microbiology and Molecular Cell Biology

700 W. Olney Road

Norfolk, VA 23507

Old Dominion University

Hampton Boulevard

Norfolk, VA 23529