IN-DEPTH PROTEOMICS AND GLYCOPROTEOMICS BY MASS SPECTROMETRY AND AFFINITY-BASED PLATFORMS FOR SELECTIVELY CAPTURING PROTEINS AND GLYCOPROTEINS FROM BREAST CANCER AND DISEASE FREE HUMAN SERUM

By

SUBHASHINI SELVARAJU

Bachelor of Science in Chemistry
Madras University
Chennai, India
2001

Master of Science in Applied Chemistry
Anna University
Chennai, India
2003

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2011

IN-DEPTH PROTEOMICS AND GLYCOPROTEOMICS

BY MASS SPECTROMETRY AND AFFINITY-BASED

PLATFORMS FOR SELECTIVELY CAPTURING

PROTEINS AND GLYCOPROTEINS FROM

BREAST CANCER AND DISEASE

FREE HUMAN SERUM

Dissertation Approved:

Dr. Ziad El Rassi

Dissertation Adviser

Dr. Richard A. Bunce

Dr. Barry K. Lavine

Dr. Kevin D. Ausman

Dr. Ramanjulu Sunkar

Outside Committee Member

Dr. Mark E. Payton

Dean of the Graduate College

ACKNOWLEDGEMENTS

First of all, I would like to extend my deepest gratitude to my research advisor Dr. Ziad El Rassi, for his guidance throughout my graduate studies and research at Oklahoma State University. I am very much grateful to for his teaching, help and continuous support. I also thank him for his patience with me. I truly believe that Dr. El Rassi has provided me with a sound knowledge for my future career.

I would like to thank Drs. Richard A. Bunce, Barry K. Lavine, Kevin D. Ausman and Ramanjulu Sunkar for agreeing to be on my graduate committee. I also thank Dr. Steve Hartson for his valuable suggestions.

I also thank our previous lab members Yazen Jmeian, Samuel Karenga and Hengwen Zhong for their help and thanks to the present members Dilani Gunasena and Erandi Mayadunne for their friendship and endless support.

A special thanks to my husband Baskar Nammalwar and our lovely son Haresh Baskar. They both have been very patient and helped me to complete my work on time. Without their endless love and support I could have not achieved what I have done.

Finally, I would like to thank my parents Selvaraju Govindasamy and Vetrichelvi Selvaraju for their blessing, endless love, care and constant support in every aspect of my life. I would also like to thank my brother Sivakumar Selvaraju for his care and encouragement. Also, my in-laws love and support during these years is also greatly appreciated.

TABLE OF CONTENTS

Chapter Page

I. BACKGROUND, RATIONALE AND SCOPE OF THE INVESTIGATION 1

Introduction and scope of the study 1
General background information pertinent to the dissertation 4
Immobilized metal ion affinity chromatography 4
Lectin affinity chromatography 5
Protein Equalizer versus immuno-depletion 6
Glycoproteomics: another complex dimension 8
Approaches for proteomics analysis by mass spectrometry 11
Overview of different techniques currently used to reduce the complexity of biological samples – sample treatment 14
Overview of depletion methods 15
Overview of the protein equalizer approach 18
Chromatographic and electrophoretic enrichment/pre-fractionation methods 21
Capturing/targeting specific glycoproteins – capturing/targeting sub-glycoproteomics by LAC and other affinity chromatography 25
Single or serial lectin columns 25
Multi-lectin affinity chromatography with high abundance protein depletion 29
Lectin and other affinity microarrays 32
Rationale of the study 35
Summary 36
References 37

II. REDUCTION OF PROTEIN CONCENTRATION RANGE DIFFERENCE FOLLOWED BY MULTICOLUMN FRACTIONATION PRIOR TO 2-DE AND LC-MS/MS PROFILING OF SERUM PROTEINS 44

Introduction 44
Experimental 46
Instrumentation 46
Reagents and materials 47
Affinity columns and reversed-phase column preparation 47

ProteoMiner™ treated serum....48
Tandem column platform - chromatographic conditions....49
Two-dimensional gel electrophoresis and spot counting....50
LC-MS/MS methodology....51
LC-MS/MS data analysis....52
Results and discussion....53
Overall strategy – combinatorial peptide ligand libraries (ProteoMiner™) followed by column fractionation....53
Evaluation of the ProteoMiner™ and comparison with the depletion approach....59
Highlights of the retention of some typical proteins on the different fractionation columns....70
Conclusions....77
References....78

III. LECTIN AFFINITY CHROMATOGRAPHY USING LECTINS OF BROAD SPECIFICITY FOR CAPTURING SUB-GLYCOPROTEOMICS FROM BREAST CANCER AND DISEASE FREE HUMAN SERA USING TANDEM MONOLITHIC COLUMNS WITH SURFACE IMMOBILIZED CONCANAVALIN A, WHEAT GERM AGGLUTININ AND *RICINUS COMMUNIS* AGGLUTININ-I....80

Introduction....80
Experimental....86
Instrumentation....86
Reagents and materials....86
Monolithic affinity columns....87
Immobilization of lectins....87
Fractionation of glycoproteins from human serum - chromatographic conditions....88
1D SDS-PAGE analysis....89
Protein assay....89
LC-MS/MS methodology....89
LC-MS/MS data Analysis....89
Results and discussion....90
Evaluation of the monolithic lectin columns....90
Effect of the serial order of the lectin columns in the tandem format....96
LC-MS/MS analysis of the lectin enriched fractions from cancer and disease-free sera....103
Specificity of the lectin columns....113
Visual differential expression of proteins using SDS-PAGE analysis of the cancer and the disease-free sera....114

Chapter Page

Spectral count as a means to assess the differentially expressed proteins......115
Hemoglobin subunit α and β......120
Conclusions......120
References......122

IV. LECTIN AFFINITY CHROMATOGRAPHY USING NARROW SPECIFICITY LECTINS FOR CAPTURING FUCOSYLATED AND SIALYLATED GLYCOPROTEINS FROM BREAST CANCER AND DISEASE FREE HUMAN SERA USING TANDEM MONOLITHIC COLUMNS WITH SURFACE IMMOBILIZED *ALEURIA AURANTIA* LECTIN, *SAMBUCCUS NIGRA* AGGLUTININ AND *LOTUS TETRAGONOLOBUS* AGGLUTININ......126

Introduction......126
Experimental......129
Instrumentation......129
Reagents and materials......129
Monolithic affinity columns......129
Immobilization of the lectins......129
Fractionation of glycoproteins from human serum - chromatographic conditions......130
Protein assay......130
LC-MS/MS Methodology......130
LC-MS/MS data analysis......130
Results and Discussion......131
Analysis of proteins in the lectin fractions......131
Identification of differentially expressed proteins in the lectin fractions using MS spectral count......141
Fibrinogens......141
Differentially expressed proteins captured by the LTA column......142
Differentially expressed proteins captured by the AAL column......144
Differentially expressed proteins captured by the SNA column......145
Narrowing on candidate biomarkers – a panel of biomarkers......148
Conclusions......151
References......152

V. TANDEM LECTIN AFFINITY CHROMATOGRAPHY FOR IN-DEPTH GLYCOPROTEOMIC ANALYSIS BY COMBINING NARROW AND BROAD SPECIFICITY LECTINS......156

Introduction......156

Chapter Page

Experimental 157
Results and discussion 158
Overall strategy 158
Analysis of the proteins captured by the lectin columns 159
Benefits of combining broad and narrow specificity lectins in a six tandem column format 168
Some of the clinically important proteins identified in the fractions captured by the lectin columns 175
Access to a wider protein concentration range 177
Conclusions 178
References 179

LIST OF TABLES

Table Page

CHAPTER II

1. LC-MS/MS results of the identified proteins from the equalized human serum without subsequent fractionation 54

2. LC-MS/MS results of the proteins identified from the solution digest of fractions from IMAC and RPC columns 61

3. Compilation of identified proteins unique to ProteoMiner™-post fractionation treated serum, identified proteins unique to depletion-post fractionation treated serum and identified proteins common to both ProteoMiner™- and depletion-post fractionation treated serum....71

CHAPTER III

1. Comparison of number of proteins that were identified on differently ordered serial lectin columns 97

2. List of proteins that were unique to WGA-CON A-RCA-I series 97

3. List of proteins that were unique to CON A–WGA-RCA-I series....98

4. List of proteins that were unique to RCA-I-CON A-WGA series 98

5. List of proteins that were common to all the three series....99

6. List of proteins identified in the lectin-bound fractions 104

7. Proteins that were differentially expressed in WGA fraction 116

8. Proteins that were differentially expressed in Con A fraction 118

9. Proteins that were differentially expressed in RCA-I fraction 119

Table Page

Chapter IV

1. List of proteins identified in lectin-bound fractions 132

2. Proteins that were differentially expressed in LTA fraction 143

3. Proteins that were differentially expressed in AAL fraction................................145

4. Proteins that were differentially expressed in SNA fraction 147

5. List of common proteins that were differentially expressed in both broad and narrow specificity lectins 150

Chapter V

1. List of proteins identified in the lectin-bound fractions 160

2. Proteins that were identified in WCR columns only in the combined LAS-WCR columns..174

LIST OF FIGURES

Figure Page

Chapter I

1. The three types of *N*-linked glycoproteins with the common core structure shown in the dashed box 9

2. Four most common types of core structures of O-linked glycoproteins 10

Chapter II

1. 2-DE electropherograms of the serum after treatment with the hexapeptide beads and fractionation/concentration on (A) IDA-Zn^{2+}, (B) IDA-Ni^{2+}, (C) IDA-Cu^{2+} and (D) the RPC-column 68

Chapter III

1. Affinity of immobilized WGA toward *N*-glycan structures....................84

2. Affinity of immobilized Con A toward *N*-glycan structures....................85

3. Sugars binding specificities of immobilized RCA-I 85

4. Chromatograms of myoglobin (A) and α-1-acid glycoprotein (B) injected into the WGA column 92

5. Chromatograms of myoglobin (A) and peroxidase (B) injected into the Con A column....................93

Figure Page

6. Chromatograms of myoglobin (A) and transferrin (B) injected into the RCA-I column....94

7. Elution profile of the bound proteins to WGA column (100 mm × 4.6 ID) when 50 μL of diluted serum (1:3 ratio) was injected....95

8. Elution profile of the bound proteins of WGA column (100 mm × 4.6 ID) when 250 μL of diluted serum (1:3 ratio) was injected....96

9. Chromatographic setup for the concentration of the serum proteins using tandem lectin columns....100

10. Venn diagram showing the number of proteins identified in each series....101

11. Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the disease-free serum....111

12. Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the breast cancer serum....112

13. Venn diagram for the differentially expressed proteins found in the lectin fractions....113

14. Analysis of bound glycoprotein fractions from the tandem lectin series on 1D SDS-PAGE....115

Chapter IV

1. Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the disease-free serum....140

2. Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the breast cancer serum....140

3. Venn diagram for the differentially expressed proteins found in the lectin fractions....142

Chapter V

1. Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the narrow specificity lectin columns....169

Figure Page

2. Venn diagram showing the number of proteins in common and the number of proteins unique to each lectin captured from the broad specificity lectin columns 169

3. Venn diagram showing the number of proteins in common and unique to narrow and broad specificity lectin columns 170

4. Comparison of the WCR series alone to the WCR series in the combined LAS-WCR series 170

5. Comparison of the LAS series alone to the LAS series in the combined LAS-WAR series 171

6. Venn diagram showing the distribution of the unique proteins of the WCR series among the three lectins 173

LIST OF ABBREVATIONS

1D	1-Dimensional
2D	2-Dimensional
2-DE	2-Dimensional gel electrophoresis
AAL	*Aleuria aurantia* lectin
ACN	Acetonitrile
AIBN	2,2-Azobis(isobutyronitrile)
CE	Capillary electrophoresis
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propane sulfonate
Con A	Concanavalin A
CPLL	Combinatorial peptide ligand libraries
DCIS	Ductal carcinoma *in situ*
DIGE	Diffrential gel electrophoresis
dPC	Digital Proteome chip
DS	Differential solubilization
DTT	Dithiothretiol
EDMA	Ethylene glycol dimethacrylate
FDA	Food and drug administration
FFE	Free flow electrophoresis

FITC	Fluorescein isothiocyanate
GlcNAC	*N*-Acetyl glucosamine
GMA	Glycidyl methacrylate
GMM	Glyceryl methacrylate
GPI	Glycosylphosphatidylinositol
HAP	High abundance proteins
HCC	Heptocellular carcinoma
His	Histidine
HMW	High molecular weight
HPLC	High-performance liquid chromatography
ID	Inner diameter
IDA	Iminodiacetic acid
IMAC	Immobilized metal affinity chromatography
IPG	Immobilized pH gradient
JAC	Jacalin
LAS	Lectin affinity chromatography
LC	Liquid chromatography
LCA	*Lens culinaris* agglutinin
llama	Lama glama
LMW	Low molecular weight
L-PHA	L-Phytohemagglutinin
LTA	*Lotus tetragonolobus* agglutinin

MAL	*Maackia amurensis* lectin
MALDI	Matrix-assisted laser desorption ionization
Micro-sol	Microscale solution
MIPS	Monoisotopic precursor selection
M-LAC	Multi-lectin affinity chromatography
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NPS	Non-porous silica
NSI	Nanospray ionization
OGE	Off-gel electrophoresis
PNA	Peanut agglutinin
PTM	Post-translational modification
RAM	Restricted access material
RCA-I	*Ricinus communis* agglutinin-I
RPC	Reversed phase chromatography
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel
SELDI	Surface-enhanced laser desorption/ionization
sLe^x	Sialylated Lewis x
SNA	Sambuccus nigra agglutinin
SSB	Solid state buffer
TED	Tris(carboxymethyl)ethylenediamine
TFA	Trifluoroacetic acid

TOF	Time of flight
Tris	Tris (hydroxymethyl)aminomethane
WGA	Wheat germ agglutinin

CHAPTER I

BACKGROUND, RATIONALE AND SCOPE OF THE INVESTIGATION

Introduction and Scope of the Study

Proteomics is the study of proteins present in a biological sample such as serum, urine, cerebrospinal fluid etc., which involves identification, quantification and characterization of the proteins present in the biological sample. Proteomics is extensively involved in the analysis of body fluids to detect disease biomarker proteins or any other proteins that are differentially expressed in response to a given disease or treatment. Human serum is one of the body fluid that reflects the physio-pathological state of person and it can be easily accessed for clinical studies. The vast complexity of serum arises primarily from its high dynamic concentration range that spans over 10 to 12 orders of magnitude. Also, the presence of a large number of proteins especially those present at very low levels further complicates the analysis of serum. Different techniques such as depletion of the high abundance proteins, electrophoretic and chromatographic pre-fractionation and equalization of the protein concentration have been reported to reduce the complexity of serum. These diverse techniques along with the advancement in mass spectrometry (MS) have allowed gaining increasing knowledge about many

biological samples. The major purpose of the work presented in this dissertation is to contribute to the introduction and evaluation of new ways and strategies for reducing the complexity of the human serum and in turn contribute to facilitating the identification of candidate biomarkers with the assistance of advanced MS techniques. In one approach, immobilized metal affinity chromatography (IMAC) along with a protein equalizer technique were combined to reduce the complexity of the human serum. In another approach, lectin affinity chromatography (LAC) was used to selectively enrich the glycoproteins present in human serum. The LAC strategy was further demonstrated for its promising biological potentials by comparing a disease-free serum with a breast cancer serum to identify the glycoproteins that were differentially expressed due to changes in their glycosylation pattern in cancer serum.

This chapter gives a brief introduction to IMAC, LAC and glycoproteins analysis. This is followed by providing (i) short reviews of the different techniques that are currently used to address the complexity of a biological sample and (ii) an overview of different strategies used in identification of glycoprotein biomarker candidates in various diseases, particularly in cancer.

Chapter II describes a strategy where the reduction of the protein dynamic concentration range in serum was obtained using protein equalizer beads. The "equalized" proteins were further fractionated on IMAC and reversed phase chromatography (RPC) columns arranged in tandem format. The fractions obtained from the IMAC and RPC columns were analyzed using liquid chromatography (LC)-MS/MS. This strategy allowed for identification of a greater number of proteins without depleting the high abundance proteins (HAP).

Chapter III investigates the potentials of three broad specificity lectins to enrich the glycoproteins present in human serum using tandem lectin affinity monolithic columns. By using the broad specificity lectins the most common type of *N*-linked glycans namely, high-mannose, complex and hybrid type glycoproteins and *O*-linked glycoproteins can be captured. This enrichment of the glycoproteins allowed reducing the sample complexity and led to the identification of many glycoproteins using LC-MS/MS. This strategy was applied to profiling the differentially expressed proteins in breast cancer serum with respect to disease-free serum.

Chapter IV focuses on selective enrichment of sialylated and fucosylated glycoproteins. Alteration of sialylation and fucosylation of glycoproteins is a common feature in the case of many diseases. Therefore, three lectins, which have affinity toward sialylated and fucosylated glycoproteins, were immobilized on macroporous monolithic columns and they were arranged in series (i.e., tandem columns) to capture the glycoproteins present in the serum. This strategy to capture targeted glycoproteins was also applied to breast cancer serum to identify the differentially expressed proteins in comparison to the disease-free serum.

Chapter V involved the combination of both broad and narrow specificity lectins for capturing a broad spectrum of glycoproteins in a single run. The lectin columns were arranged in tandem series, where the narrow specificity lectins were followed by the broad specificity lectins. In this arrangement, the binding sites of the various lectins are in principle more evenly exploited so that biding site overloading would be minimized and more efficient glycoprotein capturing may be achieved.

General Background Information Pertinent to the Dissertation

Immobilized Metal Ion Affinity Chromatography

In the year 1975, Porath and co-workers introduced IMAC by immobilizing zinc and copper ions to fractionate human serum proteins [1]. In IMAC, the underlying principle is the binding of the proteins to the metal chelated to the stationary phase such as iminodiacetic acid (IDA)-agarose. Even though cysteine and tryptophan residues on the protein surface can interact with the immobilized metal ion, it has been shown that histidine (His) on the surface of a protein is greatly involved in the protein-metal interaction [2, 3]. For IMAC, the monolithic columns used in this dissertation were bonded with iminodiacetic acid (IDA) chelated with a given transition metal ions (e.g., Zn^{2+}, Cu^{2+}, Ni^{2+}). If a given protein has two-vicinal His residues or two or more His on the surface then it is retained by the IDA-Zn^{2+} column; for a protein to be retained on an IDA-Ni^{2+} column it should contain at least two His residues on the surface and the presence of even one His residue on the surface of a protein makes it retained on a IDA-Cu^{2+} column [3]. Thus, IMAC can offer an effective fractionation of serum proteins based on the number of His residues exposed on the surface of the protein solute. This approach was used to fractionate serum proteins in Chapter II. The applications of IMAC in protein fractionation in general and in proteomics samples in particular can be found in the recent review article by Block *et al.* [4]. Also, fractionation of histidine-tagged recombinant proteins are discussed in detail.

Lectin Affinity Chromatography

Lectin affinity chromatography (LAC) involves the specific interactions between the immobilized lectins on a given support and carbohydrates and glycoconjugates such as glycoproteins, glycolipids and glycopeptides. In fact, lectins are sugar-binding proteins, which are of non-immune origin. Since lectins are capable of differentiating subtle changes in a variety of glycoforms, they have found extensive applications in the field of glycoproteomics. Depending on their specificity, lectins are classified as either broad or narrow specificity ligands. Both lectin types have been used for capturing specific glycoproteins present in biological fluids and tissues [5-7]. Although sugars have low affinity towards lectins (K_d = 10^2 - 10^6 M^{-1}), as compared to carbohydrate-specific antibodies (K_d = 10^4 - 10^8 M^{-1}), lectins have the advantage that they are readily available and the elution conditions from a given lectin column do not require any harsh conditions such as low pH elution as in the case of antibodies [8]. The use of lectins in cancer biomarker discovery has been recently reviewed by Kim *et al*. [9].

Lectins bound to agarose, silica, monoliths, membranes, polyhydroxylated polymer (POROS) support or magnetic beads have been used in many studies for theenrichment of glycoproteins [10, 11]. Monoliths, which are rigid continuous macroporous separation media have found use in proteomics studies [12, 13] due to their many advantages such as good mass transfer, low back pressure and ability to scale up or down the preparation process, stability, etc. Also, monoliths having different pore sizes with nonspecific binding can be readily prepared. Monolith based supports find applications in different chromatographic separation modes such as reversed-phase, ion-exchange, hydrophobic interaction, immobilized metal ion affinity, bioreactors, affinity

chromatography etc., [14]. The work presented in this dissertation involves polymethacrylate based monolithic supports for IMAC and LAC.

Protein Equalizer versus Immuno-depletion

Protein equalizer technology (ProteoMiner™) is a novel technique that can simultaneously decrease the HAP concentration and increase the low abundance protein concentration. Thus, it can reduce the dynamic concentration range of proteins in many biological samples or fluids. The Protein equalizer consists of peptide ligand libraries that are synthesized *via* a modified Merrifield approach using the "separate-recombine-assemble" method [15]. In brief, a batch of millions of microscopic, porous chromatographic beads is split into different batches of equal parts. Then a given amino acid is chemically attached to the beads present in a given batch. Different batches of beads are bonded with different amino acids. These batches are recombined and split again for further attachment of amino acids. Using this method, peptide ligand libraries are prepared [16]. By this process each bead would contain copies of single, unique amino acid sequence [17]. In a study, different amino acid lengths were investigated for capturing red blood cell cytoplasmic proteins. It was observed that at least four amino acids are required for selectivity of the beads and when the amino acid length was six, the largest possible population of proteins was captured [18].

The ProteoMiner™ technology works on the principle of affinity chromatography. Under large overloading conditions of a given biological sample, the proteins are captured by the specific hexapeptide beads. Proteins that are present at high concentration rapidly saturate the beads whereas the low-abundance proteins will not

saturate immediately but are gradually captured by the beads under the overloading condition of the sample. Then, the unbound proteins are washed away to ensure that all the non-captured proteins are completely eliminated and the bound proteins are eluted using appropriate buffers [16]. The adsorption of the proteins onto the beads are *via* combination of interacting forces such as ion-exchange interactions, hydrophobic associations, hydrogen bonding, structural docking and Van der Waals interactions [16].

On the other hand, depletion of HAP is a commonly practiced approach to achieve in-depth analysis of proteome. In this approach, specific HAP are removed using polyclonal antibodies, which can recognize many regions of a target and thus efficiently deplete them. Even though depletion techniques are reproducible and efficient in removing HAP, the presence of medium abundance proteins still masks the low abundance proteins. Also, there have been reports that depletion can result in co-depletion of many valuable low abundant proteins [19, 20]. Unlike the depletion approach, the protein equalizer technology does not deplete any proteins but instead it equalizes the concentration of the proteins present in a sample. The high sensitivity of the protein equalizer was demonstrated in a recent report where even 1 μg casein per liter present in white wine was detected using the protein equalizer technology [21]. This represented a 200 fold increase in sensitivity as compared to the traditional enzyme-linked immunosorbent assay test which had a sensitivity of 200 μg casein per liter [22]. Hence, the use of the protein equalizer in serum analysis to reduce the dynamic concentration range should allow for in-depth proteomic analysis. This technique is utilized in Chapter II and compared to immuno-depletion.

Glycoproteomics: Another Complex Dimension

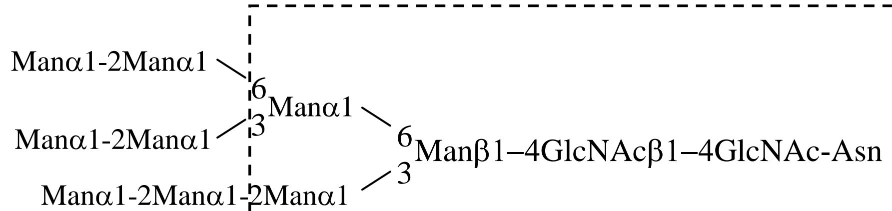
Apart from the wide dynamic concentration range of proteins, the human proteome is further complicated as a result of more than 100 post-translational modifications (PTMs) taking place in the proteome [23]. At the later stage of its biosynthesis, a given protein is subjected to covalent PTMs such as glycosylation, acetylation, phosphorylation, oxidation, methylation etc. Glycosylation is one of the most common PTMs, which plays a key role in many of the biological processes that include immune defense, cell growth and cell-cell adhesion [24, 25]. Glycosylation is a process in which addition of carbohydrates to proteins takes place in the presence of a series of enzymes [26].

Glycoproteins are glycosylated proteins, i.e., they have glycans (carbohydrates) covalently attached to them. The different types of glycans, that vary in their structures or branching types, result in various forms of glycoproteins and they are called glycoforms. There are three major types of glycoproteins namely, glycosylphosphatidylinositol (GPI) anchors, *N*-linked and *O*-linked glycoproteins. When a protein contains a GPI sequence, GPI anchors are covalently attached to a fully folded protein. In *N*-linked glycoproteins, the carbohydrate is attached to the nitrogen of the amide of an asparagine that is present in the amino acid sequence of a protein. If the carbohydrate is attached to the oxygen of a serine or threonine residue of a protein then it is termed as *O*-linked glycoproteins [24]. *N*-Linked glycans are divided into three types (i) high-mannose, (ii) complex and (iii) hybrid type (see Fig.1) The *N*-linked glycans have a similar core structure that comprises two *N*-acetylglucosamine (GlcNAc) and three mannose units (see Fig.1), found in common to all the three types (indicated by a

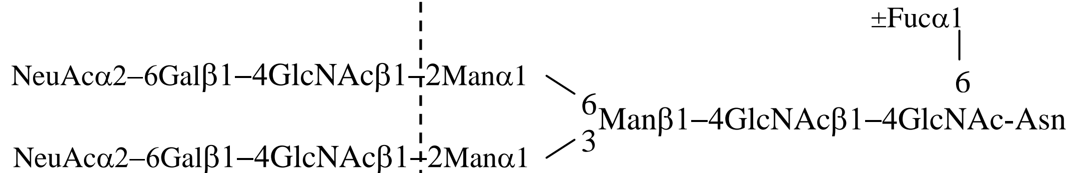
box in Fig. 1). Unlike the *N*-linked glycoproteins, the *O*-linked glycoproteins have at least 8 core structures out of which 4 are the most common types (see Fig. 2).

Over the years, many studies have established that glycoproteins are involved in a diverse array of disease conditions such as (i) infectious diseases like viral, bacterial and parasitic infections, (ii) inflammation, (iii) immune deficiency, (iv) cancer metastasis, (v) rheumatoid arthritis, (vi) inherited disorders and (vii) abnormal catabolism of

(i) High-mannose type



(ii) Complex type



(iii) Hybrid type

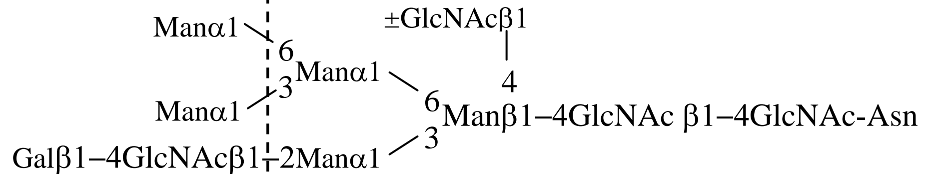


Figure 1. *The three types of N-linked glycoproteins with the common core structure shown in the dashed box.*

(i) Core 1 Galβ1-3GalNAcα1-Ser/Thr

(ii) Core 2 GlcNAcβ1-6(Galβ1-3)GalNAcα1-Ser/Thr

(iii) Core 3 GlcNAcβ1-3GalNAcα1-Ser/Thr

(iv) Core 4 GlcNAcβ1-6(GlcNAcβ1-3)GalNAcα1-Ser/Thr

Figure 2. *Four most common types of core structures of O-linked glycoproteins*

glycoconjugates [27]. Although glycoproteins are involved in many diseases, identification of glycoprotein biomarkers for cancer is an extensively active field of research.

Cancer is the second most common cause of death in the world and detection of cancer biomarkers at an early stage is a crucial goal. Some of the glycoproteins that have been approved as cancer biomarkers by the Food and Drug Administration (FDA) are carcinoembryonic antigen, prostate-specific antigen, cancer antigen 125, alpha-fetoprotein and Her-2/neu [9]. The majority of glycoproteins are synthesized in the liver and they are circulated in serum/plasma, urine and saliva [28]. Since it is understood that proteins can leak from tumor tissue into the blood [29] and as much as 50% of the human serum proteins are glycosylated [30], it is then of great importance to identify, quantify and characterize the glycoproteins present in serum. Glycoproteomics and its importance in diseases such as in breast cancer are the topics of Chapters III, IV and V.

Approaches for Proteomics Analysis by Mass Spectrometry

The two major approaches practiced in proteomics MS for protein identification and characterization involve either the fractionation of the intact proteins or their peptide fragments, which are obtained by enzymatically digesting (e.g., with trypsin) the complex proteome. In one approach, the tryptic peptides are further fractionated using different separation strategies and finally analyzed through MS to identify the parent proteins. Whereas in the second approach, the intact proteins are fractionated using various separation approaches and finally digested into peptides and analyzed through MS to identify the parent proteins. Some of the techniques in which the peptides are fractionated using different separation methods are discussed below and those techniques that involve fractionation of intact proteins are discussed in the next section.

Due to the high complexity of biological samples, a single step fractionation (i.e., single column) would not be sufficient enough to comprehensively analyze complex samples, and therefore multiple fractionation steps are required to analyze the proteome. In 1-dimensional (1D) LC, the most commonly used method for separation is based on the difference in hydrophobicity of the proteins using RPC columns [31]. In 2-dimensional (2D) LC, the commonly practiced separation approach involves strong cation exchange (SCX) in the first dimension and RPC in the second dimension [32, 33]. Isoelectric focusing [34] and hydrophilic interaction chromatography [35] have also been used as the first separation dimension in 2D LC.

In a study by Gilar *et al.* [36], a comparison between 1D and 2D-LC analysis was carried out using whole human serum (i.e., non depleted serum). In the 2D-LC, the first dimension separation was performed using either a SCX column or an RPC column, and

the second dimension separation involved the use of an RPC column. As one would expect, only 52 proteins were identified in the 1D-LC study whereas 184 and 142 proteins were identified in the RPC-RPC and SCX-RPC 2D-LC respectively, clearly indicating that the 2D-LC is necessary to reduce the complexity of the biological sample. The SCX-RPC identified mostly 7-13 long amino acid peptides whereas, the RPC-RPC identified peptides longer than 20 amino acids. The detection of longer and hydrophobic peptides by RPC-RPC indicated that these peptides were lost in the SCX separation. The authors concluded that the SCX-RPC and RPC-RPC methods were not complementary and the latter method was more suitable to hydrophobic peptides.

As stated above, in many of the multidimensional LC set ups used in proteome analysis, the first separation dimension of the tryptic peptide digest involves commonly the use of SCX chromatography. In a recent investigation, where shotgun proteome analysis of the mammalian nuclear cell lysate was carried out using 2D-LC, the traditionally used SCX chromatography in the first dimension was replaced by a mixed mode reversed-phase anion exchange (MM RP-AX) and the second dimension separation was done using RPC. The authors made a comparison between the MM RP-AX and the SCX analysis based on the peptide fractionation efficiency, distribution of peptide charge state, p*I* and hydrophobicity of the identified proteins and it was concluded that MM RP-AX offered an effective alternative to the conventionally used SCX chromatography. For example, the number of unique peptides and proteins increased by 150% when MM RP-AX was used in the first dimension. Another advantage of this method is that the use of gradient acetonitrile to elute peptides from the first dimension eliminated any desalting steps before the second dimension. Also, potential ion suppression and salt precipitation

that lead to harmful effects in RP-LC-MS/MS analysis which is usually observed in the case of SCX, was not observed in MM RP-AX [37].

In another 2D-LC study, low pH (pH 3) and high pH (pH 10) RPC concatenation fractionation approach of the tryptic peptides was performed as an off-line first dimension separation. In this proteome analysis, the high pH and low pH concatenation strategy were evaluated using human MCF10A cell sample. From the first dimension 60 fractions were collected using either high-pH or low-pH RPC. The concatenation was done by pooling early, middle and late RPC fractions of 60 total collected fractions whereby for instance the fractions 1, 16, 31 and 46 were combined together, and the fractions 2, 17, 32 and 47 were also similarly combined and so on. The high pH concatenation resulted in increase of the number of peptide and protein identification by 1.8 and 1.6 fold, respectively, when compared to the traditional SCX chromatography, whereas the low pH concatenation yielded results that were comparable to the SCX results. This indicates that concatenation strategy using high-pH RPC in the first dimension is a better alternative to the traditionally used SCX separation [38].

In a very recent investigation by Krishnan *et al*. [39], the first dimension separation was based on OFFgel electrophoretic (OGE) fractionation which allowed them to identify 1373 proteins from human platelet proteome. The second dimension analysis was achieved by RPC fractionation and the proteins were identified by nano-LC-MS/MS. In the OGE, the peptides migrate according to their pI values on an Immobiline™ DryStrip. After focusing, the samples were recovered from the sample wells and then loaded into the RPC column for MS/MS detection. This shotgun proteome analysis by

OGE fractionation and RPC fractionation in the first and second dimension allowed them to perform in-depth analysis of the human platelet proteome.

When a complex proteome mixture is enzymatically digested, it results in enormous number of peptides. As mentioned above, there are various methods available for the separation of these peptides, but still it is a major problem as too many peptides are involved in the fractionation process. Also, fractionation of these peptides based on their properties does not give much information about the nature of the protein. But, when fractionation is done at the protein level more information about the nature of the protein could be obtained. For example, when protein fractionation is done using IMAC (as shown in Chapter II), one could assess the number of His residues present on the surface of the protein or when the fractionation is done using LAC (as described in Chapters III, IV and V), information about the glycoforms of the glycoproteins could be obtained. Thus, the research work discussed in this dissertation involves fractionation at protein level.

Overview of Different Techniques Currently Used to Reduce the Complexity of Biological Samples – Sample Treatment

Proteomics analysis and profiling is a problem of sample preparation in the first place. Most of the techniques involved in proteomics involve two major areas: (i) protein separation/sample preparation and (ii) protein identification by MS. In order to analyze a proteome mixture, it is necessary to fractionate the complex proteome based on their characteristics such as solubility, p*I* value, molecular weight, hydrophobicity, etc. The various fractionation steps commonly practiced include depletion of the proteins, protein

equalizer technology and chromatographic and electrophoretic separation. The extensive growth in MS techniques along with various novel fractionation strategies have allowed for identification of increased number of proteins. In this Chapter, articles published on various fractionation techniques used in proteomics over the years 2009-to present have been reviewed. Prior to 2009, there were 3 major review articles that dealt with various aspects of fractionation/sample preparation for proteomics analysis [40-42]. One of these review articles [40] discussed the liquid-based separation system employed for in-depth proteomic analysis for the time period 2002-2009. Another review article [41] described sample preparation and fractionation associated problems along with different strategies used for cancer biomarker discovery. The review article [42] gave an overview about different technologies used in sample preparation and fractionation for biomarker discovery for the time period of 2004-2007.

Overview of Depletion methods

Depletion is one of the commonly used methods to reduce the complexity of a given proteomics sample despite reports that some of the depletion methods result in co-depletion of many other clinically important low abundant proteins [19]. Some of the depletion method has been directed towards either high molecular weight (HMW) proteins or HAP. Many other depletion methods and precipitation methods reported prior to 2009 have been reviewed by Jmeian and El Rassi [40].

Solvent precipitation methods Some of the low molecular weight (LMW) proteins that are present in plasma/serum reflect the pathological state and they could

serve as potential biomarkers [43]. Thus, the identification of these LMW proteins gains importance in proteome analysis. Kawashima *et al.* [44] described a method to concentrate the LMW proteins/peptides from serum using differential solubilization (DS) method. In the DS method, LMW proteins were isolated by diluting the serum with denaturing solution containing urea, thiourea and dithiothretiol (DTT), followed by centrifugation, and acidic treatment. They were able to analyze quantitatively more than 1500 LMW proteins/peptides from 1 μL of serum by combining the DS method with RPC separation followed by matrix assisted laser desorption/ionization (MALDI)-time of flight (TOF)-MS. This method was compared to other depletion techniques such as organic precipitation, ultrafiltration and albumin/IgG affinity removal methods. It was found that many peptides including those bound to albumin were observed in the DS method, but not the other three methods.

Warder *et al.* [45], described a protocol to precipitate the HAP which included albumin and transferrin. In brief, the serum was incubated with either DTT or tris(2-carboxyethyl)phosphine, and centrifuged to obtain the pellet which was rich in albumin. Analysis of the supernatant by MALDI-TOF-MS did not show any intact albumin ions indicating that albumin was completely precipitated. This reproducible method could be extended to plasma from other species and can also be scaled up to a larger volume of sample.

<u>Immunoaffinity depletion methods</u> Immunoaffinity methods involve the use of specific antibodies to capture one or a group of proteins. There are many commercially available immunodepletion columns to deplete one, six, twelve, fourteen or twenty HAP.

In a study done by Tu *et al.* [46], the performance of the commercially available columns to deplete 7 or 14 HAP was evaluated on the shotgun proteomic analysis of human plasma. It was observed that the depletion columns were highly reproducible, but a few non-targeted proteins were also captured by the depletion columns. Immunodepletion of top 7 or 14 proteins resulted in 25% increase in identified proteins as compared to unfractionated plasma.

In a recent study, a one-step process was reported to concentrate, purify and deplete albumin from urine. In brief, the urine proteins were first reduced, alkylated and transferred to a spin filter, where it was treated with anti-human serum albumin. The proteins were incubated and spun down to remove the waste. The supernatant solution containing the proteins were further tryptically digested, fractionated either using Off-gel electrophoresis or high pH RPC separation and analyzed by LC-MS/MS. By the Off-gel electrophoresis fractionation, 703 proteins were identified whereas using the RPC separation 499 proteins were identified. This simple, efficient and reproducible method is compatible with diverse down-stream applications and is also a potential method to study other complex body fluids [47].

A comparative study was made to see the outcome when one, six, twelve or twenty major proteins were depleted [48]. When the number of depleted proteins was increased from twelve to twenty proteins the benefits were limited, whereas when 6 proteins were depleted many low abundant proteins were detected [48]. In another study [49], 12 highly and 77 moderately abundant proteins present in serum were depleted using commercially available antibody columns. They were able to identify 222 and 71 proteins when 77 and 12 proteins were depleted, respectively. Clearly, this indicates that

the simultaneous depletion of the high and the moderate abundant proteins would increase the number of identified low-abundant proteins. These findings were in good agreement with the results presented in ref. [50] in which the same approach was followed to reduce the complexity of the proteome. The authors were able to establish the above mentioned approach to identify the differentially expressed proteins in ovarian cancer sera. It is to be noted that despite a simultaneous depletion of both high and moderate abundance proteins was performed on the sera, the flow-through fractions still contained some of the moderately abundant proteins, indicating that the depletion was not complete [49, 50]. Also, depleting the 77 moderately abundant proteins might have a disadvantage that it can result in co-depletion of some of the clinically important proteins.

Overview of the Protein Equalizer Approach

In principle, the combinatorial peptide ligand library (CPLL) reduces significantly (i.e., almost equalizes) the dynamic concentration range of proteins present in a given complex biological sample. In fact, under overloading conditions, and when a given sample of complex proteome is treated with the CPLL beads, HAP immediately saturate the corresponding ligands whereas the proteins that are present at very low level gets selectively enriched by the corresponding ligands [18]. The concept of peptide ligand libraries was combined with other fractionation methods such as differential gel electrophoresis (DIGE) and off-gel fractionation [51, 52], to identify a high number of low abundant proteins. The protein equalizer technology has allowed in-depth proteomics analysis of a complex mixture of human proteome [53]. It has also allowed

for in-depth analysis of animal plasma proteome as it does not require any antibodies that are usually specific for human plasma proteins [54].

In a novel method, the equalized proteins were further fractionated based on their differences in isoelectric points using solid-state-buffers (SSB) associated with cation exchangers [51]. In earlier methods, the equalized proteins, which resulted from treating a given serum sample with the CPLL beads, were further fractionated by 1D gel analysis. This procedure yielded a number of fractions to be analyzed. To avoid this cumbersome process, protein equalizer in combination with SSB method was proposed, whereby the equalized proteins are allowed to adsorb on a solid phase where their net charge is opposite to the ion exchange column. This proposed method, which reduced the number of fractions to be analyzed, was compared to the performance of the classical anion exchange chromatography. It was observed that the eluted fractions from SSB method had different ranges of isoelectric points, while the anion exchange chromatography did not show a good discrimination of the isoelectric points. When this method was compared to the performance of Off-Gel fractionation (p*I* based fractionation in off-gel format) after treating the serum with CPLL beads, it was observed that the SSB method detected more number of proteins spots in 2D gel electrophoresis (2-DE), although the Off-Gel method resulted in better p*I* discrimination.

In another study [55], three different methods including high abundance protein precipitation, restricted access materials (RAM) combined with IMAC and CPLL beads were evaluated to see which of these methods serves as a best fractionation step for the analysis of LMW proteins. The evaluation was performed based on the peptide/protein peaks generated from surface enhanced laser desorption/ionization-TOF-MS analysis and

on the reproducibility of the methods. Even though the authors concluded all the three methods were complementary, the peptide affinity beads efficiently depleted the HMW proteins. The IMAC-RAM method identified some additional LMW protein peptides whereas the precipitation method using organic solvents did not give any new information on the peptide/protein peaks.

In another study [56], three different commercially available protein enrichment methods were compared to see if combining these methods would allow access to the low-abundant proteins. The three different methods were immunodepletion using Seppro IgY14 (contained polyclonal antibodies raised against the 14 highest abundance proteins), a two-step immunodepletion process using Seppro IgY14 and Seppro IgY-Supermix system (that contained a mixture of antibodies raised against the proteins present in the flow-through fraction of IgY12) and the third strategy involved the use of the CPLL beads. When the bound fractions were analyzed using 2-DE, as expected differences in the protein patterns were observed for the three methods. In another experiment by the authors, the flow through from IgY14 column was treated with CPLL beads. In principle, combining two different fractionation methods should allow for identification of more number of proteins. In contrast, it was observed that the multi-step fractionation showed only slight increase in the sensitivity as compared to the one-step fractionation. It was observed that the one-step fractionation using the CPLL beads and the IgY14 fractionation increased the number of genes in 2-DE as compared to the unprocessed plasma. The author concluded that the combination of the two different fractionation methods did not show any significant increase in the number of genes in 2-

DE, but it only made the whole process very expensive leading to few thousand dollars per sample.

Chromatographic and Electrophoretic Enrichment/Pre-fractionation Methods

Chromatographic pre-fractionation using one dimension A method to deplete the high abundance proteins was proposed based on hydrophobic interaction chromatography (HIC) [57]. 56 main plasma proteins were divided into three different clusters as high, medium and low hydrophobicity based on their average hydrophobicity. Some of the highly abundant proteins fell in the category of medium hydrophobicity proteins, which represented 70% of the highly abundant proteins. This method was compared with the immuno-affinity depletion of albumin. It was concluded that HIC increased the number of detected spots on 2-DE by 80% when compared to the albumin immunoaffinity depletion.

Another method to enrich the LMW proteins present in serum was introduced by Wu. e*t al.* [58]. Serum was treated with a C_{18} absorbent with an average pore size of 100 Å under denaturing conditions using urea and DTT. After incubation and washing, the LMW proteins were eluted using 60% v/v acetonitrile solution. The eluted proteins were subjected to sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) analysis and the gel bands were further analyzed by LC-MS/MS. In the presence of the denaturing conditions, the protein-protein interactions were broken, thus allowing the identification of the proteins that were bound to the HMW proteins. Since, the procedure does not involve any addition of salts, desalting steps were avoided before analyzing the samples by 2-DE. Also, more than 110 LMW proteins were identified from the serum and the

eluted LMW fraction contained only 5% of the HMW proteins. In addition to serum proteins, this method was also demonstrated in analyzing proteins in cell and tissue extracts.

Multidimensional chromatographic pre-fractionation In an study by Cellar *et al.* [59], two-dimensional separation was carried out for sample enrichment and fractionation for mammalian proteomics. In the first dimension, IgY immunodepletion column was used to deplete the HAP from the sample. The HAP deprived sample was then passed through a trap cartridge which serves as an injection loop for the on-line second dimension. The second dimension column included an analytical C18 column with large pores. This on-line arrangement facilitated (i) on-line desalting, (ii) automatic buffer exchange, (iii) simple concentration and (iv) fractionation of the protein based on their polarity. This advantageous method offers a convenient on-line proteomic approach as compared to the traditional immunodepletion process which results in dilution of the depleted protein fractions.

In a study by Jmeian and El Rassi [12], an integrated fluidic platform was introduced to deplete the HAP in serum and to fractionate/concentrate the medium and low abundance proteins. The HAP were depleted using monolithic antibody columns, whereas the fractionation/concentration was achieved by IMAC. In brief, the serum sample deprived of albumin was injected into the tandem series of antibody columns (connected in the order: protein G′, protein A, antihuman α1-antitrypsin, anti-human transferrin, anti-human haptoglobin and anti-human α_2-macroglobulin) in order to deplete the seven HAP. This was followed by online concentration by IMAC which

consisted of monolithic stationary phases that were bound to IDA chelated with Zn^{2+}, Ni^{2+} and Cu^{2+}. The proteins that were not retained on the IDA-metal columns were captured by the RPC column that followed the IDA-metal columns. The bound fractions from the metal-chelate columns and the RPC column were further analyzed using 2-DE. 1450 protein spots were detected from the gel analysis using SYPRO fluorescent stain. Analysis of the detected gel spots resulted in the identification of 295 proteins through LC-MS/MS and MALDI-TOF analysis.

As will be discussed in Chapter II, the above mentioned method was further expanded by using protein equalizer technology to reduce the dynamic protein concentration range. The equalized proteins were further fractionated on metal-chelate columns and the RPC column. The analysis of the bound fractions using LC-MS/MS resulted in identification of 82 non-redundant proteins. This approach was compared to depletion of HAP as a sample pre-treatment before fractionation using IMAC and it was observed that protein equalizer technology resulted in identification of more number of proteins, see Chapter II for more details.

Electrophoretic fractionation methods Electrophoretic fractionation of proteins has been commonly used in proteomics, and the long-standing fractionation method used is gel electrophoresis. Some of the drawbacks of 2-DE has allowed for various electrophoretic techniques to emerge in proteomics technology. Some of them include liquid-phase isoelectric focusing, free-flow electrophoresis (FFE), capillary electrophoresis (CE) and membrane electrophoresis [60]. CE combined with MS has

been proved to be efficient tool for biomarker discovery in urinary proteomics [61] and in tumor biomarker discovery [62].

In a study by Wang *et al.* [63], two different approaches were compared using human cancer cell lysates. The first approach was GeLC-MS/MS approach, in which proteins were first separated using SDS-PAGE and analyzed using LC-MS/MS. The second approach consisted of micro-scale solution (Micro-sol) isoelectric focusing of the proteins, which was followed by SDS-PAGE separation and finally the gel slices from SDS-PAGE were analyzed using LC-MS/MS. In Micro-sol liquid phase isoelectric focusing, protein samples are divided into multiple tandem electrode chambers using isoelectric membranes. As predicted, 90% of the proteins identified in the first approach were also found in the second approach. Also, the second approach resulted in 22% increase in the number of identified proteins indicating that the inclusion of the Micro-sol isoelectric focusing step before SDS-PAGE in the workflow allowed for increase in the proteome coverage.

Unlike other liquid-based fractionation methods, which involve a solid support, FFE is done in absence of any solid support such as gel. In FFE, an electric field is applied perpendicularly to the flow, which results in the separation of charged particles according to their electrophoretic mobility or isoelectric point. As an application of FFE, it was applied to identify the proteins present in urine sample [64]. The intact urine proteins was concentrated using ultrafiltration and isoelectric focusing *via* FFE to obtain approximately 50 fractions. These fractions were tryptically digested and analyzed using RPC-MS. This strategy of including the FFE step before MS analysis allowed facilitating the in-depth analysis of urine proteome.

Capturing/Targeting Specific Glycoproteins – Capturing/Targeting Sub-glycoproteomics by LAC and other Affinity Chromatography

The following sections discuss different strategies used in capturing sub-glycoproteome, which is a major sub-proteome of serum. The section focuses on glycoproteomic analysis in the aim of identifying candidate biomarkers for the time period of the last five years (2006 – present).

Single or serial lectin columns The identification and determination of structure and functions of glycoproteins are of great value in clinical studies. Recent reviews on the clinical applications of glycoproteomics can be found in refs. [65] and [66], which describe the different glycomic approaches used for the identification of cancer biomarkers.

Alterations in glycosylation patterns are very common in diseases. For example, alterations in sialylation, fucosylation, high mannose-type, sialyl Lewis x (sLe^x) structures and changes in the degree of glycan branching have been reported in cancer patients [67-69].

In a study to target the sialic acid-rich glycoproteins in pancreatic cancer serum, three different lectins were used namely wheat germ agglutinin (WGA), *sambuccus nigra* agglutinin (SNA) and *Maackia amurensis* lectin (MAL), which have affinity towards sialic acid containing glycoconjugates. The serum sample was first depleted of the 12 HAP and the sialylated glycoproteins were subsequently enriched using the three immobilized lectins. The enriched fractions were further fractionated by RPC using non-polar non-porous silica (NPS) stationary phases. To achieve enhanced speed, resolution

and reproducibility, the RPC column was heated at 60ºC. The fractions from the RPC column were further subjected to SDS-PAGE analysis. The three glycoproteins namely, IgG, α-1-antitrypsin and plasma protease C1 inhibitor were found to be altered and were subjected to MALDI-quadrupole ion trap-TOF-MS to identify the glycan structures. This strategy allowed the detection of the altered glycosylation in pancreatic serum with respect to disease-free serum [70].

In another application [71] of LAC, fucosylated serum glycoproteins were targeted using *Aleuria aurantia* lectin (AAL) to identify biomarkers of primary hepatocellular carcinoma (HCC). Prior to LAC with agarose-bound AAL, the serum was depleted of the top 12 most abundant proteins using commercially available polyclonal antibody microbeads. It was observed that the fucosylated biantennary glycans increased from 5.8 % in disease-free patients to 10% in HCC sera and to 8.5% in cirrhotic serum. Four fucosylated proteins namely hemopexin, α-2-HS-glycoprotein, anti-1-antichymotrypsin and transferrin were further validated using lectin fluorophore-linked immunosorbent assay.

Abbott *et al*. [72] demonstrated the enrichment of β(1,6)-branched *N*-linked glycan structures using L-phytohemagglutinin (L-PHA) lectin which has affinity towards the same. The β(1,6)-branched *N*-linked glycans have been reported to serve as marker in the detection of tumor progression [73]. Disease-free and breast cancer tissues were profiled by LAC using an L-PHA column. In brief, the delipidated tissue samples were treated with biotinylated L-PHA lectin to enrich the β(1,6)-branched *N*-linked glycoproteins. The L-PHA bound glycoproteins were then captured using streptavidin particles and the captured proteins were eluted with a mobile phase consisting of

urea/DTT/ammonium bicarbonate. The enriched glycoproteins were then subjected to nanospray ionization (NSI)-MS/MS, which allowed the identification of 34 proteins that were found to be elevated in breast cancer tissue when compared to the diseased-free tissue.

In another study by Mann *et al.* [7], a label-free quantitative analysis of fucosylated serum glycoproteins was carried out using the lectins AAL and *Lotus tetragonolobus* agglutinin (LTA). First, seven HAP proteins namely, albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin and fibrinogen were depleted from the serum samples. Thereafter, a serial LAC was performed using agarose-bound AAL and LTA lectins. The glycoproteins thus enriched by LAC were further fractionated using an RPC column. The RPC fractionated proteins were tryptically digested and analyzed by NSI-MS/MS. To ensure that the same amount of proteins was subjected to RPC fractionation, a bicinchoninic acid assay was performed. The quantification of the proteins was done by the summation of the peak areas of the identified peptides. This strategy was established to identify potential glycoprotein candidates in the study of esophageal adenocarcinoma and high-grade dysplasia.

In a recent study be Cho *et al.* [74], sLe^x glycan containing glycoproteins were targeted in human plasma to identify whether sLe^x containing proteins are shed into the blood stream from cell surfaces and to determine if these proteins are related to tumorigenesis. An agarose conjugated CHO-31 antibody was used to selectively capture the sLe^x containing glycoproteins. Further analysis was done using the following two approaches: (i) the captured glycoproteins were fractionated using RPC, and the fractions obtained were tryptically digested and subsequently identified using MALDI-MS/MS,

and (ii) the proteins were tryptically digested into peptides first and then fractionated on an RPC column. This was followed by MALDI-MS/MS analysis of the peptide RPC fractions. Of these two methods the former method identified four times more number of proteins than the latter. This method allowed the authors to identify some of the Le^x and sLe^x bearing glycoproteins in the plasma. Some of them were found to be elevated more than three folds in the breast cancer plasma, which could serve as potential biomarkers for breast cancer.

In an effort to identify putative altered glycoprotein biomarkers for lung adenocarcinoma, Hongsachart *et al.* [75] developed a method in which initially the healthy and the lung cancer serum were screened using seven fluorescein isothiocyanate (FITC) labeled lectins for specific glycoprotein profile of the sera. Based on the screening results, WGA, which showed highest specific binding with the glycoproteins, was selected for the enrichment of glycoproteins from the sera. Following this, a co-immunoprecipitation of haptoglobin using anti-haptoglobin was performed. The removal of haptoglobin from the WGA bound sample allowed the identification of increased number of differentially expressed proteins, which otherwise would have been masked by haptoglobin. Further analysis was done using DIGE, and it was noted that the unbound fraction from the WGA column mostly contained the high abundance proteins like albumin and IgG. Using this strategy, three up-regulated and two down-regulated glycoproteins in lung cancer serum relative to healthy serum were identified and they were further validated by Western blot analysis.

A tandem affinity depletion which combines affinity fractionation and immunoaffinity depletion was reported to identify low-abundance proteins in human

plasma [76]. In this approach, the glycoproteins present in plasma were first enriched by LAC using WGA. In the second step, immunoaffinity depletion was carried out using antibodies that were raised in llama (Lama Glama) against the proteins that were captured in the lectin enrichment step. By using this strategy, the authors were able to selectively enrich carcinoembryonic antigen that was spiked in disease-free serum by a factor of 600-800 fold.

In a study by Jung *et al.* [77], LAC was performed using lectins with broad and narrow specificity such as concanavalin A (Con A), *Helix pomatia* agglutinin, *Lycopersicon esculentum*, AAL and *Lens culinaris* agglutinin (LCA) to analyze the changes in protein concentration of breast cancer plasma in comparison to disease-free plasma. In this study, the quantification of the protein concentration was achieved with stable isotope coding. The glycoproteins that were enriched using LAC were tryptically digested, fractionated on an RPC column and analyzed using MALDI-MS/MS. It was observed that small groups of proteins increased in concentration by 3 or more fold in the breast cancer as compared to the disease-free plasma. It was also concluded that there is no relationship between the glycan diversity and the abundance of a particular protein glycoforms.

<u>Multi-lectin affinity chromatography (M-LAC) with high abundance protein depletion</u> In M-LAC mixture of lectins having complementary specificities for different glycosylation are immobilized in a given column. In M-LAC, after treating the lectin column with the sample, the column is eluted sequentially using specific displacer for each lectin, i.e., the haptenic sugar. Some of the lectin enrichment processes which are

used in M-LAC are discussed below. Due to the high complexity of serum that is usually amplified by the presence of HAP, some of the approaches integrated the process of depletion of HAP to identify the glycoproteins present in serum.

Hancock and his co-workers extensively used M-LAC for the enrichment of glycoproteins from serum/plasma [78-82]. As the presence of the HAP will interfere with the interaction of the low-abundant glycoproteins, the HAP were first depleted and then the depleted serum was fractionated using M-LAC. In a study by Plavina *et al*. [83], albumin and IgG were first depleted from the plasma sample followed by the enrichment of the glycoproteins using M-LAC and the identification of proteins using nano-LC-MS/MS. Also, a comparison of M-LAC with and without the depletion step was made, and it was shown that the total number of identified proteins (in bound and unbound fractions) increased from 120 to 191 by including the depletion step. To demonstrate the ability of this method, it was applied to biomarker discovery from psoriasis samples. It was observed that 11 proteins had different concentrations between the control and psoriasis plasma samples, and the protein galectin-3 binding protein was further validated using enzyme linked immunosorbent assay. The authors concluded that the combination of depletion of HAP with M-LAC allowed the in-depth analysis of proteins, which had concentrations of 10-100 ng/mL.

In another application of M-LAC by Hancock and co-workers [82], changes in breast cancer serum was identified using three lectins namely Con A, WGA and Jacalin (JAC). These lectins were mixed in a ratio of 1:1:1 and packed into a column. Some of the HAP like albumin, IgM, IgA, and IgG were depleted while other HAP such as α-1-antitrypsin, transferrin and haptoglobin were not depleted as they might be involved in

cancer related changes. The depleted serum was then subjected to M-LAC and the bound fractions were further analyzed using three different orthogonal analytical platforms to identify glycoproteins that had either concentration or glycan structure changes due to the breast cancer. In the first platform, SDS-PAGE analysis was performed and three different detection methods namely, Coomassie blue staining, fluorescent staining of the glycoproteins and lectin blotting with biotinylated SNA were carried out. In the second platform the proteins were fractionated based on their pI values using ProteomeChip (dPC) to identify the breast cancer proteins, which show a change in their pI values. In the third platform, a lectin-antibody sandwich microarray was performed using AAL to detect the neutral glycan structure changes in the breast cancer serum. By all the three platforms the authors identified complement C3 beta chain, α-1-antitrypsin, transferrin and α-1B-glycoprotein as potential glycoproteins for further studies in breast cancer human serum.

In a recent report by the same group [84], an automated high-performance liquid chromatography (HPLC) platform was introduced to remove high abundance proteins and to fractionate glycoproteins using immuno-affinity depletion and M-LAC in order to facilitate the identification of the breast cancer associated serum biomarkers. The depleted and the fractionated glycoproteins obtained from M-LAC were further subjected to isoelectric focusing separation using a digital dPC, which had the operating range of pH 4.20~6.20 and 6.00~8.00. The gel plugs from the dPC were combined to get 10 fractions and these were subjected to in-gel digestion and LC-MS analysis. It was concluded that the inclusion of the isoelectric focusing using dPC fractionation after M-LAC extended the dynamic range of serum proteome and also resulted in identification of

low abundance proteins with higher sequence coverage. Finally, the proteins thrombospondin-1 and 5, α-1B-glycoprotein, serum amyloid P-component and tenascin-X were selected as promising candidates to analyze breast cancer serum.

Lectin and other affinity microarrays The expeditious growth in identification of glycoprotein biomarkers for various diseases triggered the growth of high throughput and reproducible technique such as lectin microarray, which is a newly emerging technique to analyze the alterations in glycans. Some of the applications of lectin microarray include glycoform characterization, biomarker discovery, pathogen detection, etc. [85]. In microarrays, spots of proteins are arrayed on solid supports (e.g., microscopic slides) as capture molecules. The array is then treated with complex proteome to determine the presence and/or amount of proteins present in the sample. The interaction of the array and the sample is detected through various detection techniques such as fluorescent methods, evanescent fluorescent field, etc. Some of the reports in which lectin microarrays were used in identifying biomarkers for different diseases are discussed below.

The application of lectin array was demonstrated in ref. [86], whereby the 12 HAP were first depleted from the delipidated plasma and the glycoproteins present in the plasma were then enriched using agarose-bound Con A. The captured glycoproteins were further fractionated using NPS-RP-HPLC. The fractions obtained from the NPS-RP-HPLC were then screened for changes in glycosylation patterns using five lectins namely, AAL, SNA, MAL, peanut agglutinin (PNA) and Con A. The use of these five lectins covered > 95% of reported *N*-glycan types and elevated levels of fucosylation and

sialylation in colorectal cancer and adenoma plasma samples were observed when compared to disease-free samples. The fractions which showed altered glycosylation in the lectin microarray were subjected to SDS-PAGE analysis and lectin blotting. By this method, the authors were able to identify complement C3, histidine-rich glycoprotein and kininogen-1 as potential markers of colorectal cancer which showed elevated levels of fucosylation and sialylation.

An application of antibody microarray to analyze the difference in glycosylation was done in a study by Kuno *et al.* [87]. The protein prostate-specific antigen which is an *N*-linked glycoprotein and podoplanin an *O*-linked glycoprotein was selected for profiling the differential glycan expression. In brief, (i) the protein was enriched using immuno-precipitation using a specific antibody, (ii) the protein was quantified by immunoblotting method and (iii) an antibody-overlay lectin microarray was performed for profiling the glycosylation changes. This method makes use of the antibody to its maximal potential by using it for enrichment, quantification and in microarray. By this strategy, an ultrasensitive analysis on a nanogram-scale was performed, that would lead to rapid identification of glycoprotein biomarkers present in biological samples.

In another application by Liu *et al*. [88], lectin arrays were used for identification and confirmation of biomarker candidates for distinguishing early HCC from cirrhosis. Initially, sera sample depleted of the 12 HAP was analyzed using lectin array consisting of 16 lectins to identify the differences in glycosylation patterns. Based on the lectin-array results, the lectins AAL and LCA were chosen to selectively enrich the fucosylated glycoproteins. These glycoproteins were identified by LC-MS/MS and the potential biomarkers were further validated using an AAL-antibody array. By this combined

strategy complement C3, ceruloplasmin, histidine-rich glycoprotein and CD14 were reported as biomarker candidate for early detection of HCC.

A lectin microarray analysis of the glycans present on the cell surface was done to distinguish the stem-like glioblastoma neurosphere culture from traditional adherent glioblastoma cell line [89]. Glioblastoma is a common type of malignant primary brain tumor in humans. A lectin microarray which consisted of 16 lectins was used to screen the sample for difference in glycan patterns of the two cell lines. The results from the lectin microarray indicated that two galactose specific lectins *Trichosanthes kirilowii* agglutinin (TKA) and PNA showed stronger binding capacity to the cells as compared to other lectins. Thus, TKA and PNA were selected to capture the glycoproteins from the cell cultures using affinity chromatography. These glycoproteins were tryptically digested and analyzed using LC-MS/MS. The differentially expressed glycoproteins were analyzed using label-free spectral counting method. Six glycoproteins namely, receptor-type tyrosine-protein, phosphatase zeta, tenascin-C, chondroitin sulfate proteoglycan NG2, podocalyxin-like protein 1 and CD90, and CD44 that were differentially expressed in the disease cell line were further validated by Western blotting analysis. It was concluded that further analysis of these proteins might improve the earlier diagnosis of glioblastoma.

Rationale of the Study

As discussed in the previous sections, human serum is a complex mixture and identification of proteins present in it is a major task. Although over the recent years different fractionation methods have proved efficient in reducing the complexity of the

human serum proteome, which in turn facilitated the identification of the proteins present in serum by MS, there still exists a need for more efficient fractionation and sample preparation methods in order to allow a more in-depth proteomics profiling of the human serum. In-depth proteomics refers to reaching the analysis of the low abundance proteins that are the most likely to be reflective of a person's patho-physiological state. Thus, by identifying the low abundance proteins and being able to determine their alteration and expression in certain "diseased" serum, one could obtain information about the progression of the particular disease. The primary objective of the various interrelated projects described in this dissertation is to develop an integrated approach to reduce the complexity of the human serum proteomics and to capture the part that is the most affected.

In Chapter II, the dynamic concentration range of serum will be reduced using protein equalizer technology. The equalized proteins will then be further fractionated using a series of IMAC columns followed by RPC column prior to LC-MS/MS analysis. This strategy will allow for identification of more number of proteins as compared to the immuno subtraction of the high abundance proteins (i.e., the depletion approach).

In Chapter III, three tandem lectin monolithic columns of broad specificity will be investigated for their effectiveness in capturing and enriching a given sub-glycoproteome that is thought to be associated with the progression of cancer such as breast cancer. This method will allow the capturing of a wide range of serum glycoproteins and in turn the detection of certain glycoproteins that are differentially expressed in cancer serum as compared to disease-free serum.

In Chapter IV, three tandem lectin monolithic columns of narrow specificity will be examined in the capture of a more specific set of serum glycoproteins such as sialylated and fucosylated glycoproteins that are thought to be elevated in cancer serum. Since, this strategy targets only specific glycosylation, it would therefore give more insight into specific glycoproteins that are differentially expressed in cancer serum relative to disease-free serum.

In Chapter V, the broad and narrow specificity lectins studied in Chapters III and IV will be combined and used as six tandem monolithic columns to capture the serum sub-glycoproteome over a wide range of glycosylation. By assessing the usefulness of each set of lectins separately, the combination strategy favored the detection of a much wider range of glycoproteins than when using only broad or narrow specificity lectins. More specific rationale and significance statements are given in the introductory parts of each chapter of this dissertation.

Summary

This chapter has (i) outlined the scope of the dissertation, (ii) briefly discussed the principles of IMAC, LAC and protein equalizer technology, (iii) provided a brief introduction to glycoproteomics, (iv) briefly described the fundamental approaches in proteomics profiling, (v) overviewed the current depletion methods, protein equalizer technology and chromatographic/electrophoretic fractionation methods for in-depth proteomics, (vi) overviewed the different approaches used for sub-glycoproteomics and (vii) presented the rationale of the current investigation.

References

[1] J. Porath, J. Carlsson, I. Olsson and G. Belfrage, *Nature* **1975**, *258*, 598-599.

[2] E. Sulkowski, *Trends Biotechnol.* **1985**, *3*, 1-7.

[3] E. Sulkowski, *Bioessays* **1989**, *10*, 170-175.

[4] H. Block, B. Maertens, A. Spriestersbach, N. Brinker, J. Kubicek, R. Fabis, J. Labahn and F. Schafer, *Methods Enzymol.* **2009**, *463*, 439-473.

[5] Q. Sun, X. Kang, Y. Zhang, H. Zhou, Z. Dai, W. Lu, X. Zhou, X. Liu, P. Yang and Y. Liu, *Arch. Biochem. Biophys.* **2009**, *484*, 24-29.

[6] S. Pan, Y. Wang, J. F. Quinn, E. R. Peskind, D. Waichunas, J. T. Wimberger, J. Jin, J. G. Li, D. Zhu, C. Pan and J. Zhang, *J. Proteome Res.* **2006**, *5*, 2769-2779.

[7] B. Mann, M. Madera, I. Klouckova, Y. Mechref, L. E. Dobrolecki, R. J. Hickey, Z. T. Hammoud and M. V. Novotny, *Electrophoresis* **2010**, *31*, 1833-1841.

[8] P. R. Satish and A. Surolia, *J. Biochem. Biophys. Meth.* **2001**, *49*, 625-640.

[9] Y. S. Kim, H. S. Yoo and J. H. Ko, *Protein Pept. Lett.* **2009**, *16*, 499-507.

[10] J. L. Bundy and C. Fenselau, *Anal. Chem.* **2001**, *73*, 751-757.

[11] M. C. Wiener and A. N. van Hoek, *Anal. Biochem.* **1996**, *241*, 267-268.

[12] Y. Jmeian and Z. El Rassi, *J. Proteome Res.* **2009**, *8*, 4592-4603.

[13] Y. Jmeian and Z. El Rassi, *Electrophoresis* **2008**, *29*, 2801-2811.

[14] D. Josic and J. G. Clifton, *J. Chromatogr. A* **2007**, *1144*, 2-13.

[15] R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley and J. H. Cuervo, *Nature* **1991**, *354*, 84-86.

[16] L. Guerrier, P. G. Righetti and E. Boschetti, *Nat. Protoc.* **2008**, *3*, 883-890.

[17] E. Boschetti, L. Lomas, A. Citterio and P. G. Righetti, *J. Chromatogr. A* **2007**, *1153*, 277-290.

[18] C. Simo, A. Bachi, A. Cattaneo, L. Guerrier, F. Fortis, E. Boschetti, A. Podtelejnikov and P. G. Righetti, *Anal. Chem.* **2008**, *80*, 3547-3556.

[19] J. Granger, J. Siddiqui, S. Copeland and D. Remick, *Proteomics* **2005**, *5*, 4713-4718.

[20] M. Zhou, D. A. Lucas, K. C. Chan, H. J. Issaq, E. F. Petricoin, 3rd, L. A. Liotta, T. D. Veenstra and T. P. Conrads, *Electrophoresis* **2004**, *25*, 1289-1298.

[21] A. Cereda, A. V. Kravchuk, A. D'Amato, A. Bachi and P. G. Righetti, *J. Proteomics* **2010**, *73*, 1732-1739.

[22] W. Patrick, S. Hans and P. Angelika, *J. Agric. Food Chem.* **2009**, *57*, 8399-8405.

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

[24] P. M. Rudd and R. A. Dwek, *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, 1-100.

[25] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683-720.

[26] C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto, Jr., *Angew Chem. Int. Ed. Engl.* **2005**, *44*, 7342-7372.

[27] S. A. Brooks, M. V. Dwek and U. Schumacher Functional and Molecular Glycobiology, *Vol.* BIOS Scientific Publishers Lts, **2002**, pp. 287-312.

[28] B. Blomme, C. Van Steenkiste, N. Callewaert and H. Van Vlierberghe, *J. Hepatol.* **2009**, *50*, 592-603.

[29] N. L. Anderson and N. G. Anderson, *Mol. Cell. Proteomics* **2002**, *1*, 845-867.

[30] R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta* **1999**, *1473*, 4-8.

[31] P. Xu, D. M. Duong and J. Peng, *J. Proteome Res.* **2009**, *8*, 3944-3950.

[32] R. Xiang, Y. Shi, D. A. Dillon, B. Negin, C. Horvath and J. A. Wilkins, *J. Proteome Res.* **2004**, *3*, 1278-1283.

[33] G. Mitulovic, C. Stingl, M. Smoluch, R. Swart, J. P. Chervet, I. Steinmacher, C. Gerner and K. Mechtler, *Proteomics* **2004**, *4*, 2545-2557.

[34] D. B. Wall, M. T. Kachman, S. S. Gong, S. J. Parus, M. W. Long and D. M. Lubman, *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1649-1661.

[35] M. Gilar, P. Olivova, A. E. Daly and J. C. Gebler, *Anal. Chem.* **2005**, *77*, 6426-6434.

[36] M. Gilar, P. Olivova, A. B. Chakraborty, A. Jaworski, S. J. Geromanos and J. C. Gebler, *Electrophoresis* **2009**, *30*, 1157-1167.

[37] H. L. Phillips, J. C. Williamson, K. A. van Elburg, A. P. Snijders, P. C. Wright and M. J. Dickman, *Proteomics* **2010**, *10*, 2950-2960.

[38] Y. Wang, F. Yang, M. A. Gritsenko, T. Clauss, T. Liu, Y. Shen, M. E. Monroe, D. Lopez-Ferrer, T. Reno, R. J. Moore, R. L. Klemke, D. G. Camp, 2nd and R. D. Smith, *Proteomics* **2011**, *11*, 2019-2026.

[39] S. Krishnan, M. Gaspari, A. Della Corte, P. Bianchi, M. Crescente, C. Cerletti, D. Torella, C. Indolfi, G. de Gaetano, M. B. Donati, D. Rotilio and G. Cuda, *Electrophoresis* **2011**, *32*, 686-695.

[40] Y. Jmeian and Z. El Rassi, *Electrophoresis* **2009**, *30*, 249-261.

[41] F. E. Ahmed, *J. Sep. Sci.* **2009**, *32*, 771-798.

[42] P. Matt, Z. Fu, Q. Fu and J. E. Van Eyk, *Physiol. Genomics* **2008**, *33*, 12-17.

[43] B. de Roos, *Exp. Rev. Proteomics* **2008**, *5*, 819-826.

[44] Y. Kawashima, T. Fukutomi, T. Tomonaga, H. Takahashi, F. Nomura, T. Maeda and Y. Kodera, *J. Proteome Res.* **2010**, *9*, 1694-1705.

[45] S. E. Warder, L. A. Tucker, T. J. Strelitzer, E. M. McKeegan, J. L. Meuth, P. M. Jung, A. Saraf, B. Singh, J. Lai-Zhang, G. Gagne and J. C. Rogers, *Anal. Biochem.* **2009**, *387*, 184-193.

[46] C. Tu, P. A. Rudnick, M. Y. Martinez, K. L. Cheek, S. E. Stein, R. J. Slebos and D. C. Liebler, *J. Proteome Res.* **2010**, *9*, 4982-4991.

[47] A. R. Vaezzadeh, A. C. Briscoe, H. Steen and R. S. Lee, *J. Proteome Res.* **2010**, *9*, 6082-6089.

[48] S. Roche, L. Tiers, M. Provansal, M. Seveno, M.-T. Piva, P. Jouin and S. Lehmann, *J. Proteomics* **2009**, *72*, 945-951.

[49] B. Lin, J. T. White, J. Wu, S. Lele, L. J. Old, L. Hood and K. Odunsi, *Proteomics Clin. Appl.* **2009**, *3*, 853-861.

[50] W. J. Qian, D. T. Kaleta, B. O. Petritis, H. Jiang, T. Liu, X. Zhang, H. M. Mottaz, S. M. Varnum, D. G. Camp, 2nd, L. Huang, X. Fang, W. W. Zhang and R. D. Smith, *Mol. Cell. Proteomics* **2008**, *7*, 1963-1973.

[51] U. Restuccia, E. Boschetti, E. Fasoli, F. Fortis, L. Guerrier, A. Bachi, A. V. Kravchuk and P. G. Righetti, *J. Proteomics* **2009**, *72*, 1061-1070.

[52] C. Sihlbom, I. Kanmert, H. Bahr and P. Davidsson, *J. Proteome Res.* **2008**, *7*, 4191-4198.

[53] T. Leger, D. Lavigne, J. P. Le Caer, L. Guerrier, E. Boschetti, J. Fareh, L. Feldman, O. Laprevote and O. Meilhac, *Clin. Chim. Acta* **2011**, *412*, 740-747.

[54] C. Tu, J. Li, R. Young, B. J. Page, F. Engler, M. S. Halfon, J. M. Canty and J. Qu, *Anal. Chem.* **2011**, *83*, 4802-4813.

[55] M. De Bock, D. de Seny, M.-A. Meuwis, A.-C. Servais, T. Q. Minh, J. Closset, J.-P. Chapelle, E. Louis, M. Malaise, M.-P. Merville and M. Fillet, *Talanta* **2010**, *82*, 245-254.

[56] J. E. Bandow, *Proteomics* **2010**, *10*, 1416-1425.

[57] A. Mahn, A. Reyes, M. Zamorano, W. Cifuentes and M. Ismail, *J. Chromatogr. B,* **2010**, *878*, 1038-1044.

[58] J. Wu, Y. An, H. Pu, Y. Shan, X. Ren, M. An, Q. Wang, S. Wei and J. Ji, *Anal. Biochem.* **2010**, *398*, 34-44.

[59] N. A. Cellar, A. S. Karnoup, D. R. Albers, M. L. Langhorst and S. A. Young, *J. Chromatogr. B* **2009**, *877*, 79-85.

[60] L. Ly and V. C. Wasinger, *Proteomics* **2011**, *11*, 513-534.

[61] H. Mischak, J. J. Coon, J. Novak, E. M. Weissinger, J. P. Schanstra and A. F. Dominiczak, *Mass Spectrom. Rev.* **2009**, *28*, 703-724.

[62] A. V. Simionato, E. Carrilho and M. F. Maggi Tavares, *Electrophoresis* **2010**, *31*, 1214-1226.

[63] H. Wang, T. Chang-Wong, H. Y. Tang and D. W. Speicher, *J. Proteome Res.* **2010**, *9*, 1032-1040.

[64] A. L. Foucher, D. R. Craft and C. A. Gelfand, *Methods Mol. Biol.* **2010**, *641*, 27-45.

[65] Y. Tian and H. Zhang, *Proteomics Clin. Appl.* **2010**, *4*, 124-132.

[66] D. Taylor Allen, S. Hancock William, M. Hincapie, N. Taniguchi and M. Hanash Samir, *Genome Med.* **2009**, *1*, 57.

[67] R. Saldova, L. Royle, C. M. Radcliffe, U. M. Abd Hamid, R. Evans, J. N. Arnold, R. E. Banks, R. Hutson, D. J. Harvey, R. Antrobus, S. M. Petrescu, R. A. Dwek and P. M. Rudd, *Glycobiology* **2007**, *17*, 1344-1356.

[68] J. L. Dage, B. L. Ackermann and H. B. Halsall, *Glycobiology* **1998**, *8*, 755-760.

[69] E. C. Brinkman-van der Linden, P. F. de Haan, E. C. Havenaar and W. van Dijk, *Glycoconj. J.* **1998**, *15*, 177-182.

[70] J. Zhao, D. M. Simeone, D. Heidt, M. A. Anderson and D. M. Lubman, *J. Proteome Res.* **2006**, *5*, 1792-1802.

[71] M. A. Comunale, M. Wang, J. Hafner, J. Krakover, L. Rodemich, B. Kopenhaver, R. E. Long, O. Junaidi, A. M. Bisceglie, T. M. Block and A. S. Mehta, *J. Proteome Res.* **2009**, *8*, 595-602.

[72] K. L. Abbott, K. Aoki, J. M. Lim, M. Porterfield, R. Johnson, R. M. O'Regan, L. Wells, M. Tiemeyer and M. Pierce, *J. Proteome Res.* **2008**, *7*, 1470-1480.

[73] B. Fernandes, U. Sagman, M. Auger, M. Demetrio and J. W. Dennis, *Cancer Res.* **1991**, *51*, 718-723.

[74] W. Cho, K. Jung and F. E. Regnier, *J. Proteome Res.* **2010**, *9*, 5960-5968.

[75] P. Hongsachart, R. Huang-Liu, S. Sinchaikul, F.-M. Pan, S. Phutrakul, Y.-M. Chuang, C.-J. Yu and S.-T. Chen, *Electrophoresis* **2009**, *30*, 1206-1220.

[76] N. Mortezai, S. Harder, C. Schnabel, E. Moors, M. Gauly, H. Schluter, C. Wagener and F. Buck, *J. Proteome Res.* **2010**, *9*, 6126-6134.

[77] K. Jung, W. Cho and F. E. Regnier, *J. Proteome Res.* **2009**, *8*, 643-650.

[78] Z. Yang and W. S. Hancock, *J. Chromatogr. A* **2004**, *1053*, 79-88.

[79] Z. Yang and W. S. Hancock, *J. Chromatogr. A* **2005**, *1070*, 57-64.

[80] M. Kullolli, W. S. Hancock and M. Hincapie, *J. Sep. Sci.* **2008**, *31*, 2733-2739.

[81] Z. Yang, L. E. Harris, D. E. Palmer-Toy and W. S. Hancock, *Clin. Chem.* **2006**, *52*, 1897-1905.

[82] Z. Zeng, M. Hincapie, B. B. Haab, S. Hanash, S. J. Pitteri, S. Kluck, J. M. Hogan, J. Kennedy and W. S. Hancock, *J. Chromatogr. A* **2010**, *1217*, 3307-3315.

[83] T. Plavina, E. Wakshull, W. S. Hancock and M. Hincapie, *J. Proteome Res.* **2007**, *6*, 662-671.

[84] Z. Zeng, M. Hincapie, S. J. Pitteri, S. Hanash, J. Schalkwijk, J. M. Hogan, H. Wang and W. S. Hancock, *Anal. Chem.* **2011**, *83*, 4845-4854.

[85] G. Gupta, A. Surolia and S.-G. Sampathkumar, *OMICS: J. Integ. Biol.* **2010**, *14*, 419-436.

[86] Y. Qiu, T. H. Patwa, L. Xu, K. Shedden, D. E. Misek, M. Tuck, G. Jin, M. T. Ruffin, D. K. Turgeon, S. Synal, R. Bresalier, N. Marcon, D. E. Brenner and D. M. Lubman, *J. Proteome Res.* **2008**, *7*, 1693-1703.

[87] A. Kuno, Y. Kato, A. Matsuda, M. K. Kaneko, H. Ito, K. Amano, Y. Chiba, H. Narimatsu and J. Hirabayashi, *Mol. Cell. Proteomics* **2009**, *8*, 99-108.

[88] Y.-S. Liu, J.-T. He, C. Li, R. Benitez, S. Fu, J. Marrero and D. M. Lubman, *J. Proteome Res.* **2010**, *9*, 798-805.

[89] J. He, Y. Liu, X. Xie, T. Zhu, M. Soules, F. DiMeco, A. L. Vescovi, X. Fan and D. M. Lubman, *J. Proteome Res.* **2010**, *9*, 2565-2572.

CHAPTER II

REDUCTION OF PROTEIN CONCENTRATION RANGE DIFFERENCE FOLLOWED BY MULTICOLUMN FRACTIONATION PRIOR TO 2-DE AND LC-MS/MS PROFILING OF SERUM PROTEINS

Introduction

The wide dynamic concentration range of serum proteins that extends over 10 – 12 orders of magnitude, and the many thousands of proteins that might be present in serum make the in-depth proteomic analysis of such a sample a major challenge for current analytical and separation technologies. Many protein biomarkers in the serum are present at very low concentration, and most often the high-abundance proteins mask these proteins. Thus far, three major approaches have been introduced to reduce the complexity of human serum proteomics and facilitate serum proteomic profiling, including (i) depleting high abundance proteins by immuno-subtraction [1-3], (ii)

* *The content of this chapter has been published in Electrophoresis,* ***2011****, 32, 674-685.*

minimizing concentration differences in abundance by the concept of "protein equalizer" or the ProteoMiner™, which uses peptide library beads derived from combinatorial peptide ligand libraries (CPLL) [4, 5] and (iii) electrophoretic and chromatographic pre-fractionation [1, 6].

In recent studies, Jmeian and El Rassi introduced tandem affinity columns based on protein A, protein G´ and antibodies for the depletion of the eight high abundance proteins (e.g., albumin, IgG, IgA, IgM, transferrin, α_1-antitrypsin, haptoglobin and α_2-macroglobulin) [7] and multicolumn based platforms for the depletion of high abundance proteins and subsequent concentration/fractionation of low abundance proteins and their applications to profiling protein expression in disease-free and osteoarthritis sera [8, 9]. The multicolumn platforms demonstrated the simultaneous operation of the tandem affinity depletion columns with tandem fractionation columns consisting of immobilized metal affinity chromatography (IMAC) columns in series with a reversed-phase chromatography (RPC) column. These platforms were very effective in preparing serum protein fractions for 2-dimensional gel electrophoresis (2-DE) and LC-MS/MS profiling.

Of interest to our investigation is the novel approach called "protein equalizer technology" or "ProteoMiner™" for reducing protein concentration differences that are based on the selective adsorption of proteins to peptide library beads. The ProteoMiner™ technology has been recently introduced [10] and its principle of preparation, operation and applications were reviewed, among other things, in recent articles [4-6, 11]. Briefly, the protein equalizer technology consisting of peptide affinity beads can in principle achieve simultaneously the dilution of high abundance proteins and concentration of low abundance proteins. If the hexapeptide library contains equal

numbers of affinity ligands to all proteins and the sample volume is large enough to saturate all binding sites, the abundance of all proteins in the final eluted fraction will be the same [11].

In this study, the ProteoMiner™ technology was combined with IMAC columns and RPC column for achieving protein concentration range reduction followed by concentration/fractionation of the equalized proteins. While the ProteoMiner™ equalizes, IMAC and RPC fractionation allows the simplification of the sample complexity.

Experimental

Instrumentation

A Milton Roy, LDC division, multiple solvent delivery system Model CM4000 (Riviera Beach, FL, USA) was used with a Rheodyne injector Model 7010 (Cotati, CA, USA) with a 1 mL loop along with a metering pump Model III CM from Milton Roy, LDC division, and a Model 200 UV-Vis variable wavelength detector from Linear Instruments (Reno, NV, USA).

The first dimension of the 2-DE experiments was performed on a Multiphor II IEF system from GE Healthcare (Uppsala, Sweden) while the second dimension was performed on a mini-Protean module for 8.6 x 6.8 cm gels from Bio-Rad Laboratories (Hercules, CA, USA). Fluorescent gel images were taken with Typhoon Trio Plus from GE Healthcare. All mass spectra were obtained using a hybrid LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA).

Reagents and Materials

Pooled human serum was purchased from Innovative Research (Southfield, MI, USA). Iodoacetamide, urea and dithiothreitol (DTT), were purchased from Sigma Chemical Co. (St. Louis, MO, USA). Glycidyl methacrylate (GMA), ethylene glycol dimethacrylate (EDMA), 2,2´-azobis(isobutyronitrile) (AIBN) and 1-dodecanol were purchased from Aldrich Chemical Co. (Milwaukee, WI, USA). Cyclohexanol, HPLC grade acetonitrile, zinc sulfate, cupric chloride, and nickelous nitrate were obtained from Fisher Scientific (Fair Lawn, NJ, USA). Glycine, tris(hydroxymethyl)aminomethane (Tris), sodium dodecyl sulfate (SDS), acrylamide, bromophenol blue, Bio-Safe™ Coomassie, SYPRO® ruby protein gel stain, ReadStrip™ IPG strip 7 cm pH 4-7, Bio-Lyte® 3/10, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), Tris-glycine gels 10-20% Tris-HCl (8.6 cm W x 6.8 cm L) and the ProteoMiner™ kit were from Bio-Rad Laboratories (Hercules, CA, USA). Iminodiacetic acid (IDA) was a gift from W. R. Grace (Nashua, NH, USA). Poly(styrene/divinylbenzene) reversed-phase (RP) media with a particle size of 20 μm and a pore size of 300 Å (PLRP-S, 20 μm, 300 Å) was purchased from Polymer Laboratories (Amherst, MA, USA).

Affinity Columns and Reversed-Phase Column Preparation

A well mixed and degassed polymerization mixture of 18% (w/w) GMA, 12% (w/w) EDMA, 59.5% (w/w) cyclohexanol and 10.5% (w/w) dodecanol containing 1.0% (w/w) AIBN with respect to the monomers was introduced into a 25.0 cm x 4.6 mm ID. stainless steel column that functions as a mold for the monolith. The column was heated at 50 ˚C for 24 h. The resulting monolithic column was washed extensively with

acetonitrile (ACN) followed by water and then filled with 0.9 M IDA in 2.0 M potassium carbonate and heated at 80 ˚C for 24 h to form an IDA-modified surface [12]. The IDA-column was then rinsed with water. The modified monolithic support was transferred from the 25.00 cm column to a shorter column (5.00 cm) by connecting the two columns with a ¼"-union and running water through the columns at flow rate of 3.0 mL/min until the modified monolithic support is transferred. Three IDA columns of 5.00 cm x 4.6 mm each were filled up with the IDA-monolith by this way. The metal ions were immobilized on the IDA-monolithic surface by pumping ten column volumes of 5.0 mg/mL metal solution through the column. Then the columns were washed with ten column volumes of water and 5 column volumes of loading mobile phase. The RPC column was prepared by dry packing a 3.00 cm x 4.6 mm ID. stainless steel column (having an end column fitting with a 2 μm frit attached at the outlet end) with PLRP-S particles having 20 μm mean particle diameter and 300 Å mean pore diameter. Once the dry packing was finished, a second end column fitting with a 2-μm frit was attached to the inlet end. Thereafter, the column was flushed with isopropanol at a backpressure between 1500 and 2000 psi to ensure good packing. The inlet column end fitting was removed to repack the column again to minimize any void volume. In the last step, the column was rinsed with acetonitrile and stored at room temperature.

ProteoMiner™ Treated Serum

Serum protein equalization was performed using ProteoMiner™ enrichment kit according to the manufacturer procedure. In summary, the storage solution was first washed out from the spin column containing 100 μL of peptide beads with deionized

water. Thereafter, the column was washed with the 10 mM NaH_2PO_4, 150 mM NaCl, pH 7.4 solution provided with the kit. When the spin column was ready for sample binding, 1 mL of centrifuged serum sample was added to the column and equilibrated at room temperature for 2 h. The unbound proteins were removed with the wash buffer and the captured proteins were eluted by 3 x 100 μL of 8 M urea containing 2% CHAPS dissolved in 5% acetic acid. Although this loading and elution protocol of the peptide beads has been practiced by others [13-15], different and harsher elution conditions have been reported by Sennels *et al.* [16] and Fasoli *et al.* [17].

The resulting solution containing the equalized proteins were dialyzed against (i) 0.1 M urea (ii) 0.01 M urea (iii) 0.005 M NaH_2PO_4 containing 0.005 M NaCl pH 7.0 and (iv) the binding mobile phase (0.05 M Na_2HPO_4, 0.05 M NaCl, pH 7.0) used for the tandem columns. The equalized dialyzed sample (~1.5 mL) was injected in 0.5 mL aliquot onto the IMAC/RPC tandem columns (see next section) to fractionate the equalized proteins.

Tandem Column Platform - Chromatographic Conditions

The IDA-metal columns were arranged in the following order: IDA-Zn^{2+} → IDA-Ni^{2+} → IDA-Cu^{2+} followed by the RPC column. The RPC column that was originally stored in acetonitrile was washed with 10 bed volumes of water before use. The 4 columns connected in tandem were first equilibrated with 10 column volumes of the binding mobile phase consisting of 0.05 M Na_2HPO_4, 0.05 M NaCl, pH 7.0. Then, 0.5 mL of the equalized serum was injected onto the tandem columns. This was followed by washing with the binding mobile phase for 15 min at a flow rate of 0.5 mL/min to remove

any unbound proteins. Thereafter, the tandem columns were disconnected from each other, and the proteins captured by each IDA-metal column were eluted using of 0.05 M Na_2HPO_4, 0.05 M NaCl, 100 mM imidazole, pH 7.0. The RPC column was washed with water and the bound proteins were eluted using 80:20 (v/v) ACN:H_2O containing 0.1% trifluoroacetic acid (TFA). After eluting the proteins, the RPC column was again washed with water to prepare it for the next injection.

Two-Dimensional Gel Electrophoresis (2-DE) and Spot Counting

The fractions from the three IMAC and the RPC columns were dialyzed against water at 4˚C for 24 h in Spectra/Por dialysis bags from Spectrum Laboratories, Inc (Houston, TX, USA) according to the manufacturer procedure. The samples were then evaporated to dryness with a speed vac and stored at -20˚C until use. Part of each fraction was submitted for the LC-MS/MS analysis and the remaining fractions were re-dissolved in 125 μL of the rehydration solution made of 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 0.8% (w/v) Bio-Lyte^{R} 3/10. The 7 cm isoelectric focusing strips were allowed to stay in the rehydration solution for 12 h. Then the strips were focused with a voltage ramp of 200 V for 2 h, 500 V for 1 h and 3500 V for 4 h at 20˚C. After focusing, the strips were stored at -70˚C until use. The strips were first equilibrated for 15 min with a buffer solution made of 50 mM Tris-HCl, pH 8.8, containing 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, 1% (w/v) DTT and traces of bromophenol blue, then equilibrated with a buffer solution made of 50 mM Tris-HCl, pH 8.8, containing 6 M urea, 30% (v/v) glycerol, 2% (w/v) iodoacetamide and traces of bromophenol blue for another 15 min. The strips were then sealed to the polyacrylamide gel (10-20%) with 1%

(w/v) agarose solution and the second dimension was performed according to the precast gel manufacturer procedure. Thereafter, the gels were stained with SYPRO^R Ruby Protein stain according to the manufacturer recommendation. The gels were scanned with the Typhoon Trio Plus instrument at 100 µm resolution. The 532 nm excitation laser and the 610 nm band pass emission filter were used to get the image of the gels. The software ImageMaster 2D Platinum v6.0 from GE-Healthcare (Uppsala, Sweden) was used for detecting the spots from the 2-DE electropherograms. After washing the SYPRO-stained gels with water, Coomassie blue staining was performed for 1 h according to the manufacturer protocol and then finally the gels were washed with water to prevent over-staining.

LC-MS/MS Methodology

Samples were analyzed on a hybrid LTQ-Orbitrap mass spectrometer from ThermoFisher Scientific coupled to a New Objectives PV-550 nanoelectrospray ion source and an Eksigent NanoLC-2D chromatography system. Peptides were analyzed by trapping on a 2.5 cm ProteoPrep II pre-column (New Objective) and analytical separation on a 75 µm ID fused silica column packed in house with 10-cm of Magic C18 AQ, terminated with an integral fused silica emitter pulled in house. Peptides were eluted using a 5-40% ACN/0.1% formic acid gradient performed over 40 min at a flow rate of 300 nL/min. During each one-second full-range FT-MS scan (nominal resolution of 60,000 FWHM, 300 to 2000 m/z), the three most intense ions were analyzed *via* MS/MS in the linear ion trap. MS/MS settings used a trigger threshold of 1000 counts, monoisotopic precursor selection (MIPS), and rejection of parent ions that had

unassigned charge states, were previously identified as contaminants on blank gradient runs, or were previously selected for MS/MS (data dependent acquisition using a dynamic exclusion for 150% of the observed chromatographic peak width). Column performance was monitored using trypsin autolysis fragments (m/z 421.76), and *via* blank injections between samples to assay for contamination.

LC-MS/MS Data Analysis

Centroided ion masses were extracted using the extract_msn.exe utility from Bioworks 3.3.1 and were used for database searching with Mascot v2.2.04 (Matrix Science) and X! Tandem v2007.01.01.1 (www.thegpm.org). Both Mascot and X! Tandem were set up to search the IPI_Human_022209 database (v3.55, 75554 entries) assuming the digestion enzyme trypsin. Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 15 ppm. *S*-Carbamoylmethylcysteine cyclization (*N*-terminus) of the *N*-terminus, oxidation of methionine, *N*-formylation of the *N*-terminus, acetylation of the *N*-terminus and iodoacetamide derivative of cysteine were specified in Mascot and X! Tandem as variable modifications.

Scaffold (version Scaffold-2-05-01, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted only if they could be established at greater than 80% probability as specified by the Peptide Prophet algorithm. Protein identifications were accepted if they could be established at greater than 99% probability and contained at least 2 unique identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm.

Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Results and Discussion

Overall Strategy – Combinatorial Peptide Ligand Libraries (ProteoMiner™) Followed by Column Fractionation

It is well established that CPLL is capable of reducing the concentration dynamic range of the serum proteome, thus conveniently allowing the detection of the low abundance proteins [16]. Also, the use of tandem IMAC and RPC columns has been shown in a recent article from our laboratory to serve as an effective approach to fractionate and concentrate serum proteins [9]. In this tandem arrangement, the IDA-metal chelate and RPC columns were connected in the order of decreasing selectivity (i.e., IDA-Zn^{+2} →IDA-Ni^{+2} → IDA-Cu^{+2} → RPC columns). By allowing the serum treated with peptide beads to contact first the IDA-Zn^{2+} column (the more selective column in the chain), the binding sites of IDA-Ni^{2+} are saved for the proteins to which this column has the strongest affinity. Also, IDA-Cu^{2+}, which has an affinity for proteins with even one single surface exposed histidine (His) residue would benefit from placing the other two metal chelate columns ahead so that its binding capacity could be used more efficiently for capturing other proteins. Of course, any other proteins that are not retained by the tandem IMAC columns will be trapped in the RPC column. The combination of reducing protein concentration differences provided by the peptide beads with the fractionation and concentration provided by the tandem IMAC and RPC columns should in principle facilitate the mining of serum proteome. In order to assess

TABLE 1

LC-MS/MS RESULTS OF THE IDENTIFIED PROTEINS FROM THE EQUALIZED HUMAN SERUM WITHOUT SUBSEQUENT FRACTIONATION

Identified Proteins	Accession #	Molecular weight (KDa)	Spectra count	Unique peptide
252 kDa protein	IPI00022937 (+1)	252	9	2
55 kDa protein	IPI00029863 (+2)	55	6	3
Antithrombin III variant	IPI00032179	53	112	16
Apolipoprotein A-I	IPI00021841	31	178	26
Apolipoprotein A-II	IPI00021854	11	91	9
Apolipoprotein A-IV	IPI00304273	45	101	24
Apolipoprotein B-100	IPI00022229	516	80	21
Apolipoprotein C-I	IPI00021855	9	6	2
Apolipoprotein C-II	IPI00021856	11	15	3
Apolipoprotein C-III	IPI00021857 (+1)	11	22	4
Apolipoprotein D	IPI00006662	21	39	8
Apolipoprotein E	IPI00021842	36	110	20
C4b-binding protein alpha chain	IPI00021727	67	103	17
Carboxypeptidase N subunit 2	IPI00479116	61	8	2

CD5 antigen-like	IPI00025204	38	17	5
cDNA FLJ51925, highly similar to Vitamin K-dependent protein C	IPI00908685 (+1)	58	10	2
Ceruloplasmin	IPI00017601	122	46	13
Clusterin	IPI00291262 (+2)	52	99	21
Complement C1q subcomponent subunit A	IPI00022392	26	16	3
Complement C1q subcomponent subunit C	IPI00022394	26	23	7
Complement C1s subcomponent	IPI00017696 (+1)	77	14	4
Complement C3 (Fragment)	IPI00783987	187	643	114
Complement C5	IPI00032291	188	22	6
complement component 1, q subcomponent, B chain precursor	IPI00477992 (+1)	27	21	3
Complement component 4B preproprotein	IPI00418163 (+2)	193	763	84
Complement component C9	IPI00022395	63	14	6
Complement factor H-related 5	IPI00006543 (+1)	67	7	3
Complement factor H-related protein 1	IPI00011264 (+1)	38	40	5
Fibronectin 1 isoform 4 preproprotein	IPI00414283 (+1)	257	244	51
Fibulin 1	IPI00889740	78	63	3
FLJ00385 protein (Fragment)	IPI00168728 (+6)	56	38	3

Galectin-3-binding protein	IPI00023673	65	51	12
Glutathione peroxidase 3	IPI00026199	26	12	4
IGHA1 protein	IPI00166866 (+3)	53	40	9
IGHG1 protein	IPI00448925 (+12)	60	54	13
IGHM protein	IPI00477090 (+3)	67	84	17
IGKV1-5 protein	IPI00419424 (+19)	26	37	5
IGLV3-25 protein	IPI00550162 (+1)	25	38	5
immunoglobulin J chain	IPI00178926	18	10	2
Inter-alpha-trypsin inhibitor heavy chain H1	IPI00292530	101	8	3
Inter-alpha-trypsin inhibitor heavy chain H2	IPI00305461 (+1)	106	60	14
Isoform 1 of Alpha-1-antitrypsin	IPI00553177	47	80	16
Isoform 1 of C4b-binding protein beta chain	IPI00025862 (+1)	28	9	2
Isoform 1 of Complement factor H	IPI00029739	139	170	37
Isoform 1 of C-reactive protein	IPI00022389	25	8	2
Isoform 1 of Ficolin-2	IPI00017530 (+2)	34	3	2
Isoform 1 of Ficolin-3	IPI00293925 (+1)	33	42	7
Isoform 1 of Haptoglobin-related protein	IPI00477597	39	12	3
Isoform 1 of Pregnancy zone protein	IPI00025426	164	28	9

Isoform 1 of Serum albumin	IPI00745872	69	121	27
Isoform 2 of Apolipoprotein L1	IPI00186903 (+2)	46	10	3
Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4	IPI00218192 (+3)	101	17	5
Isoform D of Fibulin-1	IPI00296534	77	67	13
Mannan-binding lectin serine protease 1 isoform 2 precursor	IPI00290283 (+1)	82	3	2
Phosphatidylcholine-sterol acyltransferase	IPI00022331	50	13	4
Plasminogen	IPI00019580	91	11	4
Properdin	IPI00021364	51	10	3
Protein Z-dependent protease inhibitor	IPI00007199	55	5	2
Prothrombin (Fragment)	IPI00019568	70	201	29
SERPINC1 protein	IPI00844156	29	66	3
Serum amyloid A-4 protein	IPI00019399	15	3	2
Serum amyloid P-component	IPI00022391	25	37	11
Serum paraoxonase/arylesterase 1	IPI00218732	40	57	11
Transthyretin	IPI00022432 (+1)	16	35	6
Vitamin K-dependent protein S	IPI00294004 (+2)	75	20	5
Vitronectin	IPI00298971	54	80	13

the benefit of further fractionating the serum treated with peptide beads, a small fraction of the concentrated proteins by the ProteoMiner™ were tryptically digested and subjected to LC-MS/MS. This yielded 66 identified proteins that are listed in Table 1.

In the strategy developed here, the fractions collected from the three IMAC columns and the RPC column were subjected to solution digest and subsequently to LC-MS/MS analysis. The number of identified proteins in the collected fractions totaled 183 proteins corresponding to 66, 29, 55 and 33 proteins for the fractions from the IDA-Zn^{2+}, IDA-Ni^{2+}, IDA-Cu^{2+} and RPC columns, respectively, and are listed in Table 2. The number of proteins identified in the IDA-metal fractions increased in the order IDA-Zn^{2+} > IDA-Cu^{2+} > IDA-Ni^{2+}. The total number of non-redundant proteins identified from the solution digests of the four fractions was 82 proteins. In these fractions, the high abundance proteins identified were complement C3, serotransferrin, isoforms of α_1-antitrypsin and serum albumin. The other high abundance proteins, such as haptoglobin, α_2-macroglobulins and the majority of immunoglobulins, were not identified, which indicates that the peptide beads may not capture certain proteins for which no partner peptide ligands are present in the library or the number of beads having the particular peptide partner is low (i.e., low copy number) requiring for more beads to capture sufficient amount of the particular target protein. This finding is in agreement with the observation reported earlier by Sennels *et al.* [16]. As will be shown below, not only were some high abundance proteins were not captured by the peptide beads but also some other proteins passed through when only 1.0 mL serum was treated by 100 μL peptide beads, as is the case in this study.

To further monitor the effectiveness of the strategy described here, aliquots of concentrated fractions from the IMAC and the RPC columns were analyzed by 2-DE, and the corresponding 2-DE electropherograms of each fraction are shown in Figs. 1A, 1B, 1C and 1D. When the gels were stained with SYPRO fluorescent stain, the number of protein spots detected in the 2-DE for the IDA-Zn^{2+}, IDA-Ni^{2+}, IDA-Cu^{2+} and RPC fractions were 207, 159, 187 and 33, respectively, totaling 586 detected protein spots. But when the gels were stained with less sensitive Coomassie Blue, only 156 protein spots were detected. In the 156 proteins spots detected, 81 were from IDA-Zn^{2+}column, 17 were from IDA-Ni^{2+} column, 48 were from IDA-Cu^{2+} column and 10 were from the RPC column. Thus, the 183 proteins identified by LC-MS/MS in the solution digests of the 4 fractions obtained from the 4 tandem columns portray a sensitivity for LC-MS/MS that is at least 3.2 folds lower than the SYPRO staining but 1.2 fold more sensitive than the Coomassie Blue staining, assuming that each spot is a single protein. Although the ProteoMiner™ as used in this study was conducted at the microscale level using only 100 µL peptide beads and 1.00 mL of human serum, the relatively high number of protein spots detected using SYPRO fluorescent stain indicates that indeed the ProteoMiner™ beads effectively concentrate low abundance proteins. However, many of these spots are below the detection threshold of the LC-MS/MS and thus do not allow accurate identification of their proteins content.

Evaluation of the ProteoMiner™ and comparison with the depletion approach

The number of the identified proteins in serum that was subjected to the combinatorial peptide library beads (i.e., ProteoMiner™ treated serum) but without any

further fractionation was found to be 66 proteins (see Table 1) as opposed to 82 identified proteins with post IMAC/RPC columns fractionation (see Table 2). In the 66 identified proteins, 53 were also identified in the IMAC/RPC columns post fractionation while the remaining 13 proteins were not identified in the IMAC/RPC post fractionation. This results in 29 (82 - 53 = 29) proteins unique to IMAC and RPC fractionation, which correspond to a 35% increase in the number of non-redundant proteins [(29/82)*100 = 35%]. In comparison to a recent study by Sihlbom *et al.* [15], where the ProteoMiner™ treated serum was analyzed by differential gel electrophoresis, the number of identified non-redundant proteins was only 49, whereas in the present investigation, which involved the fractionation of the ProteoMiner™ treated serum on three IMAC columns and one RPC column, 82 non-redundant proteins were readily identified.

In a recent study from our laboratory [9], an albumin depleted serum sample was injected onto a series of tandem depletion columns to remove the next seven abundant proteins (i.e., α_1-antitrypsin, transferrin, IgG, IgA, IgM, haptoglobin, and α_2-macroglobulin), followed by a series of tandem IDA-metal chelate columns (Zn^{2+}, Ni^{2+}, and Cu^{2+}) and RPC column for online fractionation and concentration. This resulted in identification of 58 non-redundant protein from 2-DE gels with protein identification probability greater than 99%, peptide identification greater than 80% and at least 2 unique identified peptides. The number of proteins spots identified from the IDA-Zn^{2+}, IDA-Ni^{2+}, IDA-Cu^{2+} and the RPC columns were 14, 32, 37 and 10 respectively totaling up to 93 proteins (including the overall 58 non-redundant proteins). In the present study, and as mentioned earlier, the number of identified non-redundant proteins from the direct

TABLE 2

LC-MS/MS RESULTS OF THE PROTEINS IDENTIFIED FROM THE SOLUTION DIGEST

OF FRACTIONS FROM IMAC AND RPC COLUMNS

Identified Proteins	Accession #	Mol. Wt. KDa	Zn		Ni		Cu		RP	
			Spectra count	Unique peptide	Spectra count	Unique peptide	Spectra count	Unique peptide	Spectra count	Unique peptide
252 kDa protein	IPI00022937 (+1)	252	7	2						
Alpha-1-acid glycoprotein 1	IPI00022429 (+1)	24							27	6
Alpha-1-acid glycoprotein 2	IPI00020091	24							13	2
Alpha-2-HS-glycoprotein	IPI00022431	39	17	5						
Antithrombin III variant	IPI00032179	53	98	21	52	9	57	11	25	7
Apolipoprotein A1	IPI00853525	28					120	2		
Apolipoprotein A-I	IPI00021841	31	110	20	26	6	165	30	53	16
Apolipoprotein A-II	IPI00021854	11	35	8			29	8	17	5
Apolipoprotein A-IV	IPI00304273	45	39	16	7	3	60	20	58	14
Apolipoprotein B-100	IPI00022229	516	155	43			10	5	35	9
Apolipoprotein C-II	IPI00021856	11	9	3					2	2

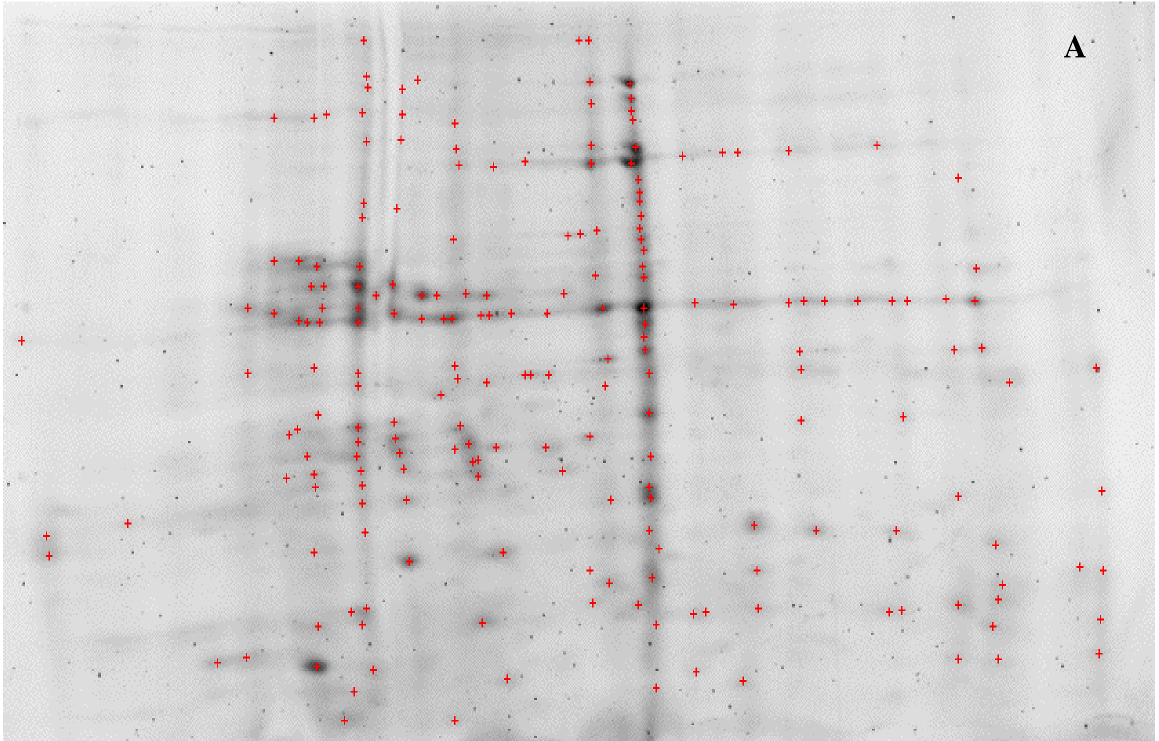
Apolipoprotein C-III	IPI00021857	11	11	2			11	2	14	3
Apolipoprotein D	IPI00006662	21	9	4			40	8	19	4
Apolipoprotein E	IPI00021842	36	99	21	47	10	52	15	49	11
C4b-binding protein alpha chain	IPI00021727	67	114	24						
CD5 antigen-like	IPI00025204	38	47	9	6	2	9	3		
cDNA FLJ58413, highly similar to Complement component C7	IPI00909594	54					29	2		
Ceruloplasmin	IPI00017601	122	12	5			9	3		
Clusterin	IPI00291262	52	237	30	61	14	122	25	18	5
Complement C1q subcomponent subunit A	IPI00022392	26	8	3						
Complement C1q subcomponent subunit C	IPI00022394	26	25	5	4	2				
Complement C1s subcomponent	IPI00017696 (+1)	77	6	3			32	7		
Complement C3 (Fragment)	IPI00783987	187	428	92	85	25	138	38	156	28
Complement C5	IPI00032291	188	23	11						
complement component 1, q subcomponent, B chain precursor	IPI00477992 (+1)	27	19	5						
Complement component 4B	IPI00887154	193	590	92	198	47	228	49	50	15

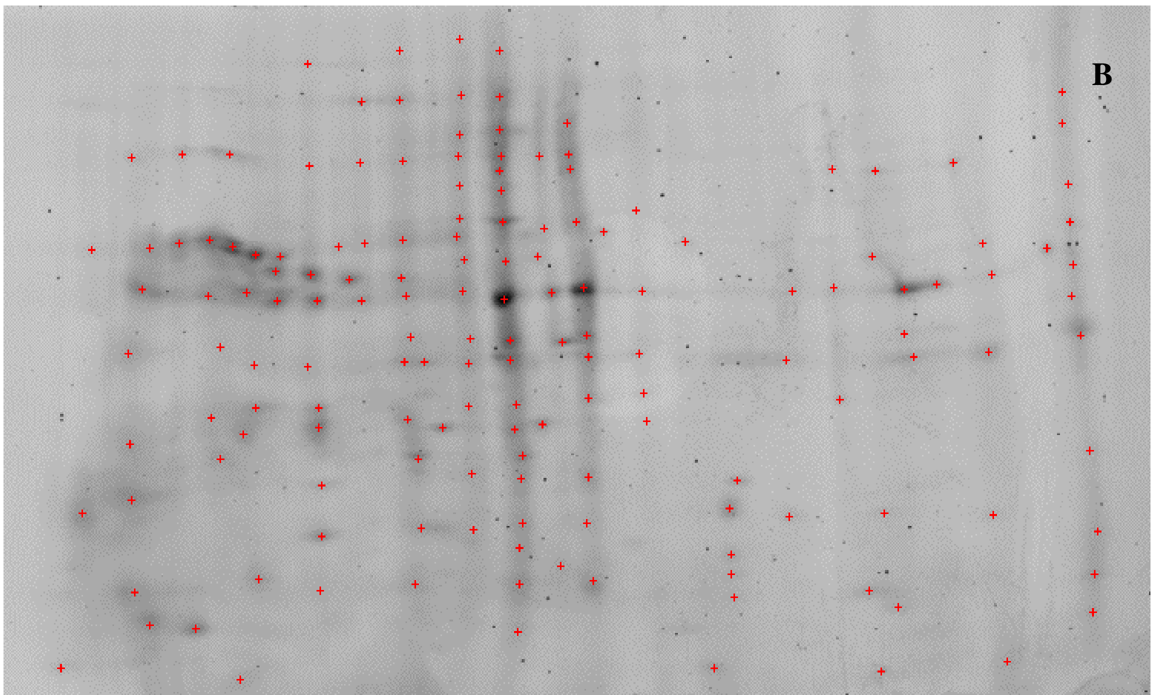
Complement component 6 precursor	IPI00879709	106	41	14						
Complement component C7	IPI00296608	94			10	3	44	12		
Complement component C9	IPI00022395	63	16	4			23	7	11	3
Complement factor H-related protein 1	IPI00011264 (+1)	38	50	7						
Complement factor H-related protein 4A	IPI00644977	65					43	9		
Complement-activating component of Ra-reactive factor	IPI00299307 (+1)	79	3	2			2	2		
Fibronectin 1 isoform 4 preproprotein	IPI00414283 (+1)	257	583	97	33	8	41	10		
FLJ00385 protein (Fragment)	IPI00168728 (+6)	56	70	6	29	5	48	5		
Galectin-3-binding protein	IPI00023673	65	40	12			28	8	7	2
Glutathione peroxidase 3	IPI00026199	26	7	3	11	2	16	4		
IGHA1 protein	IPI00430842	53	24	6	20	5	41	8		
IGHM protein	IPI00477090	67	120	24	44	10	71	16		
IGHM protein	IPI00472610	53	70	14	35	10	75	12		
IGKV1-5 protein	IPI00419424 (+2)	26	51	8	27	4	53	8		
IGL@ protein	IPI00154742	25	46	7	16	4	50	7		

IGL@ protein	IPI00719373 (+6)	23					26	2		
immunoglobulin J chain	IPI00178926	18	13	3	9	2	11	3		
Inter-alpha-trypsin inhibitor heavy chain H1	IPI00292530	101	19	5					9	2
Inter-alpha-trypsin inhibitor heavy chain H2	IPI00305461 (+1)	106	38	11			4	3	12	4
Isoform 1 of Alpha-1-antitrypsin	IPI00553177 (+1)	47					8	3		
Isoform 1 of C4b-binding protein beta chain	IPI00025862 (+1)	28	9	3						
Isoform 1 of Complement factor H	IPI00029739	139	196	42						
Isoform 1 of Extracellular matrix protein 1	IPI00003351 (+1)	61	21	7						
Isoform 1 of Ficolin-2	IPI00017530 (+2)	34	17	4			7	2		
Isoform 1 of Ficolin-3	IPI00293925	33	46	9						
Isoform 1 of Inter-alpha-trypsin inhibitor heavy chain H3	IPI00028413 (+3)	100							5	2
Isoform 1 of Pregnancy zone protein	IPI00025426 (+2)	164	15	6			13	3		
Isoform 1 of Serum albumin	IPI00745872	69	92	24	16	6	237	45	174	41
Isoform 2 of Apolipoprotein L1	IPI00186903 (+2)	46							14	2
Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain	IPI00218192 (+4)	101					5	2		

H4										
Isoform C of Fibulin-1	IPI00296537	74	96	4						
Isoform D of Fibulin-1	IPI00296534	77	97	20	30	9	11	3		
Isoform Long of Complement factor H-related protein 2	IPI00006154 (+1)	31	20	2						
Keratin, type I cytoskeletal 10	IPI00009865	60	41	14	45	17	43	12	85	21
Keratin, type I cytoskeletal 9	IPI00019359	62	25	6	45	15	16	5	36	11
Keratin, type II cytoskeletal 1	IPI00220327	66	52	13	65	17	36	10	51	17
Keratin, type II cytoskeletal 2 epidermal	IPI00021304	66	23	5	33	5	31	4	37	9
Keratin, type II cytoskeletal 5	IPI00009867 (+1)	62							8	2
Lipopolysaccharide-binding protein	IPI00032311	53	17	3						
Plasminogen	IPI00019580	91	18	7			36	12		
Properdin	IPI00021364	51	61	14						
Protein AMBP	IPI00022426	39	13	3			7	4	7	2
Protein Z-dependent protease inhibitor	IPI00007199	55	12	2						
Prothrombin (Fragment)	IPI00019568	70	131	24	43	9	115	25	162	26

Putative uncharacterized protein DKFZp686C15213	IPI00426051	51	35	3			30	3		
Secreted phosphoprotein 24	IPI00011832	24	10	2			6	2		
Serotransferrin	IPI00022463	77					31	9		
Serpin peptidase inhibitor, clade D (Heparin cofactor), member 1	IPI00292950 (+1)	60	6	2						
SERPINC1 protein	IPI00844156	29	53	3	27	3	36	4	11	2
Serum amyloid A-4 protein	IPI00019399	15					14	2		
Serum amyloid P-component	IPI00022391	25	33	8	15	5	14	7	19	3
Serum paraoxonase/arylesterase 1	IPI00218732	40	26	8			11	2		
Transthyretin	IPI00022432 (+1)	16	13	3			30	7	11	2
Vitamin K-dependent protein C	IPI00021817 (+2)	52					16	4		
Vitamin K-dependent protein S	IPI00294004 (+1)	75							7	2
Vitronectin	IPI00298971	54	133	19	77	15	92	14	55	10


A


B

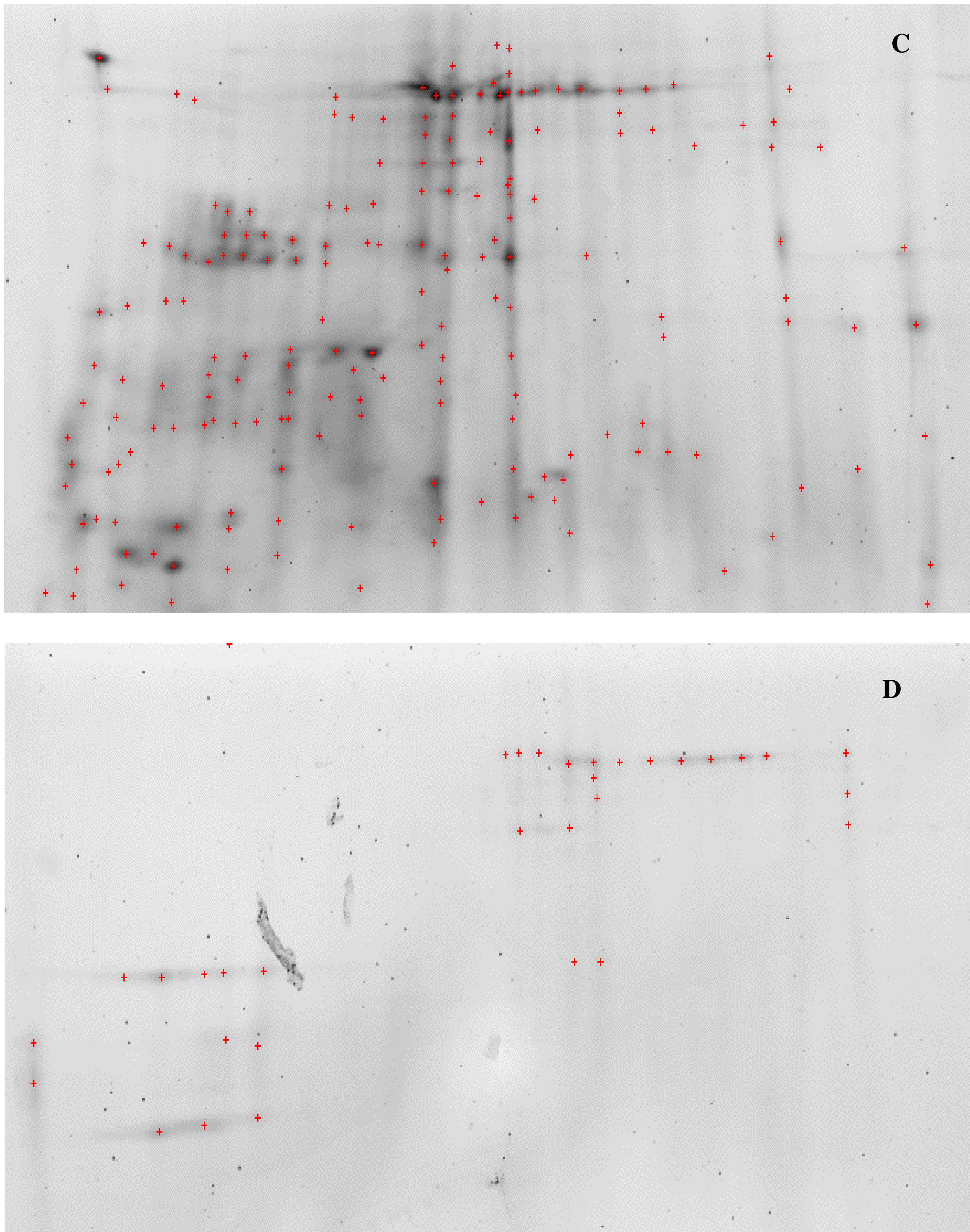


Figure 1. *2-DE electropherograms of the serum after treatment with the hexapeptide beads and fractionation/concentration on (A) IDA-Zn^{2+}, (B) IDA-Ni^{2+}, (C) IDA-Cu^{2+} and (D) the RPC-column. 2-DE experiments were performed on a 7 cm IPG strip (pH 4-7) in*

the first dimension and subsequently on a 10-20% Tris-HCl 8.6 x 6.8 cm gels in the second dimension. The gels were stained with SYPRO fluorescent stain.

solution digest of the fractions from the IDA-Zn^{2+}, IDA-Ni^{2+}, IDA-Cu^{2+} and the RPC columns increased to 66, 29, 55 and 33 respectively which totaled to 183 proteins (including the overall 82 non-redundant proteins). Although the number of non-redundant proteins identified by the depletion approach is about 30% less than those identified by the ProteoMiner™ approach, 22 identified proteins were common to both approaches while 36 identified proteins were unique to depletion and 60 identified proteins were unique to ProteoMiner™ (see Table 3). Whereas the 60 identified proteins unique to ProteoMiner™ is a significant gain over the depletion approach, the 36 identified proteins unique to depletion reflects a significant number of proteins that the ProteoMiner™ approach failed to capture them in sufficient amount in addition to some high abundance proteins (see above) to allow their identification by LC-MS/MS. In this investigation, only 100 µL peptide beads were used to treat 1.0 mL of serum. These 36 identified proteins were captured and identified by ProteoMiner™ when 1.0 mL of peptide beads was used to treat 300 mL of pooled serum [16]. However, 300 mL of pooled serum may available to obtain from species other than human. In fact, only 0.20 mL of serum can be obtained from small mice [18]. On the other hand, in the depletion approach, even though the high abundance proteins were depleted, the remaining medium abundant proteins may mask or suppress the low concentrated proteins in 2-DE and also in mass analysis, a fact that contributed to the inability of the depletion approach to allow the detection of the 60 identified proteins unique to the ProteoMiner™. In the case of

ProteoMiner™, since the high abundance proteins are diluted and the low abundant proteins are concentrated simultaneously the problem of masking or suppressing does not happen in this case.

Highlights of the retention of some typical proteins on the different fractionation columns

Although our aim in this work is mainly to evaluate the strategy under investigation in the capturing efficiency of low abundance proteins in conjunction with fractionation/concentration on three IMAC and one RPC columns, the discussion of the retention behaviors of some representative proteins would highlight at least in part the underlying retention mechanism. Also, due to the fact that a large number of proteins is involved, a highlight of the retention of some captured proteins on the various columns would be the most appropriate.

The IDA-Zn^{2+} column, which was placed first in the tandem column series, selectively captured some of the serum proteins such as 252 KDa protein, α-2-HS-glycoprotein, C4b-binding protein alpha chain, complement C1q subcomponent subunit A, complement C5, complement component 1q subcomponent B chain precursor, complement component 6 precursor, complement factor H-related protein 1, isoform 1 of C4b-binding protein beta, isoform 1 of complement factor H, isoform 1 of extracellular matrix protein, isoform 1 of ficolin-3, isoform C of fibulin-1, isoform long of complement factor H related protein 2, liposaccharide binding protein, properdin, protein Z-dependent protease inhibitor and serpin peptidase inhibitor clade D (Heparin cofactor) member 1. As shown by Sulkowski [19], IDA-Zn^{2+} column can retain proteins that have proximal His residues on the protein surface. As a typical example, α-2-HS-glycoprotein

TABLE 3

COMPILATION OF IDENTIFIED PROTEINS UNIQUE TO PROTEOMINER™-POST FRACTIONATION TREATED SERUM, IDENTIFIED PROTEINS UNIQUE TO DEPLETION-POST FRACTIONATION TREATED SERUM AND IDENTIFIED PROTEINS COMMON TO BOTH PROTEOMINER™- AND DEPLETION-POST FRACTIONATION TREATED SERUM

60 Identified Proteins Unique to ProteoMiner™		
252 kDa protein	Complement-activating component of Ra-reactive factor	Isoform 2 of Apolipoprotein L1
Alpha-1-acid glycoprotein 1	Fibronectin 1 isoform 4 preproprotein	Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4
Apolipoprotein A1	FLJ00385 protein (Fragment)	Isoform C of Fibulin-1
Apolipoprotein B-100	Galectin-3-binding protein	Isoform D of Fibulin-1
Apolipoprotein C-II	Glutathione peroxidase 3	Isoform Long of Complement factor H-related protein 2
Apolipoprotein C-III	IGHA1 protein	Keratin, type II cytoskeletal 2 epidermal
Apolipoprotein D	IGHM protein	Keratin, type II cytoskeletal 5
Apolipoprotein E	IGHM protein	Lipopolysaccharide-binding protein
C4b-binding protein alpha chain	IGKV1-5 protein	Plasminogen
CD5 antigen-like	IGL@ protein	Properdin
cDNA FLJ58413, highly similar to Complement component C7	IGL@ protein	Protein AMBP
Clusterin	Isoform 1 of Alpha-1-antitrypsin	Protein Z-dependent protease inhibitor
Complement C1q subcomponent subunit A	Isoform 1 of C4b-binding protein beta chain	Putative uncharacterized protein DKFZp686C15213
Complement C1q subcomponent subunit C	Isoform 1 of Complement factor H	Secreted phosphoprotein 24

Complement C1s subcomponent	Isoform 1 of Extracellular matrix protein 1	Serpin peptidase inhibitor, clade D (Heparin cofactor), member 1
Complement C5	Isoform 1 of Ficolin-2	SERPINC1 protein
complement component 1, q subcomponent, B chain precursor	Isoform 1 of Ficolin-3	Serum amyloid A-4 protein
Complement component 4B	Isoform 1 of Inter-alpha-trypsin inhibitor heavy chain H3	Serum paraoxonase/arylesterase 1
Complement factor H-related protein 1	Isoform 1 of Pregnancy zone protein	Vitamin K-dependent protein C
Complement factor H-related protein 4A	Isoform 1 of Serum albumin	Vitamin K-dependent protein S
36 Identified Proteins Unique to Depletion		**22 Identified Proteins Common to Depletion and** ProteoMiner™
Afamin	Heparin cofactor 2	Alpha-1-acid glycoprotein 2
Alpha-1-antichymotrypsin	Hyaluronan-binding protein 2	Alpha-2-HS-glycoprotein
Alpha-1-antitrypsin	Ig alpha-1 chain C region	Antithrombin-III
Alpha-1B-glycoprotein	Ig gamma-1 chain C region	Apolipoprotein A-I
Angiotensinogen	Ig kappa chain C region	Apolipoprotein A-II
Beta-2-glycoprotein 1	Ig kappa chain V-III region B6	Apolipoprotein A-IV
Complement C4-A	Ig lambda chain C regions	Ceruloplasmin
Complement component C8 alpha chain	Ig mu chain C region	Complement component C6
Complement component C8 beta chain	Inter-alpha-trypsin inhibitor heavy chain H4	Complement component C7
Complement factor B	Kininogen-1	Complement component C9
Complement factor H	Leucine-rich alpha-2-glycoprotein	Immunoglobulin J chain
Complement factor I	Lumican	Inter-alpha-trypsin inhibitor heavy chain H1
Corticosteroid-binding globulin	Pigment epithelium-derived factor	Inter-alpha-trypsin inhibitor heavy chain H2
C-reactive protein	Plasma protease C1 inhibitor	Keratin, type I cytoskeletal 10
Fibronectin	Retinol-binding protein 4	Keratin, type I cytoskeletal 9
Ficolin-3	Thyroxine-binding globulin	Keratin, type II cytoskeletal 1

Gelsolin	Vitamin D-binding protein	Prothrombin
Hemopexin	Zinc-alpha-2-glycoprotein	Serotransferrin
		Serum amyloid P-component
		Transthyretin
		Vitronectin

is a protein which was completely retained by the IDA-Zn^{2+} column, and this retention behavior is in agreement with the previous findings reported in ref. [20] in the sense that α-2-HS-glycoprotein has His residues that are favorable for its retention on IDA-Zn^{2+}. Also, it was reported that α-2-HS-glycoprotein has similar dissociation constants for IDA-Zn^{2+} and IDA-Cu^{2+} columns [20]. However, since it was placed first in the tandem series, the IDA-Zn^{2+} column captured all of the α-2-HS-glycoprotein, thus allowing more affinity sites on the IDA-Cu^{2+} column to be available for the retention of other proteins.

Another protein whose retention on IDA-Zn^{2+} column merits discussion is isoform C of fibulin-1. This protein, which contains 18 His residues with a protein length of 683 amino acid residues [21] was detected only in the IDA-Zn^{2+} fraction whereas isoform D of fibulin-1 (protein length of 703 amino acid residues including 20 His residues [22] was also detected in the fractions of the other 2 IDA-metal chelate columns. This may be due to the difference in the relative concentrations of isoform C and isoform D, which can be assessed from the 4 unique peptides of isoform C of fibulin-1 versus 20 unique peptides of isoform D of fibulin-1 in the IDA-Zn^{2+} fraction. The number of unique peptides for isoform D of fibulin-1 captured by the IDA-Ni^{2+} and IDA-Cu^{2+} columns were 9 and 3, respectively. The fact that the isoforms of fibulin-1 were retained primarily on the IDA-Zn^{2+} column may be indicative of the presence of vicinal His residue which favor their retention on the IDA-Zn^{2+} column that is the first column in the tandem column format. The isoform D of fibulin-1, which seems to be in larger concentration than isoform C would then overflow the sites of the IDA-Zn^{2+} column and retain on the other IMAC columns. In another example, complement factor H-related protein 1, a glycosylated protein with either one or two carbohydrate chains attached

[23], was captured by the IDA-Zn^{2+} column alone, whereas another protein from the same factor H protein family, complement factor H-related 4A, was captured only on the IDA-Cu^{2+} column. This might be due to the difference in the glycosylation at the metal binding site.

As listed in Table 2, complement component C7 was captured by both IDA-Ni^{2+} and IDA-Cu^{2+} columns, but was not captured by the IDA-Zn^{+2}column. This protein has 17 His residues and since it was not captured by the IDA-Zn^{2+}, one can envision that the protein lacks the presence of vicinal His residue on its surface. However, the spectra count in Table 2 indicates that most of the protein was captured by the IDA-Cu^{2+} column rather than by the IDA-Ni^{2+} column. Ceruloplasmin and transferrin were not captured by IDA-Ni^{2+} but were captured by the other IDA-metal columns and the RPC column. Similar results on tris(carboxymethyl)ethylenediamine (TED)-agarose columns loaded with Ni^{2+} (TED-Ni^{2+}) were reported in the sense that the TED-Ni^{2+} column did not have any affinity towards both ceruloplasmin and transferrin [24]. It was suggested [24], that the His residues on the proteins were not accessible for interactions with the Ni chelate column.

As reported by Sulkowski [19], the presence of a single His residue on the surface of a given protein is sufficient for its retention on an IDA-Cu^{2+} column. This selectivity is illustrated in the case of serum amyloid A-4 protein that was captured by the IDA-Cu^{2+} column alone. This protein has 130 amino acids in the protein sequence and has only one His residue, which is necessary for its retention on an IDA-Cu^{2+} column. This protein was neither captured by IDA-Zn^{2+} nor by the IDA-Ni^{2+} columns, which were placed before the IDA-Cu^{2+} column in the tandem series. The RPC column, which followed the

IDA-Cu^{2+} in the tandem column series captured some of the serum amyloid A-4 protein that escaped from the IDA-Cu^{2+} column. Furthermore, proteins captured by IDA-Cu^{2+} column alone were apolipoprotein A1, complement factor H-related protein 4A, IGL@protein, serotransferrin, isoform 1 of alpha-1-antitrypsin, isoform 2 of inter-alpha-trypsin inhibitor heavy chain H4, and vitamin K-dependent protein C. Previously, it was observed by Wu and Bruley that vitamin K-dependent protein C has high affinity towards IDA-Cu^{2+} [25]. This phenomenon is also seen in the present study where it was not captured by other IDA-metal columns but only by IDA-Cu^{2+} column with a small amount of the protein that escaped was captured by the RPC column. Apolipoprotein A1 protein was not captured by IDA-Zn^{2+} or IDA-Ni^{2+} columns, but was captured only by the IDA-Cu^{2+} column. This could be due to the less number of accessible His residues present on the surface of the protein.

Only the RPC column captured the proteins alpha-1-acid glycoprotein 1 and alpha-1-acid glycoprotein 2. Both proteins have only three His residues with a protein length of 201 amino acids. Also, 40% of these proteins contain carbohydrates [26] which might mask the His groups, and that could explain why these proteins were not retained on the IMAC columns but were retained by the RPC column.

The protein apolipoprotein A-II does not have any His group on its surface to have retention on the IDA-metal columns. But as can be seen in Table 2, the IDA-metal and the RPC columns captured this protein. As indicated by Blanco-Vaca, *et al* [27], apolipoproteins containing cysteine residues can form disulfide bonds with the other cysteine containing apolipoproteins. For example, apolipoprotein A-II can form a heterodimer with apolipoprotein D, which has His groups in its protein sequence [27].

This formation of dimers would enable the retention of apolipoproteins even if they lack the presence of His residues.

Conclusions

ProteoMiner™ and immuno-subtraction treatments of biological fluids can be viewed as complementary approaches for facilitating the comprehensive profiling of proteomics samples. As shown in this report, for a limited sample size of a given biological fluid, the ProteoMiner™ allows the detection of a large number of low abundance proteins whereas the immuno-subtraction permits the detection of medium abundance proteins for which the peptide beads may not have the partner peptide beads in large copy number when a limited size of peptide beads are used such as 100 µL. Both immuno-subtraction and ProteoMiner™ approaches may require post fractionation for comprehensive sample proteomics profiling.

References

[1] Jmeian, Y., El Rassi, Z., *Electrophoresis* 2009, *30*, 249-261.

[2] Wehr, T., *LCGC North America* 2008, *26*, 278-284.

[3] Zolotarjova, N., Mrozinski, P., Majors, R. E., *LCGC* 2007, *25*, 118-140.

[4] Righetti, P. G., Boschetti, E., *FEBS Journal* 2007, *274*, 897-905.

[5] Righetti, P. G., Boschetti, E., Lomas, L., Citterio, A., *Proteomics* 2006, *6*, 3980-3992.

[6] Righetti, P. G., Castagna, A., Antonioli, P., Boschetti, E., *Electrophoresis* 2005, *26*, 297-319.

[7] Jmeian, Y., El Rassi, Z., *J. Proteome Res.* 2007, *6*, 947-954.

[8] Jmeian, Y., El Rassi, Z., *Electrophoresis* 2008, *29*, 2801-2811.

[9] Jmeian, Y., El Rassi, Z., *J. Proteome Res.* 2009, *8*, 4592-4603.

[10] Thulasiraman, V., Lin, S., Gheorghiu, L., Lathrop, J., Lomas, L., Hammond, D., Boschetti, E., *Electrophoresis* 2005, *26*, 3561-3571.

[11] Boschetti, E., Lomas, L., Citterio, A., Righetti, P. G., *J. Chromatogr. A* 2007, *1153*, 277-290.

[12] Luo, Q., Zou, H., Xiao, X., Guo, Z., Kong, L., Mao, X., *J. Chromatogr. A* 2001, *926*, 255-264.

[13] Keidel, E.-M., Ribitsch, D., Lottspeich, F., *Proteomics* 2010, *10*, 2089-2098.

[14] Li, L., Sun, C. J., Freeby, S., Yee, D., Kieffer-Jaquinod, S., Guerrier, L., Boschetti, E., Lomas, L., *J. Proteomics Bioinform.* 2009, *2*, 485-494.

[15] Sihlbom, C., Kanmert, I., Bahr, H. v., Davidsson, P., *J. Proteome Res.* 2008, *7*, 4191-4198.

[16] Sennels, L., Salek, M., Lomas, L., Boschetti, E., Righetti, P. G., Rappsilber, J., *J. Proteome Res.* 2007, *6*, 4055-4062.

[17] Fasoli, E., Farinazzo, A., Sun, C. J., Kravchuk, A. V., Guerrier, L., Fortis, F., Boschetti, E., Righetti, P. G., *J. Proteomics* 2010, *73*, 733-742.

[18] Hoff, J., *Lab Animal* 2000, *29*, 47-53.

[19] Sulkowski, E., *Trends Biotechnol.* 1985, *3*, 1-7.

[20] Goubran-Botros, H., Nanak, E., Abdul Nour, J., Birkenmeir, G., Vijayalakshimi, M. A., *J. Chromatogr.* 1992, *597*, 357-364.

[21] Argraves, W. S., Tran, H., Burgess, W. H., Dickerson, K., *J. Cell Biology* 1990, *111*, 3155-3164.

[22] Tran, H., Mattei, M., Godyna, S., Argraves, W. S., *Matrix Biology* 1997, *15*, 479-493.

[23] Zipfel, P. F., Jokiranta, T. S., Hellwage, J., Koistinen, V., Meri, S., *Immunopharmacology* 1999, *42*, 53-60.

[24] Porath, J., Olin, B., *Biochemistry* 1983, *22*, 1621-1630.

[25] Wu, H., Bruley, D. F., *Biotechnol. Prog.* 1999, *15*, 928-931.

[26] Urien, S., Bree, F., Testa, B., Tillement, J. P., *Biochemical J.* 1991, *280*, 277-280.

[27] Blanco-Vaca, F., Via, D. P., Yang, C. Y., Massey, J. B., Pownall, H. J., *J. Lipid Res.* 1992, *33*, 1785-1796.

CHAPTER III

LECTIN AFFINITY CHROMATOGRAPHY USING LECTINS OF BROAD SPECIFICITY FOR CAPTURING SUB-GLYCOPROTEOMICS FROM BREAST CANCER AND DISEASE FREE HUMAN SERA USING TANDEM MONOLITHIC COLUMNS WITH SURFACE IMMOBILIZED CONCANAVALIN A, WHEAT GERM AGGLUTININ AND *RICINUS COMMUNIS* AGGLUTININ-I

Introduction

In many aspects, glycosylation is one of the most important post-translational modifications of proteins. In one biological implication, glycosylation undergoes significant alteration in many diseases, especially in cancer [1, 2]. More than 50% of human serum proteins are glycosylated, thus making this easily accessible biological fluid a rich source of information about the patho-physiological state of a person. Identification of serum proteins that are up or down regulated, as a result of disease or response to a given treatment in comparison to the disease-free serum represent an important undertaking in clinical studies. Since a few cancer biomarkers such as

carcinoembryonic antigen, cancer antigen 125 and prostate-specific antigen, which are serum glycoproteins, are currently reliable markers in clinical diagnosis, the search of other glycoprotein-based cancer biomarkers is an intensive theme of investigation by researchers.

Based on the above consideration, the availability of well-developed and tested analytical approaches for specific capturing of the glycoprotein information represent an important theme of research in the biological and biomedical areas. In particular, the development of liquid phase separation methods for fractionating and concentrating glycoproteins that are differentially expressed in cancerous serum relative to disease-free serum are badly needed in the area of glycoprotein biomarkers. However, the wide dynamic range of protein concentration and the many thousands of proteins present in serum poses a major challenge to analyze the differentially expressed glycoproteins.

In this chapter, the complexity of human serum was addressed by using lectin-based affinity pre-fractionation before identifying the proteins using LC-MS/MS. Even though depletion of high abundance proteins serves to be a good pre-fractionation approach it has the disadvantage that it can result in co-depletion of some of the low abundance proteins that might have valuable information about the disease in question [3].

In addition to enrichment of glycoproteins using a single lectin column [4], both serial-lectin affinity chromatography (LAC) [5-8] and multi-LAC (M-LAC) [9-13] have been reported for the enrichment of glycoproteins in human serum/plasma. Serial-LAC is a separation approach where multiple lectin columns are used in series or in a sequential order to capture glycoproteins, whereas in M-LAC, the lectins to be used are mixed in

proportions and immobilized onto a given stationary phase. Thereafter, the lectin column is eluted with the specific displacer required or the so-called haptenic sugar for each lectin. Madera *et al.* [14] used 4 different lectins immobilized on macroporous silica for enrichment of the glycoproteins from serum. The immobilized lectins were Concanavalin A (Con A), *Sambuccus nigra* agglutinin (SNA), *Ulex europaaeus* agglutinin-I (UEA-I) and *Phaseolus vulgaris* agglutinin-L (PHA-L). An evaluation of serial-LAC and M-LAC was made, and it was found that serial-LAC resulted in the identification of higher number of proteins than M-LAC [14]. Some of the serial-LAC studies performed so far targeted specific types of glycoconjugates such as the *O*-linked glycopeptides, sialic acid containing *N*-linked complex-type glycoforms and fucosylated glycoproteins that are present in human serum [6-8]. However, other glycoforms are also involved in most of the diseases, thus a method that will capture a broad range of glycoforms of both *N*-linked and *O*-linked glycoproteins from human serum is required. With this objective in mind, a serial-LAC (referred to as tandem lectin columns here) using three broad specificity lectins, including Con A, wheat germ agglutinin (WGA) and *Ricinus communis* agglutinin-I (RCA-I) was performed using human serum to specifically capture a range of glycoforms thus reducing the glycoproteome complexity. This platform was demonstrated in identifying the serum proteins that were altered in their glycosylation due to breast cancer.

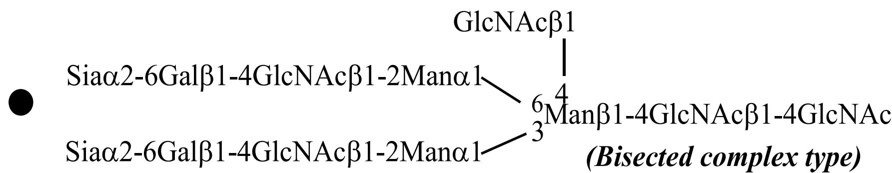
The chromatographic separation media used in proteomics/glycoproteomics is an important factor for rapid analysis of samples. In most of the lectin enrichment studies, the solid supports used have been agarose, macroporous silica and polyhydroxylated polymer (POROS). Even though agarose has been frequently used, it has certain

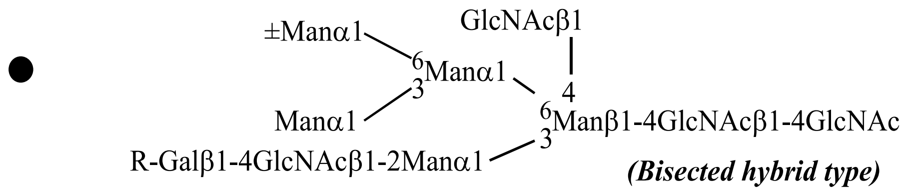
disadvantages like high backpressure and low flow rate, which leads to increased analysis time whereas silica has limitations due to its pH compatibility. Polymer based monolithic supports for LAC offer high flow rates, low backpressure, and consequently reduced analysis time. Bedair and El Rassi reported lectins immobilized on polymethacrylate monoliths for rapid and nano-scale separation of glycoproteins [15]. In this study polymethacrylate based monolith was used as the chromatographic separation medium in LAC.

This study involved the immobilization of three different lectins of the broad specificity type on the surface of glyceryl methacrylate (GMM)/ethylene glycol dimethacrylate (EDMA) monolith. They were WGA, Con A and RCA-I. While WGA and Con A have specificities directed toward the core portion of *N*-glycans, RCA-I specifically interacts with the non-reducing terminal moieties of the outer chain structures of *N*-glycans. Immobilized WGA interacts strongly with *N,N'*-diacetylchitobiose, bisected hybrid type and complex type *N*-glycans and glycoproteins with clustered sialic acid residues. Oligosaccharides with poly-*N*-acetyllactosamine chains have weak affinity and are retarded in the column (see Fig. 1) [16, 17]. The haptenic monosaccharide for WGA is *N*-acetyl-D-glucosamine (GlcNAc). Immobilized Con A binds strongly to high mannose type and hybrid type *N*-glycans. Presence of bisecting GlcNAc markedly changes the conformation of the trimannosyl core and weakens the interaction with a Con A column. Presence of α1→ 3 linked fucose at the GlcNAc residue of Galβ1 → 4 GlcNAc group or NeuNAcα2 → 6 at the GlcNAc of the Galβ1 → 3 GlcNAc group in the outer chain moieties interferes with the binding of *N*-glycans to a Con A column by steric hindrance. Con A has weak binding to biantennary complex type *N*-glycans (see Fig. 2)

[16, 17]. The haptenic monosaccharide for Con A is methyl-α-D-mannopyranoside (Me-α-D-Man). RCA-I interacts with oligosaccharide chains by recognizing the β-galactosyl residue of *N*- and *O*-glycans. The Galβ1 → 4 GlcNAc group strongly binds to immobilized RCA-I and is eluted with 10 mM lactose in the eluting mobile phase (see Fig. 3) whereas oligosaccharides with Galβ1 → 3 GlcNAcβ1 → 6 or Galβ1 → 6 residues at their non-reducing termini are only retarded in the column (i.e., weak binding affinity) [16, 18, 19]. The haptenic sugar for RCA-I is lactose.

➡ ***Structures exhibiting strong binding and requiring 0.1 M N-acetylglucosamine for elution***

- GlcNAcβ1-4GlcNAc
- GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc, structural determinant for lectin recognition
- 

- 

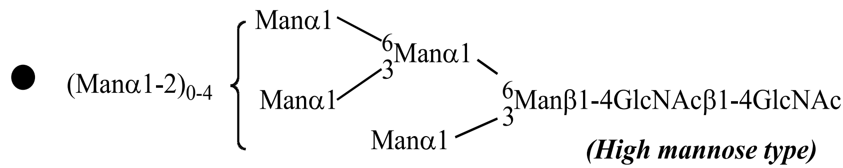
- Cluster of sialic acid residues

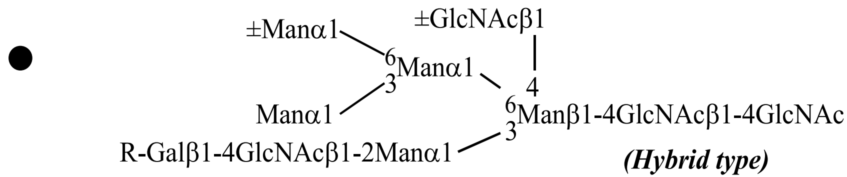
➡ ***Structures possessing weak binding***

- Poly-N-acetyllactosamine chains

Figure 1. *Affinity of immobilized WGA toward N-glycan structures.*

Structures exhibiting strong binding and requiring 0.2 M methyl-α-D-mannopyranoside for elution





Structures possessing weak binding and requiring 5 mM methyl-α-D-glucopyranoside for elution

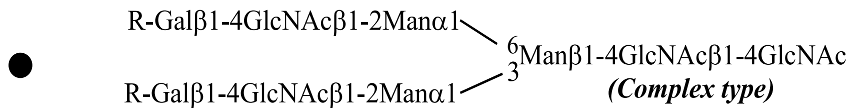


Figure 2. *Affinity of immobilized Con A toward N-glycan structures.*

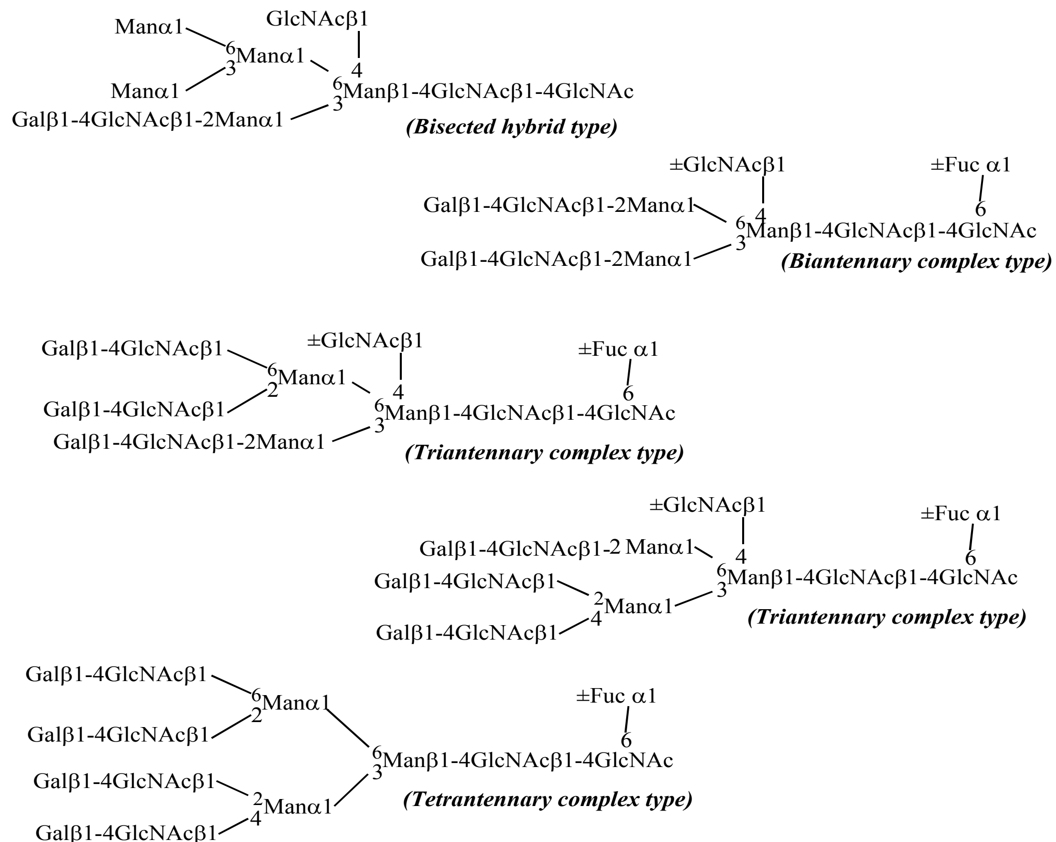


Figure 3. *Sugars binding specificities of immobilized RCA-I.*

Experimental

Instrumentation

The instrumental setup was made using a solvent delivery system Model CM3500 and CM4000 and two metering pumps Model III CM from Milton Roy, LDC division (Riviera Beach, FL, USA) with a Rheodyne injector Model 7010 (Cotati, CA, USA) equipped with a 100 μL loop and a Model 200 UV-Vis variable wavelength detector from Linear Instruments (Reno, NV, USA). All mass spectra were obtained using a hybrid LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA).

Reagents and Materials

Breast cancer of infiltrating ductal carcinoma (IDC) type cancer from one donor and pooled disease-free human serum from 3 different donors (same age group and race as the cancer serum) was purchased from Innovative Research (Southfield, MI, USA). GMM, EDMA, 2,2´-azobis(isobutyronitrile) (AIBN), 1-dodecanol, sodium periodate, sodium cyanoborohydride, *N*-acetyl-D-glucosamine and methyl α-D-mannopyranoside were purchased from Aldrich Chemical Co. (Milwaukee, WI, USA). Cyclohexanol and HPLC grade acetonitrile were obtained from Fisher Scientific (Fair Lawn, NJ, USA). Bio-safe Coomassie stain and a Bradford protein assay kit were purchased from Bio-Rad (Hercules, CA). The three unconjugated lectins namely, Con A, RCA-I and WGA were purchased from Vector Laboratories (Burlingame, CA, USA)

Monolithic Affinity Columns

A well mixed and degassed polymerization mixture of 18% GMM, 12% EDMA, 35% cyclohexanol and 35% dodecanol containing 1.0% (w/w) AIBN with respect to the monomers [20] was introduced into a 25.0 cm x 4.6 mm ID stainless steel column that functions as a mold for the monolith and was heated at 50 ˚C for 24 h. The resulting monolithic column was washed extensively with acetonitrile followed by water. The monolithic support was transferred from the 25.0 cm column to a shorter column (10 cm) by connecting the two columns with a ¼"-union and running water through the columns at flow rate of 3.0 mL/min until the modified monolithic support is transferred.

Immobilization of Lectins

The 10-cm monolithic columns were allowed to react with freshly prepared 0.1 M $NaIO_4$ for 7 h at room temperature. The columns were washed with water and the lectins were immobilized immediately. The immobilization was done on the column by passing a solution of 10 mg of Con A in 1 mL of 0.1 M sodium acetate at pH 6.4 containing 1mM of $CaCl_2$, 1mM $MgCl_2$, 1 mM $MnCl_2$, 0.1 M Me-α-D-Man and 50 mM sodium cyanoborohydride through the column for overnight at room temperature. WGA was immobilized using the same procedure by passing 1 mL of a solution containing 10 mg of WGA in 0.1 M sodium acetate at pH 6.4 containing 50 mM sodium cyanoborohydride and 0.1 M GlcNAc. RCA-I was immobilized by passing 3 mL of solution containing 10 mg of RCA-I in 0.1 M sodium acetate at pH 6.4 containing 50 mM sodium cyanoborohydride and 0.1 M lactose. To the resulting column, reductive amination of any unreacted aldehyde group was done by passing a solution of 0.4 M Tris-HCl, pH 7.2

containing 50 mM sodium cyanoborohydride for 3 h at room temperature. These immobilized lectin columns were stored with the mobile phase containing 20 mM Tris-HCl containing 100 mM NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM $CaCl_2$, pH 6.0 and 0.08% NaN_3 at 4 °C until use.

<u>Fractionation of Glycoproteins from Human Serum - Chromatographic Conditions</u>

The lectin columns were arranged in tandem series. In all the experiments, serum was used in 1:3 dilution ratio and a total of 100 μL (2 x 50 μL) was injected for the investigation of effect of serial order of the lectin columns or 250 μL (5 x 50 μL) was injected for comparison of disease-free and breast cancer. The columns were first equilibrated with 10 column volumes of the binding mobile phase consisting 20 mM Tris-HCl containing 100 mM NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM $CaCl_2$, pH 6.0. Then, 50 μL of the diluted serum was injected onto the tandem columns and the unbound serum proteins were washed away with the binding mobile phase at flow rate 1 mL/min until the absorbance reached zero. Then, the bound fractions from the RCA-I, Con A and the WGA were eluted using the hapten sugars 0.1 M lactose, 0.1 M Me-α-D-Man and 0.1 M GlcNAc in the binding mobile phase, respectively. The lectin columns were again equilibrated with 30 column volumes of the binding mobile phase to prepare them for the next injection. The experiments were conducted at ambient temperature at a flow rate of 1.0 mL/min and the baseline was monitored at λ = 280 nm. The chromatograms were recorded using a digital recorder. The bound fractions from the lectin columns were subjected to dialysis using Spectra/Por dialysis bag from Spectrum Laboratories, Inc. (Houston, TX, USA) and the dialysis was done against water at 4 °C for 24 h.

1D SDS-PAGE Analysis

The SDS-PAGE analysis of the lectin bound fractions was performed on 10% Tris-glycine home case gels [21] (14 × 16 cm, 1.5 mm thickness) for 50 min at 200 V. After the completion of the electrophoresis, the gel was rinsed with water for 3 × 5 min and stained with Bio-safe Coomassie blue stain for 2h. Later the gel was washed with water to remove the background staining.

Protein Assay

The amount of protein that was captured by the lectin columns was determined using Bradford protein assay kit. The assay was done according to the manufacturer's protocol. Briefly, the standard or the sample were mixed with Coomassie protein assay reagent and incubated at room temperature for 10 min. The absorbances of the samples were then measured at a wavelength of 595 nm using a UV spectrophotometer.

LC-MS/MS Methodology

The conditions were same as in Chapter II.

LC-MS/MS Data Analysis

Mascot and X! Tandem were set up to search the SwissProt_082510 database (selected for Homo sapiens, 20359 entries) assuming the digestion enzyme trypsin. Other conditions were same as in Chapter II.

Scaffold (version Scaffold-3-00-07, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted only if they could be established at greater than 95% probability as specified by the Peptide Prophet algorithm. Protein identifications were accepted if they could be established at greater than 99% probability and contained at least 2 unique identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

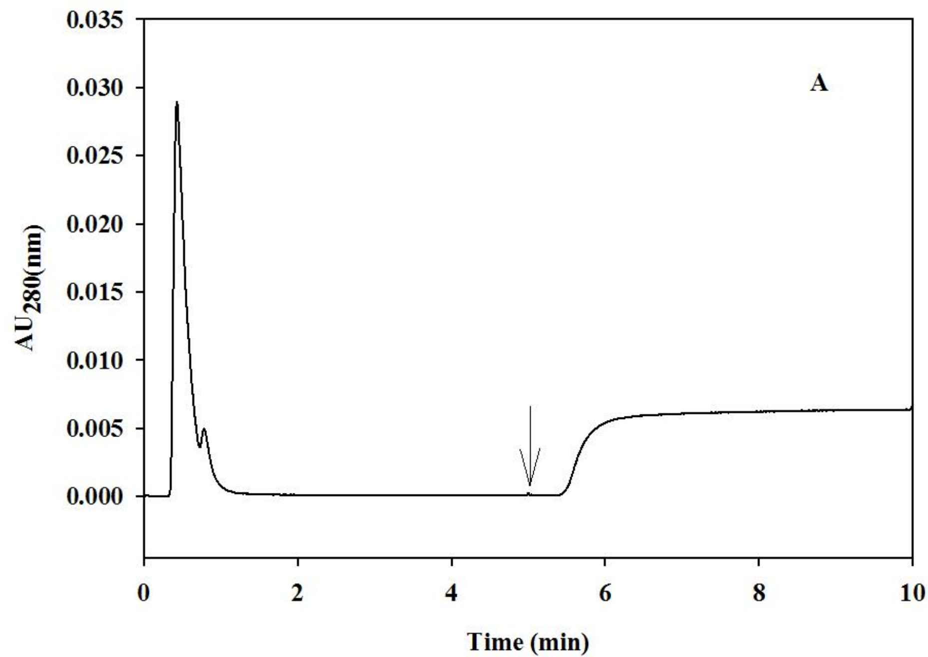
Results and Discussion

Evaluation of the Monolithic Lectin Columns

The lectins under investigation were immobilized onto a monolithic support according to the procedure discussed earlier (see Experimental). The monolithic support was prepared from a pre-polymerization solution containing 30 wt% monomers to 70 wt% porogens and 1 wt% initiator with respect to the monomers. This composition was chosen as it has been shown that this composition yields sufficiently moderate specific surface area, which can afford moderate and sufficient retention towards large bio-macromolecules such as glycoproteins [20].

The lectin columns thus prepared were evaluated for their specific and non-specific binding toward proteins. To verify that the lectin columns specifically capture glycoproteins, they were tested with standard glycoproteins and non-glycoproteins. The WGA column did not capture any of the injected myoglobin, which is a non-glycoprotein (Fig. 4A) but the column completely retained the α-1-acid glycoprotein that was eluted

only in the presence of the eluting haptenic sugar (Fig. 4B). Also, myoglobin was not at all retained by the Con A column (Fig. 5A), but when excess peroxidase (a glycoprotein) was injected onto the column, the excess was seen in the pass through fraction and the remaining retained peroxidase was eluted only in the presence of the haptenic sugar (Fig 5B). The RCA-I column also showed similar behaviors. When myoglobin was injected into the lectin column everything eluted in the pass through fraction (Fig. 6A) whereas when transferrin (a glycoprotein) was injected the excess amount was seen in the pass through fraction but the remaining eluted in presence of the eluting mobile phase (Fig. 6B). Therefore, the lectin columns did not have any non-specific binding due to the support with the standard proteins tested.



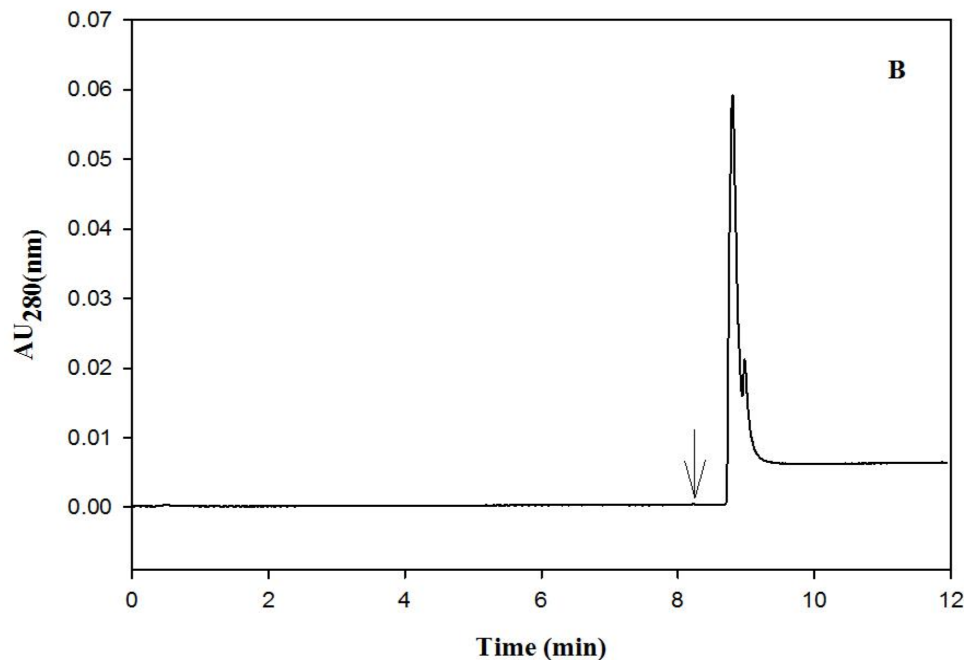
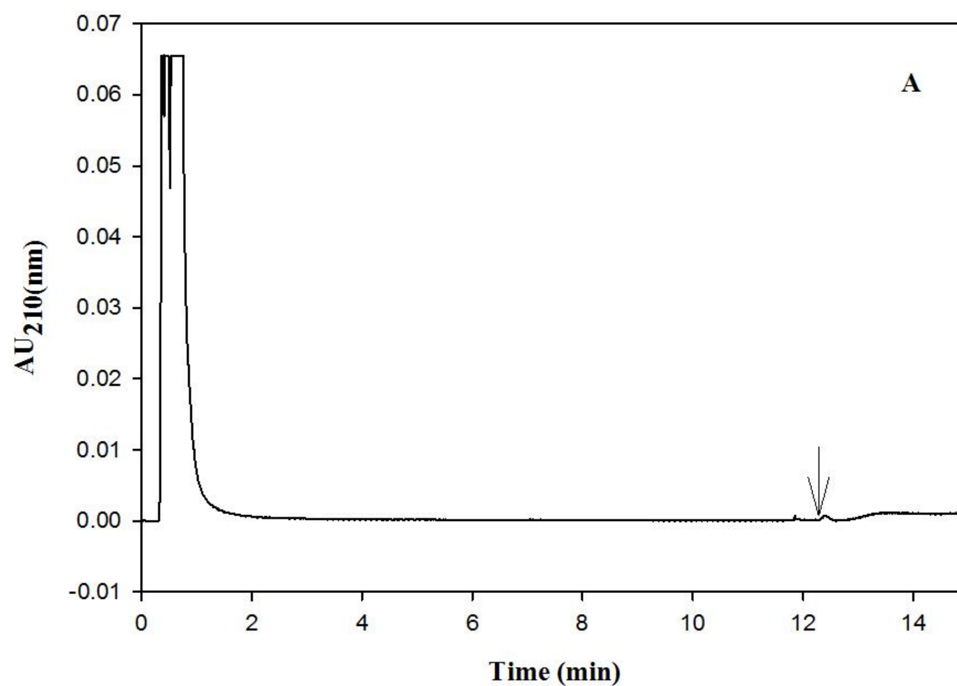


Figure 4. *Chromatograms of myoglobin (A) and α-1-acid glycoprotein (B) injected into the WGA column (50 mm × 4.6 mm ID). Binding mobile phase, 20 mM Tris containing 100 mM NaCl, 1mM of Mn^{2+}, Mg^{2+} and Ca^{2+}, pH = 6.0; eluting mobile phase, 0.1 M GlcNAc in the binding mobile phase; flow rate, 1 mL/min; wavelength, 280 nm. The, arrow indicates the change to eluting phase.*



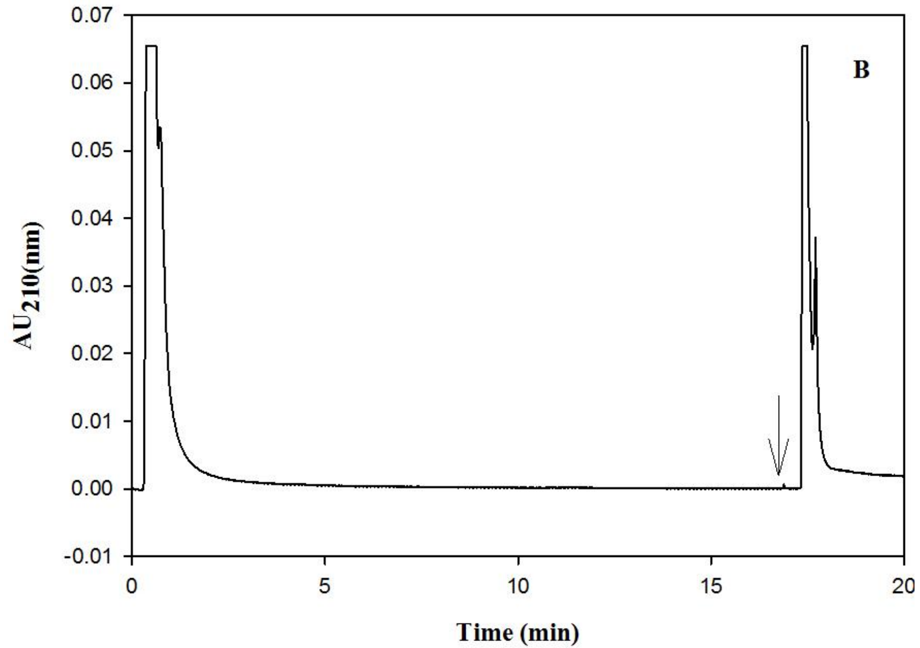
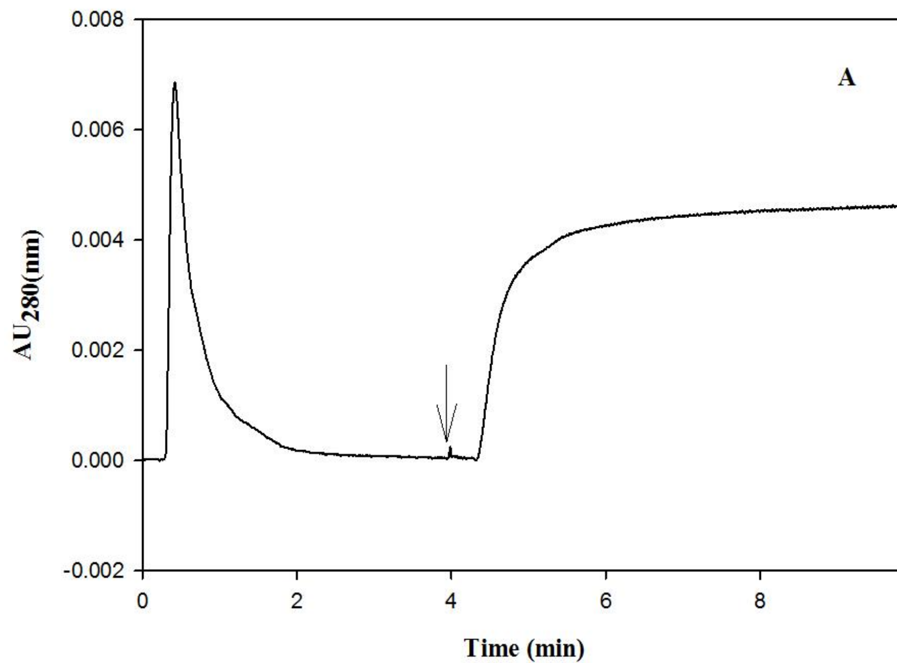


Figure 5. *Chromatograms of myoglobin (A) and peroxidase (B) injected into the Con A column (50 mm × 4.6 mm ID). Binding mobile phase, 20 mM Tris containing 100 mM NaCl, 1mM of Mn^{2+}, Mg^{2+} and Ca^{2+}, pH = 6.0; eluting mobile phase, 0.1 M Me-α-D-Man in the binding mobile phase; flow rate, 1 mL/min; wavelength, 210 nm. The arrow indicates the change to eluting phase.*



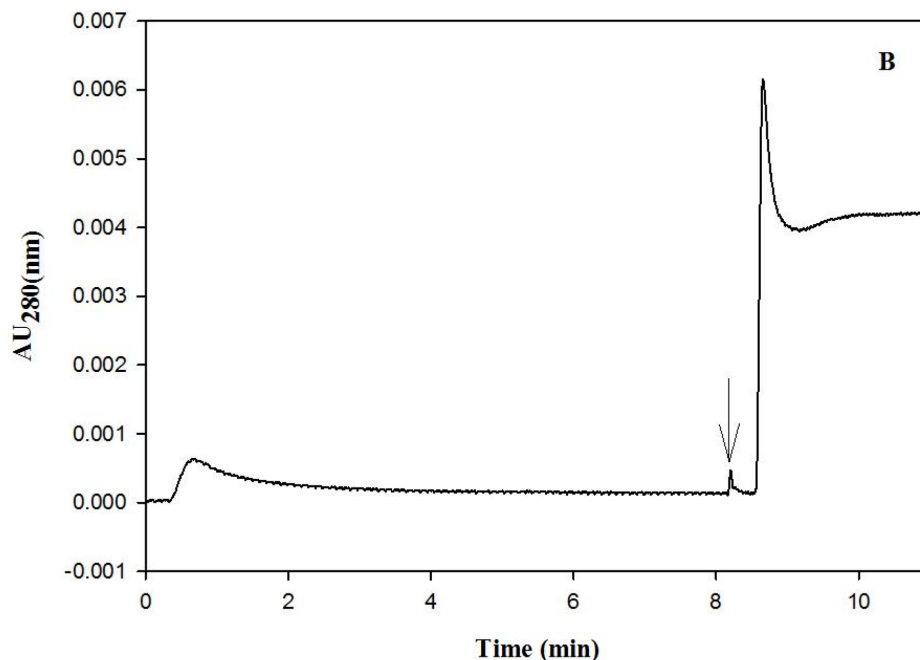


Figure 6. *Chromatograms of myoglobin (A) and transferrin (B) injected into the RCA-I column (50 mm × 4.6 mm ID). Binding mobile phase, 20 mM Tris containing 100 mM NaCl, 1mM of Mn^{2+}, Mg^{2+} and Ca^{2+}, pH = 6.0; eluting mobile phase, 0.1 M lactose in the binding mobile phase; flow rate, 1 mL/min; wavelength, 280 nm. The arrow indicates the change to eluting phase.*

The reproducibility of the immobilization process of the lectin was evaluated by preparing two sets of monolithic columns for each lectin. Each set of lectin columns thus obtained was tested under similar conditions using standard glycoproteins. It was found that the peak areas were reproducible from column-to-column. For instance, the WGA column yielded a peak area of 0.25 Vsec for the first column and 0.26 Vsec for the second column. The same was observed in the case of the Con A column with the peak areas for first and second column being 5.55 and 5.32 Vsec, respectively. The RCA-I

column also showed similar behavior with the peak areas being 0.23 and 0.24 Vsec for the first and second column, respectively.

In order to test their loading capacity, different amounts of human serum were injected onto the lectin columns. For instance, the elution profiles of the WGA column when 50 μL and 250 μL of diluted serum (1:3 ratio) were injected into the lectin column are shown in Figs. 7 and 8. The areas under the peaks were 0.43 Vsec and 2.38 Vsec for 50 and 250 μL diluted serum injected, respectively, showing an increase in peak area by a factor of 5.5. Also, the height of the peak increased by a factor of 4.3 from ~ 0.015 AU for 50 μL to ~ 0.065 AU for 250 μL of the diluted serum. Clearly, the column was able

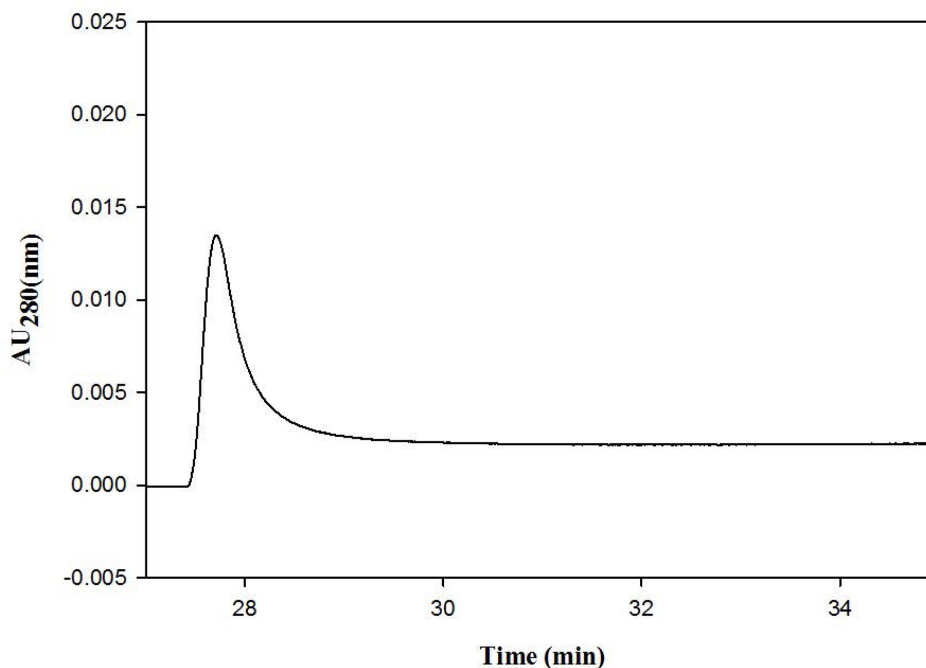


Figure 7. *Elution profile of the bound proteins from the WGA column (100 mm × 4.6 ID) when 50 μL of diluted serum (1:3 ratio) was injected. Conditions are the same as in Fig. 4.*

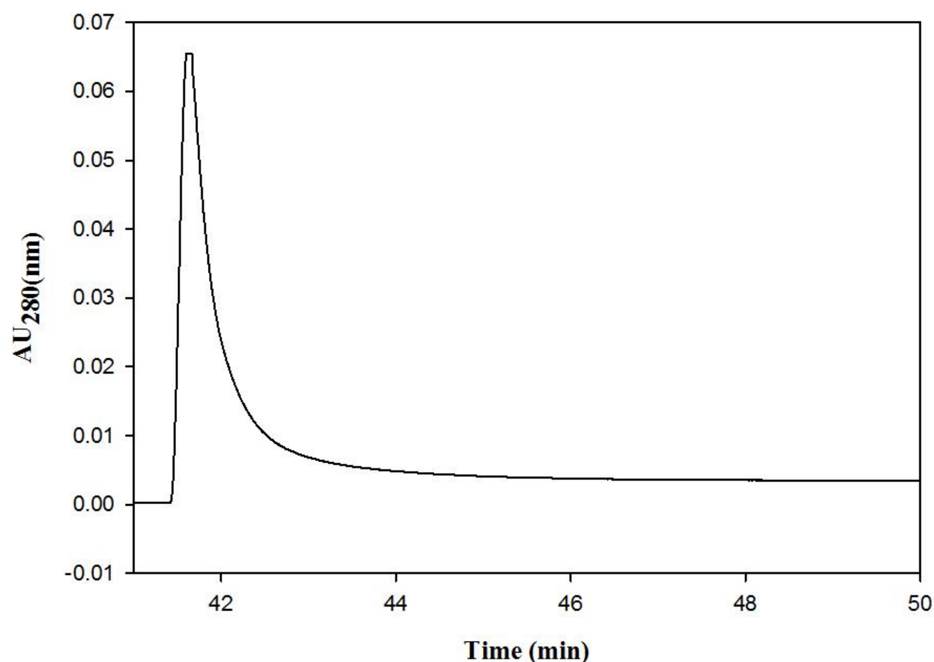


Figure 8. *Elution profile of the bound proteins from the WGA column (100 mm × 4.6 ID) when 250 μL of diluted serum (1:3 ratio) was injected. Conditions are the same as in Fig. 4.*

to capture a commensurate amount of serum proteins when the amount of serum injected onto the lectin column was increased by 5 fold from 50 µL to 250 µL. Despite the fact that the loading capacity of the lectin columns was quite high only 50 µL serum were injected in further experiments in order to keep the column operating in the loading range, and to avoid a false identification of the differentially expressed proteins in the cancer serum with respect to disease-free serum.

Effect of the Serial Order of the Lectin Columns in the Tandem Format

To investigate the effects of the order in which the lectin columns are arranged in the tandem series on the number of captured proteins by each lectin column, three different arrangements were considered including WGA → Con A → RCA-I (denoted as

WCR hereafter), Con A → WGA → RCA-I (CWR) and RCA-I → Con A → WGA (RCW). Fifty microliter of the 1:3 diluted human serum were injected consecutively twice onto the tandem lectin columns and the bound fractions were eluted, desalted, concentrated and analyzed by LC-MS/MS to identify the proteins captured by each lectin column in the three different series. The chromatographic setup is shown in Fig. 9.

A Venn diagram, which indicates the numbers of unique and common proteins that were captured by the three series, is shown in Fig. 10. A comparison of number of proteins that were captured by each lectin column in the three different column series is listed in Table 1. The WCR series has 9 proteins that were not identified in the other two series and also this series has the maximum number of identified non-redundant proteins. The proteins that were unique to each series are listed in Tables 2, 3 and 4. There were 60 non-redundant proteins that were common to all the series, and these are listed in Table 5.

TABLE 1

COMPARISON OF NUMBER OF PROTEINS THAT WERE IDENTIFIED ON DIFFERENTLY ORDERED SERIAL LECTIN COLUMNS

Series	WGA	Con A	RCA-I	Total proteins	# non-redundant proteins	# of unique proteins
WCR	70	40	30	140	80	9
CWR	52	60	31	143	73	6
RCW	5	22	73	100	74	4

TABLE 2

LIST OF PROTEINS THAT WERE UNIQUE TO WGA-CON A-RCA-I SERIES

Identified Proteins	Accession Number	Molecular Weight
Apolipoprotein D	APOD_HUMAN	21 kDa
Carboxypeptidase N subunit 2	CPN2_HUMAN	61 kDa
Clusterin	CLUS_HUMAN	52 kDa
Complement C1q subcomponent subunit C	C1QC_HUMAN	26 kDa
Complement component C9	CO9_HUMAN	63 kDa
Fibronectin	FINC_HUMAN	263 kDa
Ficolin-3	FCN3_HUMAN	33 kDa
Kallistatin	KAIN_HUMAN	49 kDa
Serum paraoxonase/arylesterase 1	PON1_HUMAN	40 kDa

TABLE 3

LIST OF PROTEINS THAT WERE UNIQUE TO CON A – WGA - RCA-I SERIES

Identified Proteins	Accession Number	Molecular Weight
Ig heavy chain V-III region BRO *	HV305_HUMAN	13kDa
Ig kappa chain V-III region POM *	KV306_HUMAN (+1)	12kDa
Keratin, type I cytoskeletal 14*	K1C14_HUMAN (+1)	52 kDa
Keratin, type II cytoskeletal 6B*	K2C6B_HUMAN	60 kDa
L-selectin	LYAM1_HUMAN	42 kDa
Semenogelin-1*	SEMG1_HUMAN	52 kDa

* Non-glycoproteins

TABLE 4

LIST OF PROTEINS THAT WERE UNIQUE TO RCA-I - CON A - WGA SERIES

Identified Proteins	Accession Number	Molecular Weight
Heparin cofactor 2	HEP2_HUMAN	57 kDa
Ig heavy chain V-III region GAL *	HV320_HUMAN	13 kDa
Xin actin-binding repeat-containing protein 2*	XIRP2_HUMAN	382 kDa
Ig lambda chain V-I region WAH *	LV106_HUMAN	12 kDa

*Non-glycoproteins

TABLE 5

LIST OF PROTEINS THAT WERE COMMON TO ALL THE THREE SERIES

Identified Proteins	Accession Number	Molecular Weight
Afamin	AFAM_HUMAN	69 kDa
Alpha-1-acid glycoprotein 1	A1AG1_HUMAN	24 kDa
Alpha-1-acid glycoprotein 2	A1AG2_HUMAN	24 kDa
Alpha-1-antichymotrypsin	AACT_HUMAN	48 kDa
Alpha-1-antitrypsin	A1AT_HUMAN	47 kDa
Alpha-1B-glycoprotein	A1BG_HUMAN	54 kDa
Alpha-2-HS-glycoprotein	FETUA_HUMAN	39 kDa
Alpha-2-macroglobulin	A2MG_HUMAN	163 kDa
Angiotensinogen	ANGT_HUMAN	53 kDa
Antithrombin-III	ANT3_HUMAN	53 kDa
Apolipoprotein A-I	APOA1_HUMAN	31 kDa
Apolipoprotein A-II*	APOA2_HUMAN	11 kDa
Attractin	ATRN_HUMAN	159 kDa
Beta-2-glycoprotein 1	APOH_HUMAN	38 kDa
CD5 antigen-like*	CD5L_HUMAN	38 kDa
Ceruloplasmin	CERU_HUMAN	122 kDa
Complement C3	CO3_HUMAN	187 kDa
Complement C4-B	CO4B_HUMAN	193 kDa
Complement C5	CO5_HUMAN	188 kDa
Complement factor B	CFAB_HUMAN	86 kDa
Complement factor H	CFAH_HUMAN	139 kDa
Complement factor I	CFAI_HUMAN	66 kDa
Corticosteroid-binding globulin	CBG_HUMAN	45 kDa
Haptoglobin	HPT_HUMAN	45 kDa
Hemoglobin subunit alpha	HBA_HUMAN	15 kDa
Hemoglobin subunit beta	HBB_HUMAN	16 kDa
Hemopexin	HEMO_HUMAN	52 kDa
Histidine-rich glycoprotein	HRG_HUMAN	60 kDa
Ig alpha-1 chain C region	IGHA1_HUMAN	38 kDa
Ig alpha-2 chain C region	IGHA2_HUMAN	37 kDa
Ig gamma-1 chain C region	IGHG1_HUMAN	36 kDa
Ig gamma-2 chain C region	IGHG2_HUMAN	36 kDa
Ig gamma-3 chain C region	IGHG3_HUMAN	41 kDa
Ig gamma-4 chain C region	IGHG4_HUMAN	36 kDa
Ig kappa chain C region*	IGKC_HUMAN	12 kDa
Ig kappa chain V-II region RPMI 6410*	KV206_HUMAN	15 kDa
Ig kappa chain V-III region WOL*	KV305_HUMAN (+2)	12kDa
Ig lambda-2 chain C regions*	LAC2_HUMAN	11 kDa
Ig lambda chain V-III region LOI*	LV302_HUMAN	12 kDa

Ig lambda chain V-III region SH *	LV301_HUMAN	11 kDa
Ig mu chain C region	IGHM_HUMAN	49 kDa
Immunoglobulin J chain	IGJ_HUMAN	18 kDa
Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1_HUMAN	101 kDa
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2_HUMAN	106 kDa
Inter-alpha-trypsin inhibitor heavy chain H4	ITIH4_HUMAN	103 kDa
Keratin, type I cytoskeletal 10*	K1C10_HUMAN	59 kDa
Keratin, type I cytoskeletal 9*	K1C9_HUMAN	62 kDa
Keratin, type II cytoskeletal 1*	K2C1_HUMAN	66 kDa
Keratin, type II cytoskeletal 2 epidermal*	K22E_HUMAN	65 kDa
Kininogen-1	KNG1_HUMAN	72 kDa
Leucine-rich alpha-2-glycoprotein	A2GL_HUMAN	38 kDa
N-acetylmuramoyl-L-alanine amidase	PGRP2_HUMAN	62 kDa
Plasma protease C1 inhibitor	IC1_HUMAN	55 kDa
Protein AMBP	AMBP_HUMAN	39 kDa
Prothrombin	THRB_HUMAN	70 kDa
Serotransferrin	TRFE_HUMAN	77 kDa
Serum albumin*	ALBU_HUMAN	69 kDa
Thyroxine-binding globulin	THBG_HUMAN	46 kDa
Transthyretin	TTHY_HUMAN	16 kDa
Zinc-alpha-2-glycoprotein	ZA2G_HUMAN	34 kDa

*Non-glycoproteins

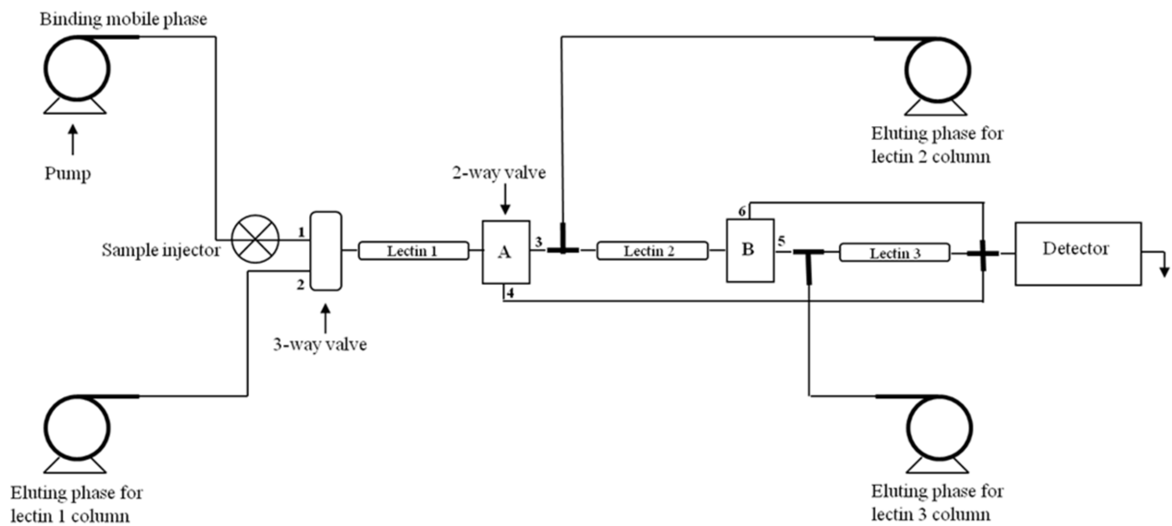


Figure 9. *Chromatographic setup for the concentration of the serum proteins using tandem lectin columns. When the diluted serum was injecte,d the 3-way valve was in 1-position. The two-way valves (A and B) are in positions 3 and 5, so that the binding*

mobile phase (20 mM Tris-HCl (pH 6.0) containing 100 mM NaCl, 1 mM $MgCl_2$, 1mM $MnCl_2$, 1 mM $CaCl_2$) flows through the three lectin columns. Under this condition the unbound proteins are washed away. The Lectin 1 column was eluted when the three-way valve was in position 2 and the two-way valve (A) is in position 4, where it by-passes the other two columns and directly reaches the detector. After complete elution the position of two-way valve (A) was again changed to 3. To elute the lectin 2 column, the two-way valve was in position 6, so that the bound fraction reaches the detector by-passing lectin 3 column. The lectin 3 column was eluted when the two-way valve was in position 5 and the bound fraction was collected from the detector.

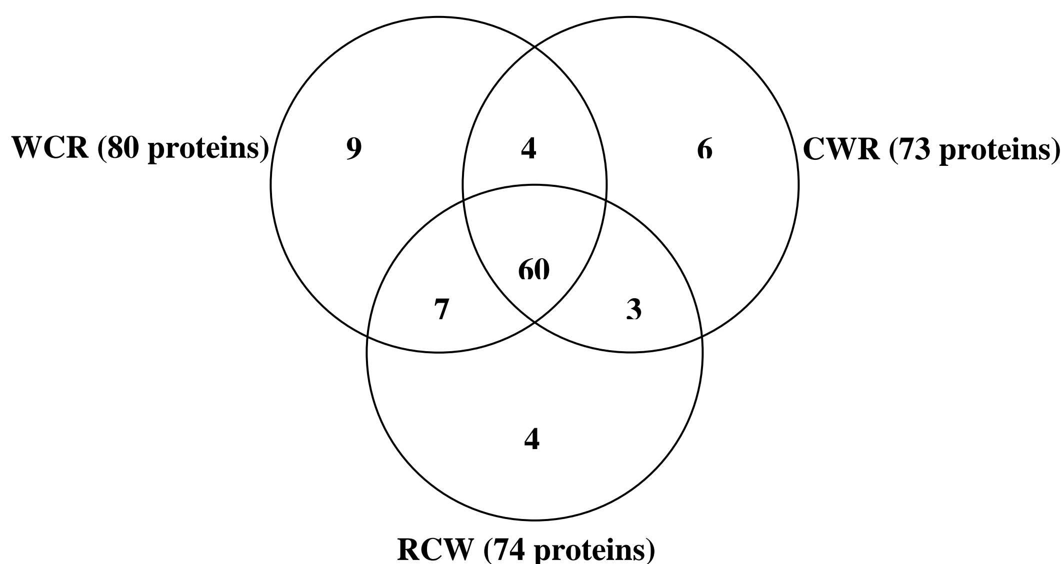


Figure 10. *Venn diagram showing the number of proteins identified in each series. The total numbers of non-redundant proteins for each series are written in parentheses.*

The number of proteins captured by the last lectin column (which is the RCA-I column) in the WCR and CWR series were 30 and 31 proteins, respectively, whereas only 5 proteins were identified in the last column (which is the WGA column) in the

RCW series. This is due to the fact that the RCA-I column offers a wider sugar recognition than the other two lectin columns (see Figs. 1 – 3) in the sense that any glycoprotein that has a higher amount of complex type tri- and tetrantennary glycoforms will pass through the WGA and the Con A columns and be preferentially retained by the RCA-I column that was placed at the end of the series. In the WCR and CWR arrangements, the RCA-I column would have more binding sites available for highly branched glycans. In the case of the RCW series, the RCA-I column, which was placed first in the series, will offer all of its binding sites to the complex type biantennary structures and to hybrid structures in addition to the tri- and tetrantennary glycans, and consequently captured the majority of proteins. In this serial set up, only the glycoproteins that carry the sugar determinants that are not recognized by RCA-I are able to pass through and be captured either by the Con A or the WGA that follows the RCA-I column. This might be one of the possible reasons why RCW series had the lowest number of total proteins.

Based on the above results, the WCR tandem arrangement was found to be better than the other two series. In fact, by placing the WGA column first, the glycoproteins with bisected complex and hybrid type glycans will bind to the WGA thus freeing the sites of the Con A column for the non-bisected hybrid type as well as for the high mannose type. In addition, by placing the WGA column first followed by the Con A column, the binding sites of the RCA-I column would be saved in major part for the binding of tri- and tetraantennary glycans.

Furthermore, the 9 unique proteins that were captured by the WCR series were all glycoproteins, whereas in the CWR series, out of the 6 unique proteins, only one was

found to be a glycoprotein, and in the RCW series, only 1 out of 4 was a glycoprotein. In the 9 glycoproteins that were captured by the WCR series, some were reported to be clinically important. For example, the glycoprotein kallistatin has been reported to be up-regulated in type 1 diabetic patients with microvascular complications and with hypertension [22]. On the other hand, the proteins that were unique to CWR series were mostly immunoglobulin chains and keratin and had two low abundance proteins namely, L-selectin and semenogelin-1, where only L-selectin was a glycoprotein. In summary, the WCR series gave better results and this series was used in further work.

LC-MS/MS Analysis of the Lectin Enriched Fractions from Cancer and Disease-free Sera

The WCR series was investigated for its usefulness in capturing specific glycoproteins from both disease-free and breast cancer serum for their subsequent identification by LC-MS/MS. The advantage of using broad specificity lectins is that they should in principle allow the enrichment/capturing of glycoproteins with the widest aberrant glycosylation possible in a given cancerous serum sample. This should allow the determination of the differentially expressed glycoproteins in cancer serum with respect to disease-free serum. In this regard, cancer serum (n = 1) and disease-free serum (n = 3) were enriched by the tandem WCR columns. As mentioned in the Experimental, only proteins that exhibited identification probability greater than 99% with peptide identification probability greater than 95% containing at least two unique peptides were considered, and they are reported in Table 6. The total number of non-redundant proteins that were identified from the disease-free serum using the three tandem lectin columns

TABLE 6

LIST OF PROTEINS IDENTIFIED IN THE LECTIN-BOUND FRACTIONS

#	Identified Proteins	Accession Number	Mol Wt.	Disease-free serum			Breast cancer serum		
				WGA	Con A	RCA-I	WGA	Con A	RCA-I
				Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]
1	Afamin	AFAM_HUMAN	69 kDa	8/5	C		6/4	11/7	
2	Alpha-1-acid glycoprotein 1	A1AG1_HUMAN	24 kDa	66/12		19/8	62/12		9/4
3	Alpha-1-acid glycoprotein 2	A1AG2_HUMAN	24kDa	31/5		11/3	30/5		8/3
4	Alpha-1-antichymotrypsin	AACT_HUMAN	48 kDa	40/15			39/16		
5	Alpha-1-antitrypsin	A1AT_HUMAN	47 kDa	50/19	61/20	54/21	52/21	70/28	54/22
6	Alpha-1B-glycoprotein	A1BG_HUMAN	54 kDa	9/6	31/14	12/7	10/6	33/14	6/3
7	Alpha-2-antiplasmin	A2AP_HUMAN	55 kDa	4/2			W		
8	Alpha-2-HS-glycoprotein	FETUA_HUMAN	39 kDa	25/12			22/11		
9	Alpha-2-macroglobulin	A2MG_HUMAN	163 kDa	234/75	2/1	6/4	291/85	2/2	10/6
10	Angiotensinogen	ANGT_HUMAN	53 kDa	12/7	5/3	3/3	4/3	2/2	R

11	Antithrombin-III	ANT3_HUMAN	53 kDa		9/4	1/1		11/7	2/1
12	Apolipoprotein A-I	APOA1_HUMAN	31 kDa	18/10		15/9	11/6		2/1
13	Apolipoprotein A-II*	APOA2_HUMAN	11 kDa	8/5		1/1	4/3		R
14	Apolipoprotein B-100	APOB_HUMAN	516 kDa	66/39			35/22		
15	Apolipoprotein D	APOD_HUMAN	21 kDa	5/3			7/5		
16	Attractin	ATRN_HUMAN	159 kDa	8/4			5/3		
17	Beta-2-glycoprotein 1	APOH_HUMAN	38 kDa	24/12	4/3		7/4	11/5	
18	Carboxypeptidase N subunit 2	CPN2_HUMAN	61 kDa	4/2					
19	CD5 antigen-like*	CD5L_HUMAN	38 kDa	20/11			7/4		
20	Ceruloplasmin	CERU_HUMAN	122 kDa	87/36	7/3	21/11	69/31	14/8	25/13
21	Complement C2	CO2_HUMAN	83 kDa	2/1			4/2		
22	Complement C3	CO3_HUMAN	187 kDa		48/26			41/24	
23	Complement C4-B	CO4B_HUMAN	193 kDa	13/8	3/2	2/2	15/8	4/2	4/2
24	Complement C5	CO5_HUMAN	188 kDa			3/2			R
25	Complement factor B	CFAB_HUMAN	86 kDa	6/3	7/5	25/16	2/2	3/3	R

26	Complement factor H	CFAH_HUMAN	139 kDa	52/28			48/27		
27	Complement factor I	CFAI_HUMAN	66 kDa	13/8			6/4		
28	Corticosteroid-binding globulin	CBG_HUMAN	45 kDa	7/5			W		
29	Fibrinogen alpha chain	FIBA_HUMAN	95 kDa	W			2/1		
30	Fibrinogen beta chain	FIBB_HUMAN	56 kDa	W			1/1		
31	Fibrinogen gamma chain	FIBG_HUMAN	52 kDa	W			3/1		
32	Haptoglobin	HPT_HUMAN	45 kDa	130/31	C	11/6	144/27	8/6	16/10
33	Haptoglobin-related protein*	HPTR_HUMAN	39 kDa	50/4			45/3		
34	Hemoglobin subunit alpha	HBA_HUMAN	15 kDa	11/4	C		37/11	7/3	
35	Hemoglobin subunit beta	HBB_HUMAN	16 kDa	13/8	C	R	59/17	3/2	3/2
36	Hemoglobin subunit delta*	HBD_HUMAN	16 kDa	W			33/4		
37	Hemopexin	HEMO_HUMAN	52 kDa	72/17	C	6/5	69/20	2/1	24/12
38	Heparin cofactor 2	HEP2_HUMAN	57 kDa	3/2			W		
39	Histidine-rich glycoprotein	HRG_HUMAN	60 kDa	W			4/3		
40	Ig alpha-1 chain C region	IGHA1_HUMAN	38 kDa	56/14	C	5/4	50/15	4/3	5/3

41	Ig alpha-2 chain C region	IGHA2_HUMAN	37 kDa	42/3			W		
42	Ig gamma-1 chain C region	IGHG1_HUMAN	36 kDa	10/6	15/7	43/16	5/4	22/9	45/15
43	Ig gamma-2 chain C region	IGHG2_HUMAN	36 kDa		C	32/3		11/2	27/2
44	Ig gamma-3 chain C region	IGHG3_HUMAN	41 kDa		3/1	38/6		16/4	35/6
45	Ig gamma-4 chain C region	IGHG4_HUMAN	36 kDa		3/1	30/6		10/4	25/5
46	Ig heavy chain V-III region BRO*	HV305_HUMAN	13 kDa	5/2			W		
47	Ig heavy chain V-III region GAL*	HV320_HUMAN	13 kDa	4/2			1/1		
48	Ig heavy chain V-III region VH26*	HV303_HUMAN	13 kDa	5/2			W		
49	Ig kappa chain C region*	IGKC_HUMAN	12 kDa	33/8	4/3	21/7	13/5	2/1	14/4
50	Ig kappa chain V-I region AG*	KV101_HUMAN	12 kDa			2/1			R
51	Ig kappa chain V-I region EU*	KV106_HUMAN	12 kDa	4/2			W		
52	Ig kappa chain V-III region HAH*	KV312_HUMAN	14 kDa	5/3		6/4	W		2/2
53	Ig kappa chain V-III region POM*	KV306_HUMAN (+1)	12 kDa	1/1		1/1	1/1		R
54	Ig lambda chain V-I region WAH*	LV106_HUMAN	12 kDa	3/2			W		
55	Ig lambda chain V-III region LOI*	LV302_HUMAN	12 kDa	6/3			6/3		

56	Ig lambda-1 chain C regions*	LAC1_HUMAN	11 kDa	16/2		8/2	9/1		2/1
57	Ig lambda-2 chain C regions*	LAC2_HUMAN	11 kDa	24/6	8/3	16/6	21/5	5/3	12/4
58	Ig mu chain C region	IGHM_HUMAN	49 kDa	82/19			40/15		
59	Immunoglobulin J chain	IGJ_HUMAN	18 kDa	7/4			8/4		
60	Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1_HUMAN	101 kDa	34/17			13/8		
61	Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2_HUMAN	106 kDa	51/23			26/16		
62	Inter-alpha-trypsin inhibitor heavy chain H4	ITIH4_HUMAN	103 kDa	36/19			24/15		
63	Keratin, type I cytoskeletal 10*	K1C10_HUMAN	59 kDa	W		56/26	5/4		6/4
64	Keratin, type I cytoskeletal 14*	K1C14_HUMAN	52 kDa			12/4			R
65	Keratin, type I cytoskeletal 9*	K1C9_HUMAN	62 kDa			35/17			3/2
66	Keratin, type II cytoskeletal 1*	K2C1_HUMAN	66 kDa	W		62/27	5/3		8/5
67	Keratin, type II cytoskeletal 2 epidermal*	K22E_HUMAN	65 kDa			53/25			5/2
68	Keratin, type II cytoskeletal 5*	K2C5_HUMAN	62 kDa			18/6			R
69	Kininogen-1	KNG1_HUMAN	72 kDa	19/13			13/9		
70	Leucine-rich alpha-2-glycoprotein	A2GL_HUMAN	38 kDa	27/13			19/11		

71	N-acetylmuramoyl-L-alanine amidase	PGRP2_HUMAN	62 kDa	7/3			W		
72	Plasma kallikrein	KLKB1_HUMAN	71 kDa	3/2			4/2		
73	Plasma protease C1 inhibitor	IC1_HUMAN	55 kDa	2/1			5/4		
74	Pregnancy zone protein	PZP_HUMAN	164 kDa	56/10			W		
75	Protein AMBP	AMBP_HUMAN	39 kDa	11/8	C		3/2	1/1	
76	Serotransferrin	TRFE_HUMAN	77 kDa	7/5	168/52	147/58	3/2	128/44	166/57
77	Serum albumin*	ALBU_HUMAN	69 kDa	43/24	C	4/3	37/22	6/4	13/7
78	Serum amyloid P-component	SAMP_HUMAN	25 kDa	3/2		R	4/2		3/2
79	Thyroxine-binding globulin	THBG_HUMAN	46 kDa	W	C		2/1	1/1	
80	Vitronectin	VTNC_HUMAN	54 kDa	3/2			W		
81	Zinc-alpha-2-glycoprotein	ZA2G_HUMAN	34 kDa		29/14	10/7		23/12	20/10

[a] Average of spectral counts in triplicate runs
[b] Average of number of unique peptides from triplicate runs
* Non-glycoproteins

W, C and R represents that the protein was only present in either disease-free or cancer serum for the WGA, Con A and RCA-I columns respectively

(WCR) was 75 proteins. The WGA column captured 61 proteins whereas the Con A and the RCA-I columns captured 17 and 35 proteins, respectively, which totaled 113 proteins. The percentages of glycoproteins that were captured by each lectin column were 72%, 88% and 60% for the WGA, Con A and RCA-I columns, respectively, corresponding to 28%, 12% and 40% non-glycosylated proteins, respectively. There were 11 common proteins among all the three lectin columns. A Venn diagram listing the number of identified proteins on the three lectin columns is shown in Fig. 11. The number of proteins captured by the WGA, Con A and the RCA-I columns from the cancer serum was 56, 27 and 29, respectively, totaling 112 proteins. A Venn diagram indicating the number of common proteins and unique proteins is shown in Fig. 12. The percentages of glycoproteins that were captured by each lectin column were 77%, 89% and 69%, for the WGA, Con A and RCA-I columns, respectively, corresponding to 27%, 11% and 31% non-glycosylated proteins, respectively. The total number of non-redundant proteins identified from three lectin columns using the cancer serum was 65 proteins. The number of common proteins among the three lectin columns was 14. Although, some proteins were found to be common among all the three lectins, it should be noted that the proteins did not overflow due to overloading from one column to another as all the three columns were operated belowf their capacities. This overlap between the lectin columns might be due to the presence of different glycoforms present in a single protein that has affinity towards all the three lectins. For example, afamin was detected in both disease-free and cancer serum by the WGA column but it was over-expressed only in the cancer serum fraction from the Con A column which was placed right after the WGA column in the tandem column format.

As mentioned above the percentage of glycoproteins captured in each lectin column is in the range of 69-89%. This non-specific binding of the lectin column may possibly be due to the glycoprotein and non-glycoprotein interactions. For example, transferrin which is one of the high abundant glycoproteins can interact with some of the other proteins such as apolipoprotein-I, fibrinonectin, immunoglobulin kappa light chain, transthyretin and albumin. This non-specific binding to lectin columns has been also reported with other solid supports. For example, agarose bound lectin column also has been reported to capture non-glycoproteins [23], and another report which used POROS 20-AL support resulted in only 75% of glycoprotein capture [12]. In the current study, the major portion of the non-glycoproteins consisted of some immunoglobulin chains and keratin, which were also observed in some of the other reports that involved lectin capturing of glycoproteins from blood plasma and serum [14, 23].

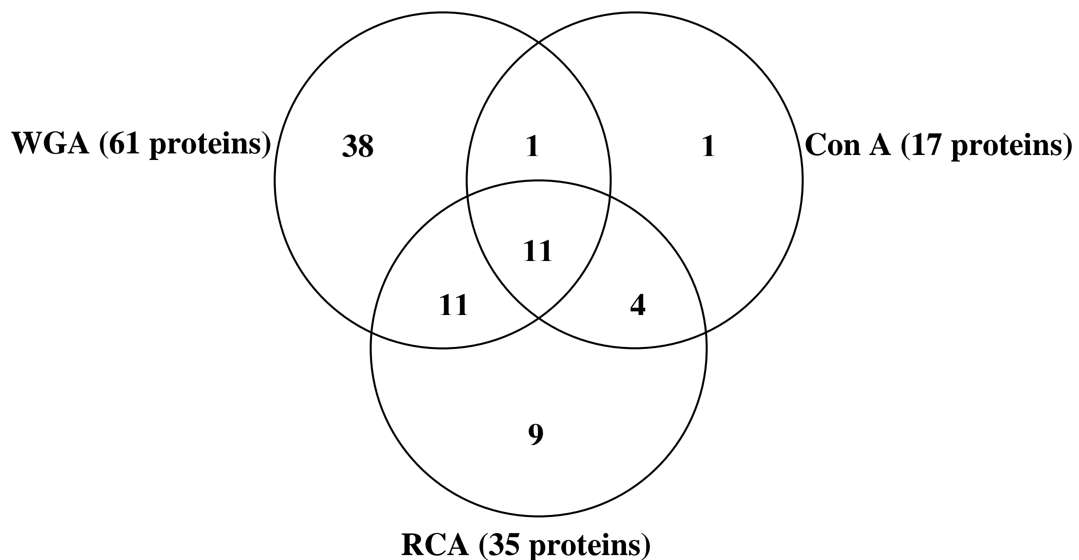


Figure 11. *Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the disease-free serum.*

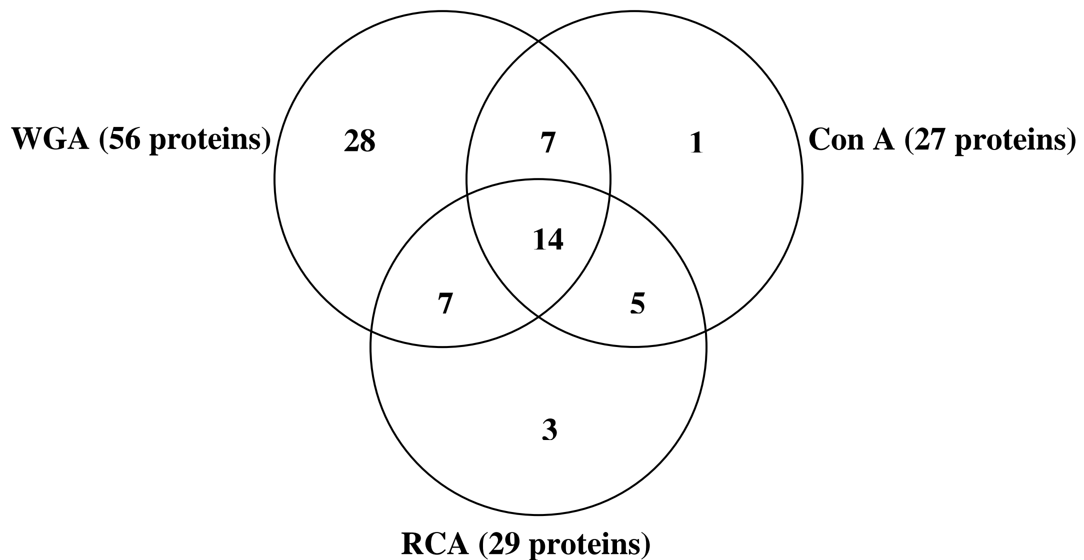


Figure 12. *Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the breast cancer serum.*

There were 6 unique proteins in the cancer serum including fibrinogen-α–chain, fibrinogen-β-chain, fibrinogen-γ-chain, hemoglobin subunit delta, histidine-rich glycoprotein and thyroxine-binding globulin. An elevated level of fibrinogen was found in 44% of the early stage and 22 % of the advanced stage breast cancer patients [24]. It should be noted that the spectrum count of the fibrinogen was less than 4 (see Table 6), a value that may not be significant. There were 16 unique proteins in disease-free serum but not in cancer serum. These unique proteins present in the disease-free and the cancer serum, indicate that certain proteins were differentially expressed in both sera, see Fig. 13. The number of identified proteins common to the disease-free and the cancer serum was 59 non-redundant proteins. As can be seen in Fig. 13, there were no differentially

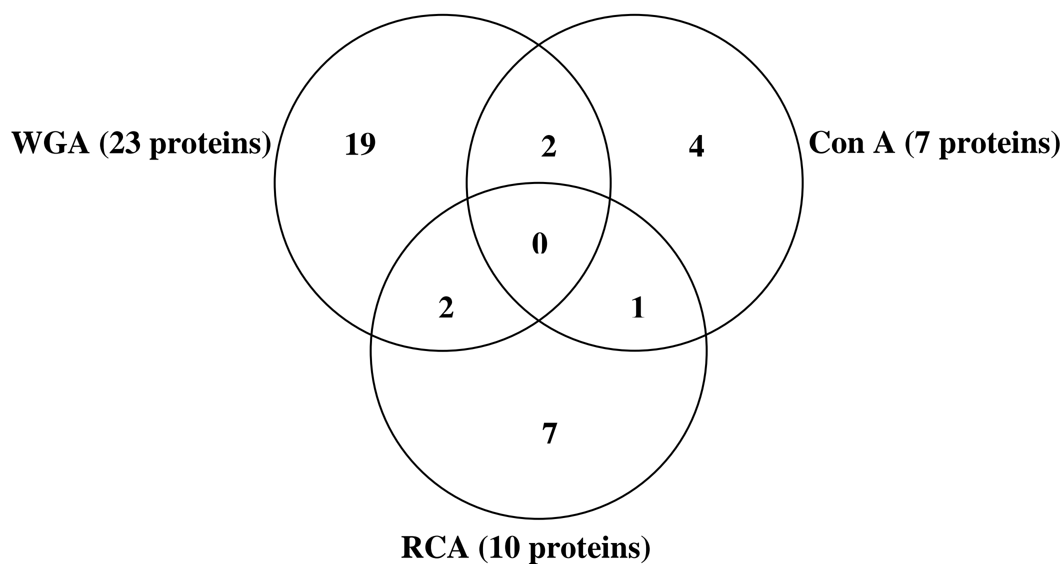


Figure 13. *Venn diagram for the differentially expressed proteins found in the lectin fractions.*

expressed proteins common to the three lectin columns. There were two common proteins between WGA and Con A, 2 common proteins between WGA and RCA-I and 1 common protein between WGA and RCA-I.

Specificity of the Lectin Columns

The specificity of the lectin columns can be illustrated with some captured glycoproteins. For instance, α-1-acid glycoprotein is one of the medium abundance proteins present in serum. It contains a higher amount of sialylated tri- and tetraantennary complex type compared to the biantennary complex type glycans. It is also known that Con A has very weak binding towards the bianntenary complex type glycans, whereas WGA has affinity towards the sialic acid residue and RCA-I has affinity

towards the tri- and the tetrantennary complex type glycan. Both α-1-acid glycoprotein 1 and α-1-acid glycoprotein 2 were captured only by the WGA and RCA-I columns in both disease-free and the cancer serum, but not captured by the Con A column. Clearly, the specificity of the lectins is well obeyed here. Another protein that merits discussion is complement C3. This was the only protein that was unique to the Con A column in both the disease-free and cancer serum. The α chain of the protein has the glycan compositions $(Man)_9(GlcNAc)_2$-Asn and $(Man)_8(GlcNAc)_2$-Asn while the β chain has $(Man)_5(GlcNAc)_2$-Asn and $(Man)_6(GlcNAc)_2$-Asn [25]. The high mannose content of complement C3 reported [25, 26], made it specific towards Con A column and was thus retained only by the Con A column and not the other two columns. It has to be noted that even though complement C3 is one of high abundance proteins, it was specifically captured by the Con A and not by WGA or RCA-I columns indicating the absence of any non-specific binding.

Visual Differential Expression of Proteins Using SDS-PAGE Analysis of the Cancer and the Disease-free Sera

Aliquots of equal volumes taken from each lectin column bound fraction of disease-free and cancer sera were further analyzed by SDS-PAGE and the results are shown in Fig. 14. By visually examining the SDS-PAGE gel, it can be seen that some proteins (marked in circles) were either over-expressed or down-expressed in cancer serum. By comparing the WGA lanes (lanes 1 and 2) and the Con A lanes (lanes 3 and 4) lanes, it can be seen that some gel bands (shown by arrows) were observed in only that particular lectin, indicating that each lectin captured glycoproteins that were only specific

to it (see Figs. 1 - 3 for details of the three lectin specificities). Also, some of the bands that are common to Con A and WGA indicate the overlapping in their affinity towards some of the glycoforms. Since RCA-I lanes (lanes 5 and 6) were too faint they were not considered for comparison with the other two lectins.

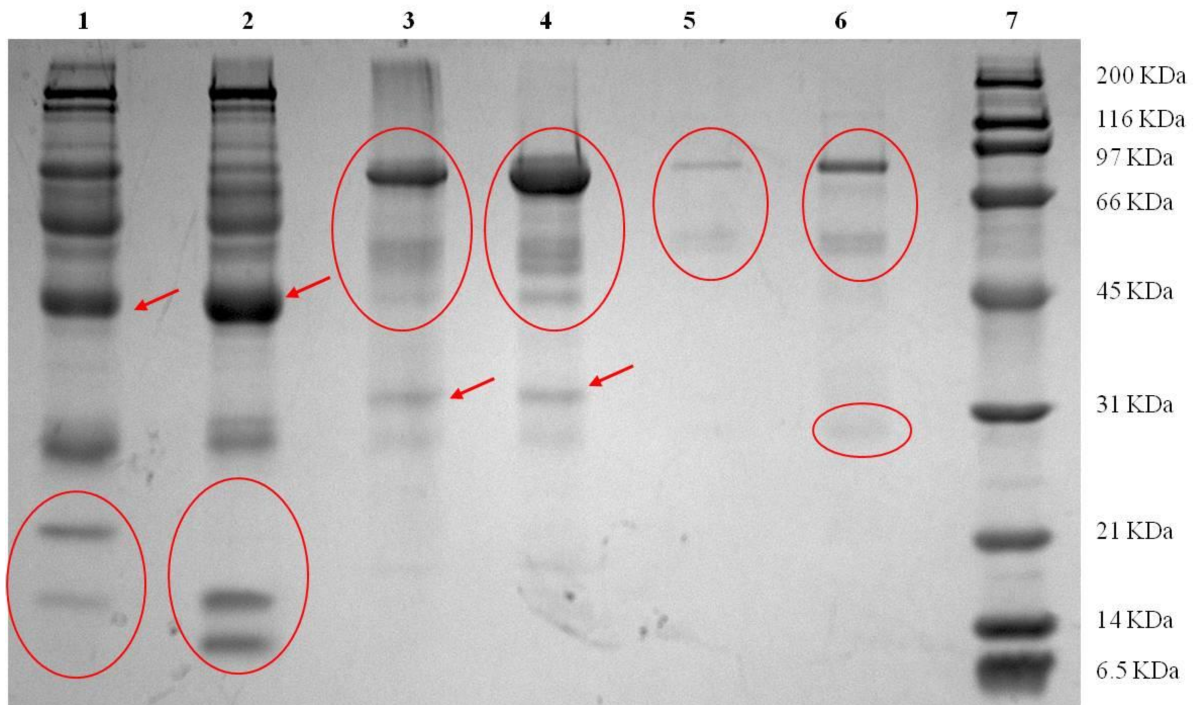


Figure 14. *Analysis of bound glycoprotein fractions from the tandem lectin series on 1D SDS-PAGE. Lanes 1, 3 and 5 represent the fractions from WGA, Con A and RCA-I from disease-free serum, respectively, whereas 2, 4, 6 represents the fractions from WGA, Con A and RCA-I from the breast cancer serum, respectively. Lane 7 represents the molecular weight markers.*

Spectral Count as a Means to Assess the Differentially Expressed Proteins

The spectral count is the total number of spectra that were obtained for an identified protein from LC-MS/MS analysis. The spectral count allows the determination

of the differentially expressed proteins between the disease-free and the cancer samples without any labeling techniques. The spectral count has been used in a number of comparative proteomic studies and shown to perform better than stable-isotope labeling methods [27, 28].

In this study, an average spectral count of less than 4 was not considered to make comparison between the disease-free and the cancer serum. A t-test was performed and only those that had a p-value < 0.05 were listed in tables 7, 8 and 9. Only those proteins that showed at least two fold changes in their spectral count ratios were considered as significantly changed.

TABLE 7

PROTEINS THAT WERE DIFFERENTIALLY EXPRESSED IN WGA FRACTION

Identified proteins	Average spectral count		Ratio	
	Disease-free	Cancer	Cancer/ Disease-free	Disease-free /Cancer
Alpha-2-antiplasmin	4	0	L	H
Angiotensinogen	12	4	0.37	2.69
Beta-2-glycoprotein 1	24	7	0.27	3.65
Carboxypeptidase N subunit 2	4	0	L	H
CD5 antigen-like*	20	7	0.37	2.68
Complement factor I	13	6	0.48	2.11
Corticosteroid-binding globulin	7	0	L	H
Hemoglobin subunit alpha	11	37	3.39	0.29
Hemoglobin subunit beta	13	59	4.56	0.22
Hemoglobin subunit delta*	0	33	H	L
Histidine-rich glycoprotein	0	4	H	L
Ig alpha-2 chain C region	42	0	L	H
Ig heavy chain V-III region BRO*	5	0	L	H
Ig heavy chain V-III region VH26*	5	0	L	H
Ig kappa chain C region*	33	13	0.38	2.63
Ig kappa chain V-I region EU*	4	0	L	H

Ig kappa chain V-III region HAH*	5	0	L	H
Ig mu chain C region	82	40	0.48	2.08
Inter-alpha-trypsin inhibitor heavy chain H1	34	13	0.38	2.62
Keratin, type I cytoskeletal 10*	0	5	H	L
Keratin, type II cytoskeletal 1*	0	5	H	L
N-acetylmuramoyl-L-alanine amidase	7	0	L	H
Pregnancy zone protein	56	0	L	H

*Non-glycoproteins
H and L represents that the protein is either present in high or low amount, either in disease-free or cancer serum.

The proteins hemoglobin subunit delta, histidine-rich protein, keratin type I cytoskeletal 10 and keratin type II cytoskeletal 1 were identified in the WGA cancer serum fraction but not from the WGA fraction from the disease-free serum (see Table 7). Histidine-rich glycoprotein has been reported to have elevated levels of sialylation and fucosylation that can distinguish colorectal cancer from adenoma and disease-free human plasma [29]. This was also observed in the current study where histidine-rich-protein was over-expressed in the cancer serum only in the WGA fraction, indicating its elevated level of sialylation.

The proteins that were either up- or down-regulated by more than two fold were not common for the three-lectin columns (see Fig. 13), thus allowing the detection of more differentially expressed glycoproteins than one could detect by using only a single lectin column. A recent article states that high-mannose glycans which has affinity towards Con A are elevated in the progression of breast cancer [30]. Table 8 shows the differentially expressed proteins in the Con A fractions where one can see that all the proteins were over-expressed by more than two-fold change in the cancer serum. β-2-glycoprotein has already been reported to have altered glycosylation and over expressed

in breast cancer and pancreatic cancer [31, 32]. The same was observed in the current study were it was found to be over expressed by 2.6 fold in the cancer serum. Also, the Con A column captured six other proteins only from the cancer serum but not from the disease-free serum as shown in Table 8. Out of these, afamin was reported to be down-expressed in ovarian cancer serum and also was reported as a potential biomarker for ovarian cancer [33]. But in the current study afamin was found to be more than two-fold over expressed in the breast cancer serum, this might be due to the difference in the localization of the tumor. Surprisingly, serum albumin a non-glycoprotein was also found to be elevated in the breast cancer serum. Since albumin was observed only in the breast cancer serum and not in the normal serum of the Con A fraction, the elevated level could be due its interaction with other highly elevated glycoproteins present in the breast cancer serum. This elevated level of albumin in breast cancer serum is in agreement with the findings reported [23] where different lectins were used to make comparison between breast cancer and disease-free sera.

TABLE 8

PROTEINS THAT WERE DIFFERENTIALLY EXPRESSED IN CON A FRACTION

Identified proteins	Average spectral count		Ratio	
	Disease-free	Cancer	Cancer/ Disease-free	Disease-free /Cancer
Afamin	0	11	H	L
Beta-2-glycoprotein 1	4	11	2.67	0.38
Haptoglobin	0	8	H	L
Hemoglobin subunit alpha	0	7	H	L
Ig alpha-1 chain C region	0	4	H	L
Ig gamma-2 chain C region	0	11	H	L
Serum albumin*	0	6	H	L

*Non-glycoproteins
H and L represents that the protein is either present in high or low amount, either in disease-free or cancer serum.

The proteins that were differentially expressed in the RCA-I column are listed in Table 9. As can be seen in this table, the proteins that were over expressed by two-fold in the RCA-I fraction were hemopexin, serum albumin and zinc-α-2-glycoprotein. Recently, hemopexin glycan has been shown to have altered glycosylation in patients with hepatocellular carcinoma and it can serve as a complementary test to α-fetoprotein to identify hepatocellular carcinoma with cirrhosis [34]. In the present study, hemopexin was found to be four times over expressed in the RCA-I fraction of the cancer serum. Zinc-α-2-glycoprotein was reported to be a potential biomarker for breast cancer and prostate cancer [35, 36]. It has also been reported to be in elevated levels in urinary bladder cancer [37]. The elevated level of serum albumin has been already discussed in the preceding section.

TABLE 9

PROTEINS THAT WERE DIFFERENTIALLY EXPRESSED IN RCA-I FRACTION

Identified proteins	Average spectral count		Ratio	
	Disease-free	Cancer	Cancer/ Disease-free	Disease-free /Cancer
Alpha-1-acid glycoprotein 1	19	9	0.50	2.00
Complement factor B	25	0	L	H
Hemopexin	6	24	4.29	0.23
Keratin, type I cytoskeletal 10*	56	6	0.11	8.84
Keratin, type I cytoskeletal 14*	12	0	L	H
Keratin, type II cytoskeletal 1*	62	8	0.12	8.04
Keratin, type II cytoskeletal 2 epidermal*	53	5	0.10	9.88
Keratin, type II cytoskeletal 5*	18	0	L	H
Serum albumin*	4	13	3.17	0.32
Zinc-alpha-2-glycoprotein	10	20	2.03	0.49

*Non-glycoproteins
H and L represents that the protein is either present in high or low amount, either in disease-free or cancer serum.

Hemoglobin subunit α and β Hemoglobin subunits α and β were found to be over expressed in the cancer serum (3.39 and 4.56 fold change) fraction of the WGA column. Also, the α subunit of hemoglobin was found to be over-expressed in the cancer serum fraction from the Con A column. These two proteins have been reported to be potential biomarkers for ovarian cancer [38]. Another article reported that α- and β hemoglobin to be as putative markers for ovarian and head and neck cancer [39]. The high level of α and β subunits of hemoglobin in the current study is in agreement with the results of a recent article, in which the ductal carcinoma *in situ* (DCIS) type cancer serum was compared with the control samples using SELDI-TOF [40]. Also, the level of hemoglobin β−chain isoforms in nipple aspirate fluid have been found to be correlated to breast cancer [41]. As reported [42] hemoglobin can form an irreversible and non-covalent complex with haptoglobin in the presence of sialic acid and galactose present in the carbohydrate moiety of haptoglobin. The elevated level of hemoglobin (in WGA and Con A) and haptoglobin (in Con A) suggest that the hemoglobin can serve as a potential indicator of the breast cancer. Thus, the high-elevated level of hemoglobin subunits α and β can be attributed to the changes taking place in blood serum as a result of tumor progression.

Conclusions

The strategy introduced here to capture the wide range of glycoproteins allowed the enrichment of several glycoproteins, which were differentially expressed in cancer sera with respect to disease free sera. Even though, the number of human subjects used by the supplier to obtain the breast cancer serum was very small, the strategy discussed

here allows identification of the differentially expressed proteins in breast cancer serum with respect to the disease-free serum by avoiding the depletion of high abundance proteins. Also, by using spectral counts as a tool to identify the differentially expressed proteins in serum, other expensive and laborious techniques such as isotope labeling methods were avoided. Since the strategy investigated here is a proof-of-concept for enrichment of glycoproteins from a complex mixture such as serum, further validation needs to be done to validate the biomarker candidates identified in this study.

References

[1] Y. S. Kim, H. S. Yoo and J. H. Ko, *Protein Pept. Lett.* **2009**, *16*, 499-507.

[2] I. Brockhausen, J. Schutzbach and W. Kuhns, *Acta Anat. (Basel)* **1998**, *161*, 36-78.

[3] J. Granger, J. Siddiqui, S. Copeland and D. Remick, *Proteomics* **2005**, *5*, 4713-4718.

[4] M. Madera, B. Mann, Y. Mechref and M. V. Novotny, *J. Sep. Sci.* **2008**, *31*, 2722-2732.

[5] T. Schwientek, U. Mandel, U. Roth, S. Muller and F. G. Hanisch, *Proteomics* **2007**, *7*, 3264-3277.

[6] M. Durham and F. E. Regnier, *J. Chromatogr. A* **2006**, *1132*, 165-173.

[7] B. Mann, M. Madera, I. Klouckova, Y. Mechref, L. E. Dobrolecki, R. J. Hickey, Z. T. Hammoud and M. V. Novotny, *Electrophoresis* **2010**, *31*, 1833-1841.

[8] R. Qiu and F. E. Regnier, *Anal. Chem.* **2005**, *77*, 7225-7231.

[9] Z. Yang and W. S. Hancock, *J. Chromatogr. A* **2004**, *1053*, 79-88.

[10] Z. Yang and W. S. Hancock, *J. Chromatogr. A* **2005**, *1070*, 57-64.

[11] Z. Yang, L. E. Harris, D. E. Palmer-Toy and W. S. Hancock, *Clin. Chem.* **2006**, *52*, 1897-1905.

[12] M. Kullolli, W. S. Hancock and M. Hincapie, *J. Sep. Sci.* **2008**, *31*, 2733-2739.

[13] Z. Zeng, M. Hincapie, B. B. Haab, S. Hanash, S. J. Pitteri, S. Kluck, J. M. Hogan, J. Kennedy and W. S. Hancock, *J. Chromatogr. A* **2010**, *1217*, 3307-3315.

[14] M. Madera, Y. Mechref, I. Klouckova and M. V. Novotny, *J. Chromatogr. B* **2007**, *845*, 121-137.

[15] M. Bedair and Z. El Rassi, *J. Chromatogr. A* **2005**, *1079*, 236-245.

[16] A. Kobata and K. Yamashita Glycobiology A Practical Approach, *Vol.* (Eds.): M. Fukuda and A. Kobata, IRL Press, Oxford, **1993**, pp. 103-125.

[17] K. Yamamoto, T. Tsuji, I. Matsumoto and T. Osawa, *Biochem. J.* **1981**, *20*, 5894-5899.

[18] J. U. Baenziger and D. Fiete, *J. Biol. Chem.* **1979**, *254*, 9795-9799.

[19] H. Lis and N. Sharon Biology of Carbohydrates, *Vol. 2* (Eds.): V. Ginsburg and P. W. Robbins, John Wiley & Sons, New York, **1984**, pp. 9-22.

[20] H. W. Zhong and Z. El Rassi, *J. Sep. Sci.* **2009**, *32*, 10-20.

[21] Bollag, Rozycki, M.D., Edelstein, S., *Wiley-Liss, New York* **1996**, 107-154.

[22] A. Jenkins, J. McBride, A. Januszewski, C. Karschimkus, B. Zhang, D. O'Neal, C. Nelson, J. Chung, C. Harper, T. Lyons and J.-X. Ma, *J. Angiogenes. Res.* **2010**, *2*, 19.

[23] K. Jung, W. Cho and F. E. Regnier, *J. Proteome Res.* **2009**, *8*, 643-650.

[24] W. P. Mielicki, M. Tenderenda, P. Rutkowski and K. Chojnowski, *Cancer Lett.* **1999**, *146*, 61-66.

[25] S. Hirani, J. D. Lambris and H. J. Muller-Eberhard, *Biochem. J.* **1986**, *233*, 613-616.

[26] M. D. Crispin, G. E. Ritchie, A. J. Critchley, B. P. Morgan, I. A. Wilson, R. A. Dwek, R. B. Sim and P. M. Rudd, *FEBS Lett.* **2004**, *566*, 270-274.

[27] J. M. Asara, H. R. Christofk, L. M. Freimark and L. C. Cantley, *Proteomics* **2008**, *8*, 994-999.

[28] C. W. Turck, A. M. Falick, J. A. Kowalak, W. S. Lane, K. S. Lilley, B. S. Phinney, S. T. Weintraub, H. E. Witkowska and N. A. Yates, *Mol. Cell. Proteomics* **2007**, *6*, 1291-1298.

[29] Y. H. Qiu, T. H. Patwa, L. Xu, K. Shedden, D. E. Misek, M. Tuck, G. Jin, M. T. Ruffin, D. K. Turgeon, S. Synal, R. Bresalier, N. Marcon, D. E. Brenner and D. M. Lubman, *J. Proteome Res.* **2008**, *7*, 1693-1703.

[30] M. L. de Leoz, L. J. Young, H. J. An, S. R. Kronewitter, J. Kim, S. Miyamoto, A. D. Borowsky, H. K. Chew and C. B. Lebrilla, *Mol. Cell. Proteomics* **2011**, *10*, M110 002717.

[31] J. Zhao, T. H. Patwa, W. Qiu, K. Shedden, R. Hinderer, D. E. Misek, M. A. Anderson, D. M. Simeone and D. M. Lubman, *J. Proteome Res.* **2007**, *6*, 1864-1874.

[32] E. A. Alsabti and K. Muneir, *Jpn. J. Exp. Med.* **1979**, *49*, 235-240.

[33] D. Jackson, R. A. Craven, R. C. Hutson, I. Graze, P. Lueth, R. P. Tonge, J. L. Hartley, J. Nickson, S. J. Rayner, C. Johnston, B. Dieplinger, M. Hubalek, N. Wilkinson, T. J. Perren, S. Kehoe, G. D. Hall, G. Daxenbichler, H. Dieplinger, P. J. Selby and R. E. Banks, *Clin. Cancer Res.* **2007**, *13*, 7370-7379.

[34] E. N. Debruyne, D. Vanderschaeghe, H. Van Vlierberghe, A. Vanhecke, N. Callewaert and J. R. Delanghe, *Clin. Chem.* **2010**, *56*, 823-831.

[35] V. Dubois, L. Delort, F. Mishellany, T. Jarde, H. Billard, C. Lequeux, O. Damour, F. Penault-Llorca, M. P. Vasson and F. Caldefte-Chezet, *Anticancer Res.* **2010**, *30*, 2919-2925.

[36] L. P. Hale, D. T. Price, L. M. Sanchez, W. Demark-Wahnefried and J. F. Madden, *Clin. Cancer Res.* **2001**, *7*, 846-853.

[37] S. Irmak, D. Tilki, J. Heukeshoven, L. Oliveira-Ferrer, M. Friedrich, H. Huland and S. Ergun, *Proteomics* **2005**, *5*, 4296-4304.

[38] W. S. Ahn, S. P. Park, S. M. Bae, J. M. Lee, S. E. Namkoong, G. H. Nam, Y. L. Cho, H. S. Choi, H. J. Jun, C. K. Kim, Y. W. Kim, B. D. Han and H. S. Jin, *Cancer Sci.* **2005**, *96*, 197-201.

[39] M. Roesch-Ely, M. Nees, S. Karsai, A. Ruess, R. Bogumil, U. Warnken, M. Schnolzer, A. Dietz, P. K. Plinkert, C. Hofele and F. X. Bosch, *Oncogene* **2007**, *26*, 54-64.

[40] J. Solassol, P. Rouanet, P. J. Lamy, C. Allal, G. Favre, T. Maudelonde and A. Mange, *Oncogene* **2010**, *29*, 550-560.

[41] H. Alexander, A. L. Stegner, C. Wagner-Mann, G. C. Du Bois, S. Alexander and E. R. Sauter, *Clin. Cancer Res.* **2004**, *10*, 7500-7510.

[42] V. Kaartinen and I. Mononen, *Biochimi. Biophys. Acta* **1988**, *953*, 345-352.

CHAPTER IV

LECTIN AFFINITY CHROMATOGRAPHY USING NARROW SPECIFICITY LECTINS FOR CAPTURING FUCOSYLATED AND SIALYLATED GLYCOPROTEINS FROM BREAST CANCER AND DISEASE FREE HUMAN SERA USING TANDEM MONOLITHIC COLUMNS WITH SURFACE IMMOBILIZED *ALEURIA AURANTIA* LECTIN, *SAMBUCCUS NIGRA* AGGLUTININ AND *LOTUS TETRAGONOLOBUS* AGGLUTININ

Introduction

According to recent statistics, about 1 in 8 women in USA will develop invasive breast cancer and there were about 2.5 million breast cancer survivors in the year 2010 in the USA [1]. Detection of cancer at early stage can profoundly increase the survival rate of cancer patients. But early stage detection remains a major challenge as only minimal symptoms are observed at that stage of cancer. During metastasis, tumor cells dissociate from a primary site, enter blood stream and proliferate in another site. During this process, the tumor-associated proteins can shed from the cell surface and enter the blood stream. Thus, blood can reflect the patho-physiological state of a person [2]. Even though, the Food and Drug Administration (FDA) has approved the proteins CA 15.3 and

Her-2/neu as cancer biomarkers for breast cancer, to decrease the death rate and to detect cancer at an early stage, more biomarker candidates are needed using an easily accessible clinical sample such as human serum.

It is well known that the alteration in glycosylation of proteins takes place in several diseases including cancer [3-5]. Especially aberrations in sialylation and fucosylation of glycoproteins are thought to be associated with tumor progression. For example, Warren *et al*. [6], proposed that sialylation is an important step that leads to the alteration of cell surface sugar chains as a result of malignant transformations. Also, sialyl Lewis x (Le^x) glycans have been identified in tumor cells [7]. Sialyltransferase is an enzyme that is involved in addition of sialic acid to sugars, and an increased activity of this enzyme can lead to an increase in sialylation of tumor cell surfaces [2]. Increase in fucosylation is another observed change in the tumor. For example, increased activity of α-(1,3)-fucosyl transferase was observed in tumor, and that resulted in elevation of fucosylation in haptoglobin [8]. Also, changes in fucosylation of glycoproteins in breast cancer tissues have been reported [9].

There have been some reports in the literature regarding the targeting of specific type of glycoproteins [10-14] to identify differentially expressed glycoproteins in cancer serum as compared to disease-free serum. In one report, sialylated serum glycoproteins were targeted using wheat germ agglutinin, *Sambuccus nigra* agglutinin (SNA) and *Maackia amurensis* lectin [11], which have specificity toward sialylated glycoproteins. In other reports, fucosylated proteins were targeted using the lectins *Aleuria aurantia* lectin (AAL) [12] and *Lotus tetragonolobus* agglutinin (LTA) [13], whereas the β(1,6)-branched glycoforms of *N*-linked glycoproteins were captured by L-phytohemagglutinin

lectin [10]. None of these methods targeted both fucosylated and sialylated glycoproteins simultaneously. Since both fucosylation and sialylation of glycoprotein can be altered as a result of malignant transformations, a method to selectively enrich both fucosylated and sialylated serum glycoproteins is urgently needed.

With the intention of capturing the fucosylated and the sialylated serum glycoproteins, three narrow specificity lectins namely, AAL, LTA and SNA were immobilized onto the surface of a glyceryl methacrylate (GMM)/ethylene glycol dimethacrylate (EDMA) monolith. AAL has a strong affinity towards core fucosylated glycans (i.e.,) where a fucose residue is attached to the innermost *N*-acetyl-D-glucosamine (GlcNAc) of the *N*-linked-core structure represented as Fuc α1→6GlcNAc→R and has weak binding towards fucose in the outer arm such as Fuc α1→2Galβ1→4GlcNAcβ1→R, Galβ1→4(Fucα1→3)GlcNAc→R and Galβ1→3(Fucα1→4)GlcNAc→R, where R = H or sugar [15]. LTA can bind to glycans having fucose present in the outer arm Fucα1→3/1→4GlcNAc and Fucα1→2Gal. LTA also has an affinity for glycans containing the Le^x determinant represented as Galβ1→4(Fucα1→3)GlcNAc β1→R [16]. The haptenic sugar for AAL and LTA is α-L-fucose. SNA lectin has strong binding towards NeuAcα2→6Galβ1→4GlcNAcβ1→R in *N*-glycans and NeuAcα2→6GalNAc→Ser(Thr) in *O*-glycans, where R = H or sugar. The haptenic sugar for SNA is lactose [15]. In this current study depletion of high abundance proteins was not done as it is known that depletion might result in co-depletion of other clinically important proteins. For, example, one study states that 210 proteins were removed with the depletion of six high abundant proteins [17].

Experimental

Instrumentation

The instruments used were same as those described in Chapter III.

Reagents and Materials

The three unconjugated lectins namely, AAL, LTA and SNA were purchased from Vector Laboratories (Burlingame, CA, USA). All other chemicals and reagents used are the same as in Chapter III.

Monolithic Affinity Columns

The monolith preparation was prepared as described in Chapter III and the monolithic support was transferred from the 25.0 cm column to a 3 cm column.

Immobilization of the Lectins

The immobilization process was similar to that described in Chapter III. The 3 cm monolithic columns were allowed to react with freshly prepared 0.1 M $NaIO_4$ for 2 h at room temperature. The immobilization was done on the column by passing a solution of 1 mg of AAL or LTA in 0.5 mL of 0.1 M sodium acetate at pH 6.4 containing 0.1 mM of Ca^{2+}, 0.1 M of fucose and 50 mM of sodium cyanoborohydride through the column for 12 h at room temperature. SNA was immobilized using the same procedure, but in absence of Ca^{2+} and using lactose as the haptenic sugar. These immobilized lectin columns were stored with the mobile phase containing 20 mM of Tris-HCl (pH 6.0) containing 100 mM of NaCl, 0.1 mM of Ca^{2+} and 0.08% of NaN_3 at 4 °C until use.

Fractionation of Glycoproteins From Human Serum - Chromatographic Conditions

In all the experiments, serum was used in 1:3 dilution ratio and a total of 600 μL was injected into the tandem lectin columns. The columns were first equilibrated with 10 column volumes of the binding mobile phase consisting of 20 mM of Tris-HCl (pH 6.0) containing 100 mM of NaCl, 0.1 mM of Ca^{2+}. The diluted serum (600 μL) was injected onto the tandem columns and the unbound serum proteins were washed with the binding mobile phase at flow rate 1 mL/min. The bound fractions from the LTA and AAL columns were eluted using 5 mM fucose in the binding mobile phase. The SNA column was eluted using 0.1 M lactose in the binding mobile phase. The chromatographic setup was same as in Chapter III.

Protein Assay

The procedure was same as that described in Chapter III.

LC-MS/MS Methodology

The conditions were same as in Chapter III, except that in this analysis, the six most intense ions were analyzed *via* MS/MS in the linear ion trap.

LC-MS/MS Data Analysis

Conditions were same as in Chapter III, except for the parent ion tolerance, which was 20 ppm.

Results and Discussion

Analysis of Proteins in the Lectin Fractions

The lectins LTA, AAL and SNA which have narrow selectivity towards glycans were used in this investigation to identify the differentially expressed proteins in breast cancer serum relative to disease-free serum. These lectins were selected based on the fact the LTA and AAL can capture the fucosylated glycoproteins at different sugar residues while the SNA column can capture the sialylated glycoproteins. The lectin columns were arranged in the tandem series in the order LTA→AAL→SNA. This order was chosen based on the fact that LTA is more specific in its affinity than AAL in the sense that LTA recognizes glycans with fucose residues in the outer arm whereas AAL binds to glycans with fucose residues in the inner core and to a lesser extent in the outer arm. Thus, by placing LTA first in the tandem column format, the AAL will have all its binding sites available for the glycoproteins bearing glycans with inner core fucosylation. On the other hand, SNA will capture the sialylated glycoproteins that pass through the first two lectin columns. The chromatographic set-up containing the tandem lectin columns was the same as in Chapter III (see Fig. 9).

Disease-free serum (n = 3) and cancer serum (n = 1) were injected into the tandem lectin columns and the bound fractions were eluted with the haptenic sugars, and subsequently subjected to dialysis. An aliquot of the dialyzed fractions thus obtained were analyzed using LC-MS/MS. As mentioned in Experimental, only proteins that exhibited identification probability greater than 99% with peptide identification probability greater than 95% containing at least two unique peptides were considered, and they are reported in Table 1.

. TABLE 1

LIST OF PROTEINS IDENTIFIED IN THE LECTIN-BOUND FRACTIONS

Identified Proteins	Accession Number	Mol. Wt.	Disease-free serum			Breast cancer serum		
			LTA	AAL	SNA	LTA	AAL	SNA
			Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]	Avg SC[a]/ Avg UP[b]
6-phosphofructokinase, muscle type*	K6PF_HUMAN	85 kDa			S			2/2
Actin, alpha skeletal muscle*	ACTS_HUMAN	42 kDa			S			61/16
Afamin	AFAM_HUMAN	69 kDa	4/2			L		
Alpha-1-acid glycoprotein 1	A1AG1_HUMAN	24 kDa	16/7	26/11	4/2	15/7	10/6	S
Alpha-1-acid glycoprotein 2	A1AG2_HUMAN	24 kDa	9/2	21/6		6/1	8/3	
Alpha-1-antichymotrypsin	AACT_HUMAN	48 kDa	7/5	9/6	S	9/6	9/5	3/1
Alpha-1-antitrypsin	A1AT_HUMAN	47 kDa	52/20	23/14	10/8	48/21	21/13	12/8
Alpha-1B-glycoprotein	A1BG_HUMAN	54 kDa	9/6		2/1	8/6		2/2
Alpha-2-HS-glycoprotein	FETUA_HUMAN	39 kDa	2/2	A	S	6/4	2/1	2/2
Alpha-2-macroglobulin	A2MG_HUMAN	163 kDa	99/44	196/73	16/10	101/47	183/69	40/25
Alpha-actinin-2*	ACTN2_HUMAN	104 kDa			S			11/6
Alpha-actinin-3*	ACTN3_HUMAN	103 kDa			S			14/9
Angiotensinogen	ANGT_HUMAN	53 kDa	10/6	A	S	6/5	1/1	1/1
Antithrombin-III	ANT3_HUMAN	53 kDa	4/3			8/5		
Apolipoprotein A-I	APOA1_HUMAN	31 kDa	46/20	28/14	S	50/21	28/16	9/8
Apolipoprotein A-II*	APOA2_HUMAN	11 kDa	7/4	6/3		10/4	5/3	
Apolipoprotein A-IV*	APOA4_HUMAN	45 kDa	L			2/1		
Apolipoprotein B-100	APOB_HUMAN	516 kDa	25/18	56/40	10/9	46/32	62/42	65/49
Apolipoprotein C-I*	APOC1_HUMAN	9 kDa	0			1		
Apolipoprotein D	APOD_HUMAN	21 kDa	3/2	1	S	4/3	2/2	3/2

Apolipoprotein E	APOE_HUMAN	36 kDa	L	A		6/5	2/2	
Apolipoprotein L1	APOL1_HUMAN	44 kDa		A			4/2	
Apolipoprotein(a)	APOA_HUMAN	501 kDa		2/1	3/2		4/2	5/2
Attractin	ATRN_HUMAN	159 kDa		6/5			A	
Beta-2-glycoprotein 1	APOH_HUMAN	38 kDa	22/11		S	17/9		3/2
Beta-enolase*	ENOB_HUMAN	47 kDa			S			13/8
Carboxypeptidase N catalytic chain	CBPN_HUMAN	52 kDa		7/3			1/1	
Carboxypeptidase N subunit 2	CPN2_HUMAN	61 kDa		9/6			5/3	
CD44 antigen	CD44_HUMAN	82 kDa		A			1/1	
CD5 antigen-like*	CD5L_HUMAN	38 kDa	11/7	22/13	4/2	2/2	14/9	5/4
Ceruloplasmin	CERU_HUMAN	122 kDa	25/14	38/21	5/4	24/14	43/24	21/12
Clusterin	CLUS_HUMAN	52 kDa	2/1	16/8		2/1	12 /6	
Coagulation factor X	FA10_HUMAN	55 kDa		1/1			2/2	
Complement C1q subcomponent subunit C	C1QC_HUMAN	26 kDa		1/1			A	
Complement C1r subcomponent	C1R_HUMAN	80 kDa		1/1	2/2		4/3	6/5
Complement C1r subcomponent-like protein	C1RL_HUMAN	53 kDa		1/1			A	
Complement C1s subcomponent	C1S_HUMAN	77 kDa	L	4/3	7/5	1/1	5/3	8/6
Complement C2	CO2_HUMAN	83 kDa			S			1/1
Complement C3	CO3_HUMAN	187 kDa	97/52	12/8	S	117/62	64/38	35/24
Complement C4-A	CO4A_HUMAN (+1)	193 kDa	16/11		14/10	27/18		39/23
Complement C4-B	CO4B_HUMAN	193 kDa		13/10			41/28	
Complement C5	CO5_HUMAN	188 kDa	1/1			L		
Complement component C6	CO6_HUMAN	105 kDa			S			1/1
Complement factor B	CFAB_HUMAN	86 kDa	18/12		7/5	17/10		14/10
Complement factor H	CFAH_HUMAN	139 kDa	26/17	5/5	19/13	37/26	13/10	66/37
Complement factor H-related protein 1	FHR1_HUMAN	38 kDa			S			9/3

Complement factor H-related protein 3	FHR3_HUMAN	37 kDa		A			1/1	
Corticosteroid-binding globulin	CBG_HUMAN	45 kDa	1/1	1		L	A	
Creatine kinase M-type*	KCRM_HUMAN	43 kDa			S			19/10
Cryptochrome-1*	CRY1_HUMAN	66 kDa		1/1			A	
Desmin*	DESM_HUMAN	54 kDa			S			5/5
Fibrinogen alpha chain	FIBA_HUMAN	95 kDa	L	A	S	14/7	9/7	16/8
Fibrinogen beta chain	FIBB_HUMAN	56 kDa	L	A	S	22/12	14/9	16/9
Fibrinogen gamma chain	FIBG_HUMAN	52 kDa	L	A	S	25/13	11/7	21/12
Fibronectin	FINC_HUMAN	263 kDa			S			1/1
Ficolin-3	FCN3_HUMAN	33 kDa		16/8	1/1		26/11	4/3
Filamin-C*	FLNC_HUMAN	291 kDa			S			5/4
Fructose-bisphosphate aldolase A*	ALDOA_HUMAN	39 kDa			S			17/12
Gelsolin*	GELS_HUMAN	86 kDa	1/1			L		
Glyceraldehyde-3-phosphate dehydrogenase*	G3P_HUMAN	36 kDa			S			6/2
Glycogen phosphorylase, muscle form*	PYGM_HUMAN	97 kDa			S			11/8
Haptoglobin	HPT_HUMAN	45 kDa	67/25	89/29	40/20	59/22	30/16	43/20
Haptoglobin-related protein*	HPTR_HUMAN	39 kDa		41/2			11/2	
Hemoglobin subunit alpha	HBA_HUMAN	15 kDa	1/1	8/3	S	24/8	7/3	6/2
Hemoglobin subunit beta	HBB_HUMAN	16 kDa	5/3	10/5	S	39/15	8/4	10/6
Hemopexin	HEMO_HUMAN	52 kDa	29/14	17/10	6/4	28/13	21/11	20/8
Heparin cofactor 2	HEP2_HUMAN	57 kDa	1/1		1/1	1/1		8/5
Histidine-rich glycoprotein	HRG_HUMAN	60 kDa		6/3	S		12/7	3/2
Ig alpha-1 chain C region	IGHA1_HUMAN	38 kDa	28/11	46/17	15/7	28/13	50/18	18/8
Ig alpha-2 chain C region	IGHA2_HUMAN	37 kDa		34/3			36/3	
Ig gamma-1 chain C region	IGHG1_HUMAN	36 kDa	45/16	24/11	5/4	47/16	32/14	20/10
Ig gamma-2 chain C region	IGHG2_HUMAN	36 kDa	47/10	16/6	6/2	50/10	18/6	15/4
Ig gamma-3 chain C region	IGHG3_HUMAN	41 kDa	40/5		5/1	41/5		16/4

Ig gamma-4 chain C region	IGHG4_HUMAN	36 kDa	29/4	A		28/4	17/2	
Ig heavy chain V-I region HG3*	HV102_HUMAN	13 kDa	2/1	1/1	S	5/2	4/2	1/1
Ig heavy chain V-II region ARH-77*	HV209_HUMAN	16 kDa	2/2	A		3/2	2/1	
Ig heavy chain V-II region WAH	HV206_HUMAN	14 kDa		A			1/1	
Ig heavy chain V-III region BRO*	HV305_HUMAN	13 kDa	8/2		2/2	7/2		4/2
Ig heavy chain V-III region BUT*	HV306_HUMAN	12 kDa		4/2			4/2	
Ig heavy chain V-III region GA*	HV308_HUMAN	13 kDa	1/1	A		L	2/1	
Ig heavy chain V-III region GAL*	HV320_HUMAN	13 kDa	3/2	3/2	1/1	2/2	6/2	2/1
Ig heavy chain V-III region KOL*	HV311_HUMAN	14 kDa	L			1/1		
Ig heavy chain V-III region VH26*	HV303_HUMAN	13 kDa	2/2	2	2/2	3/2	2/1	3/2
Ig kappa chain C region*	IGKC_HUMAN	12 kDa	30/7	28/8	7/3	22/6	23/7	19/5
Ig kappa chain V-I region EU*	KV106_HUMAN	12 kDa	5/2	5/2	S	4/2	4/2	2/1
Ig kappa chain V-I region WEA*	KV118_HUMAN	12 kDa	2/1			L		
Ig kappa chain V-III region HAH*	KV312_HUMAN (+1)	14 kDa	5/3	8/4		7/4	9/4	
Ig kappa chain V-III region POM*	KV306_HUMAN (+1)	12 kDa			S			1/1
Ig kappa chain V-III region SIE*	KV302_HUMAN (+3)	12 kDa			3/2			1/1
Ig kappa chain V-III region VG (Fragment) *	KV309_HUMAN	13 kDa	L	1		4/3	2/1	
Ig kappa chain V-III region VH (Fragment) *	KV310_HUMAN	13 kDa		1			1/1	
Ig kappa chain V-IV region (Fragment) *	KV401_HUMAN (+1)	13 kDa	3/2	5/3		L	4/3	
Ig lambda chain V-I region NEWM*	LV105_HUMAN	11 kDa		A			1/1	
Ig lambda chain V-I region WAH*	LV106_HUMAN	12 kDa	1/1	1/1	S	1/1	2/2	1/1
Ig lambda chain V-III region LOI*	LV302_HUMAN	12 kDa	6/3	4/3	S	8/4	7/3	2/1
Ig lambda chain V-III region SH*	LV301_HUMAN	11 kDa	2/1	2/2		1/1	1/1	
Ig lambda-1 chain C regions*	LAC1_HUMAN	11 kDa	10/2	15/2		13/2	16/2	

Ig lambda-2 chain C regions*	LAC2_HUMAN	11 kDa	21/6	18/5	9/5	19/5	19/6	13/5
Ig lambda-7 chain C region*	LAC7_HUMAN	11 kDa	L			4/1		
Ig mu chain C region	IGHM_HUMAN	49 kDa	38/14	68/21	30/13	24/13	57/16	42/13
Ig mu heavy chain disease protein	MUCB_HUMAN	43 kDa		A			12/1	
IgGFc-binding protein	FCGBP_HUMAN	572 kDa		7/6			2/1	
Immunoglobulin J chain	IGJ_HUMAN	18 kDa	5/3	8/4	4/3	5/3	9/5	6/4
Insulin-like growth factor-binding protein 3	IBP3_HUMAN	32 kDa		A			1/1	
Insulin-like growth factor-binding protein complex acid labile subunit	ALS_HUMAN	66 kDa		15/10			21/14	
Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1_HUMAN	101 kDa	4/3		S	3/2		4/3
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2_HUMAN	106 kDa	8/5	A	S	9/7	3/3	14/8
Inter-alpha-trypsin inhibitor heavy chain H4	ITIH4_HUMAN	103 kDa	6/4	7/4	4/3	117	5/3	11/8
Intercellular adhesion molecule 2	ICAM2_HUMAN	31 kDa		A			2/1	
Kallistatin	KAIN_HUMAN	49 kDa		2/2			A	
Keratin, type I cytoskeletal 10*	K1C10_HUMAN	59 kDa	25/14	41/18	33/19	24/15	56/23	37/20
Keratin, type I cytoskeletal 14*	K1C14_HUMAN	52 kDa		9/2	21/9		44/6	10/2
Keratin, type I cytoskeletal 16*	K1C16_HUMAN	51 kDa		11/6	8/1		73/21	S
Keratin, type I cytoskeletal 17*	K1C17_HUMAN	48 kDa		A			30/32	
Keratin, type I cytoskeletal 9*	K1C9_HUMAN	62 kDa	19/12	38/19	28/15	14/9	41/21	31/15
Keratin, type II cytoskeletal 1	K2C1_HUMAN	66 kDa	34/21	55/26	45/21	28/16	76/32	48/24
Keratin, type II cytoskeletal 2 epidermal*	K22E_HUMAN	65 kDa	17/9	27/12	30/4	21/11	45/16	24/13
Keratin, type II cytoskeletal 5*	K2C5_HUMAN	62 kDa		8/2	18/4		35/7	7/1
Keratin, type II cytoskeletal 6A*	K2C6A_HUMAN	60 kDa		10/8			75/38	
Keratin, type II cytoskeletal 6B*	K2C6B_HUMAN	60 kDa	L	A		7/1	73/3	
Keratin, type II cytoskeletal 6C*	K2C6C_HUMAN	60 kDa		A	18/10		50/1	9/2

Kininogen-1	KNG1_HUMAN	72 kDa	8/5	29/18	S	10/6	18/13	9/6
Leucine-rich alpha-2-glycoprotein	A2GL_HUMAN	38 kDa	5/3	1/1		6/5	A	
Lipopolysaccharide-binding protein	LBP_HUMAN	53 kDa			S			1/1
L-lactate dehydrogenase A chain*	LDHA_HUMAN	37 kDa			S			5/3
Low affinity immunoglobulin gamma Fc region receptor III-B	FCG3B_HUMAN	26 kDa		1/1			1/1	
L-selectin	LYAM1_HUMAN	42 kDa		104			8/4	
Myosin light chain 1/3, skeletal muscle isoform *	MYL1_HUMAN	21 kDa			S			11/5
Myosin regulatory light chain 2, skeletal muscle isoform*	MLRS_HUMAN	19 kDa			S			12/8
Myosin-1*	MYH1_HUMAN	223 kDa			S			148/71
Myosin-2*	MYH2_HUMAN	223 kDa			S			118/4
Myosin-4*	MYH4_HUMAN	223 kDa			S			89/3
Myosin-8 *	MYH8_HUMAN	223 kDa			S			86/3
Myosin-binding protein C, fast-type*	MYPC2_HUMAN	128 kDa			S			2/2
Neutrophil defensin 1*	DEF1_HUMAN (+1)	10 kDa	L			3//2		
Phosphoglucomutase-1*	PGM1_HUMAN	61 kDa			S			2/1
Plasma kallikrein	KLKB1_HUMAN	71 kDa	1/5	2/2	5/3	1/1	3/2	13/8
Plasma protease C1 inhibitor	IC1_HUMAN	55 kDa	6/4	21/10	3/3	11/8	39/16	12/7
Plasminogen	PLMN_HUMAN	91 kDa	5/4			9/7		
Platelet glycoprotein Ib alpha chain	GP1BA_HUMAN	69 kDa		3/2			4/3	
Protein AMBP	AMBP_HUMAN	39 kDa	5/4	3/3	S	5/4	4/3	1/1
Prothrombin	THRB_HUMAN	70 kDa	5/3	7/5	1/1	6/4	85	20/13
Pyruvate kinase isozymes M1/M2*	KPYM_HUMAN	58 kDa			S			11/8
Retinol-binding protein 4*	RET4_HUMAN	23 kDa	1/1			6/4		
Sarcoplasmic/endoplasmic reticulum calcium ATPase 1*	AT2A1_HUMAN	110 kDa			S			8/5

Scavenger receptor cysteine-rich type 1 protein M130	C163A_HUMAN	125 kDa		A			7/6	
Serotransferrin	TRFE_HUMAN	77 kDa	151/54	11/8	14/9	143/53	9/7	17/12
Serum albumin*	ALBU_HUMAN	69 kDa	248/62	46/25	13/10	219/57	67/32	42/22
Serum amyloid A-4 protein	SAA4_HUMAN	15 kDa	L	A		3/1	1/1	
Serum amyloid P-component	SAMP_HUMAN	25 kDa	13/7	20/8	9/5	5/4	A	9/6
Serum paraoxonase/arylesterase 1	PON1_HUMAN	40 kDa	1/1			L		
Sulfhydryl oxidase 1	QSOX1_HUMAN	83 kDa		2/2			8/6	
Thyroxine-binding globulin	THBG_HUMAN	46 kDa	2/1			2/2		
Titin*	TITIN_HUMAN	3816 kDa			S			9/7
Transthyretin	TTHY_HUMAN	16 kDa	3/2	A		8/4	1/1	
Triosephosphate isomerase*	TPIS_HUMAN	27 kDa			S			5/3
Tropomyosin alpha-1 chain*	TPM1_HUMAN	33 kDa			S			21/4
Tropomyosin beta chain*	TPM2_HUMAN	33 kDa			S			31/16
Troponin C, skeletal muscle*	TNNC2_HUMAN	18 kDa			S			3/2
Vitamin D-binding protein	VTDB_HUMAN	53 kDa	37/18			20/13		
Vitronectin	VTNC_HUMAN	54 kDa	L	2/2	S	2/2	5/3	5/3
Von Willebrand factor	VWF_HUMAN	309 kDa		A			1/1	
Zinc-alpha-2-glycoprotein	ZA2G_HUMAN	34 kDa	12/7	2/2	0	13/7	7/5	5/4

[a] Average of spectral counts in triplicate runs

[b] Average of number of unique peptides from triplicate runs

* Non-glycoproteins

L, A and S represents that the protein was only present in either disease-free or cancer serum for the LTA, AAL and SNA columns, respectively

.

The number of proteins identified in the bound fractions of LTA, AAL and SNA from the pooled disease-free sera were 76, 81 and 44 proteins, respectively, totaling 201 proteins of which 108 proteins were non-redundant proteins. The Venn diagram shown in Fig. 1 indicates the numbers of common and unique proteins identified from the pooled disease-free sera. The percentages of glycoproteins captured by each lectin column were 67%, 65% and 63% for LTA, AAL and SNA columns, respectively, which corresponded to 33%, 35% and 37% of non-glycoproteins, respectively. From the breast cancer serum, 82, 98 and 98 proteins were identified from the LTA, AAL and SNA columns, respectively, thus totaling 278 proteins of which 153 proteins were non-redundant proteins. The numbers of common and unique proteins from the breast cancer serum are shown in the Venn diagram in Fig. 2. The percentages of glycoproteins captured by each lectin column were 66%, 66% and 52% for LTA, AAL and SNA columns, respectively, which corresponded to 34%, 34% and 48% of non-glycoproteins, respectively. The number of common proteins for all the three-lectin columns was 48 in the case of cancer serum versus 29 common proteins for disease-free serum. There was an increase in the number of proteins identified in LTA (76 *vs.* 82), AAL (81 *vs.* 98) and SNA (44 *vs.* 98) columns when going from disease-free serum to cancer serum indicating that many proteins were altered in their expression in cancer serum as compared to disease-free sera. The percentages of glycoproteins identified in disease-free and cancer serum in LTA (67% *vs.* 66%) and in AAL (65% *vs.* 66%) remained similar, but in the case of SNA the glycoprotein percentage decreased greatly (63% *vs.* 52%). Although SNA column captured more proteins from cancer serum (54 proteins in excess), many of these proteins were non-glycoproteins. This may indicate that some of the non-glycoproteins captured by the SNA column interact with the glycoproteins that are

altered in the cancer serum. As mentioned in Chapter III, the non-specific binding of the non-glycoproteins might be due to protein-protein interaction.

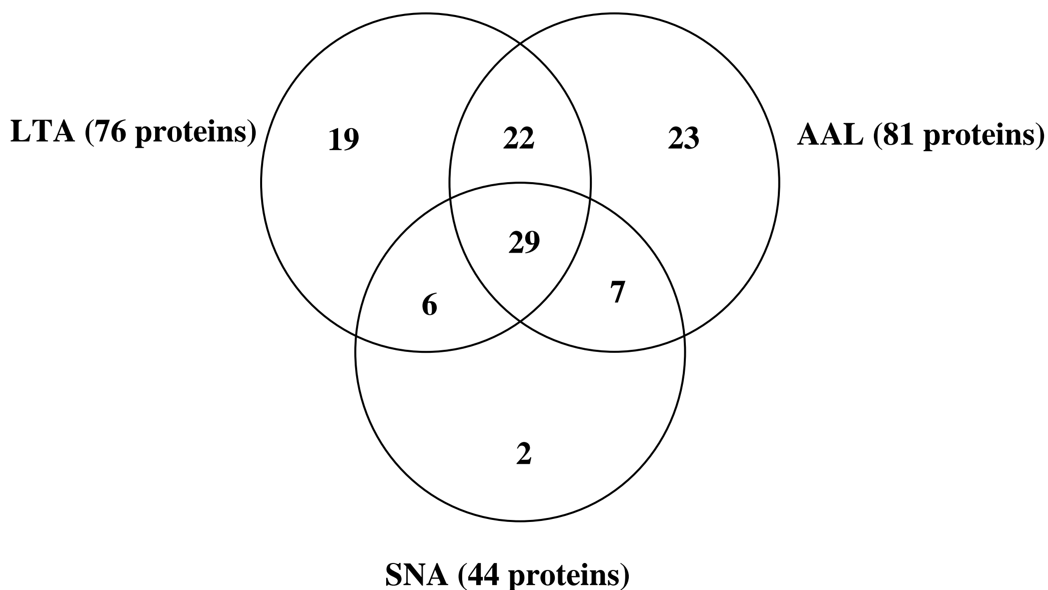


Figure 1. *Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the disease-free serum.*

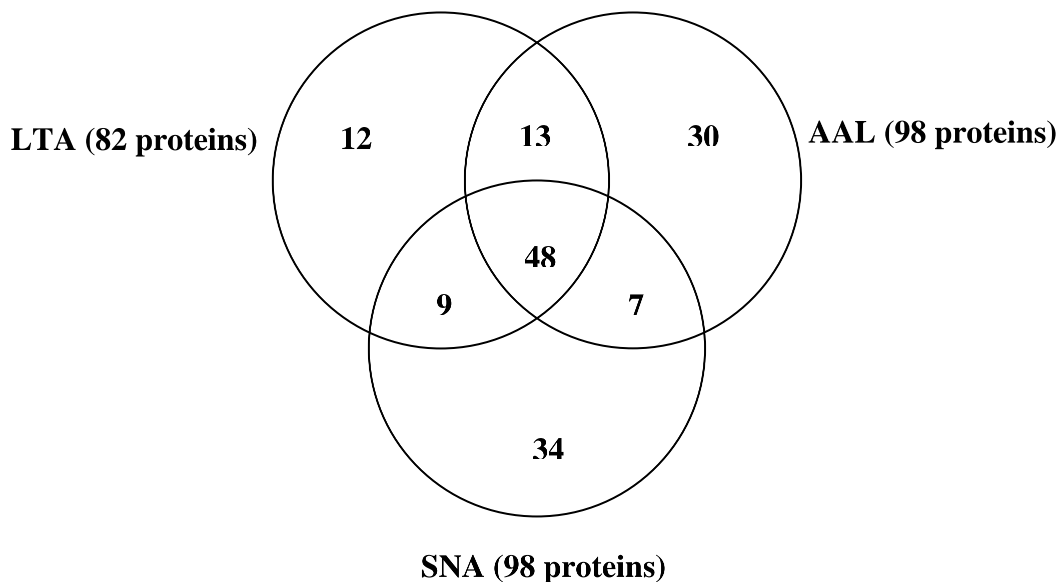


Figure 2. *Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the breast cancer serum.*

Identification of Differentially Expressed Proteins in the Lectin Fractions Using MS Spectral Count

The comparison between the regulation of proteins in disease-free serum and the cancer serum was made using spectral counts obtained from the LC-MS/MS analysis results. A statistical analysis was made using Student's t-test, and only those with p-values < 0.05 are listed in Tables 2, 3 and 4. Moreover, only those proteins with more than 4 average spectral counts were considered for the differential analysis. The numbers of differentially expressed proteins that were common between the columns and unique to each column are indicated in the Venn diagram shown in Fig.3. It can be seen that the number of common proteins between LTA and AAL was 1, LTA and SNA was also 1 and there were 3 common proteins between AAL and SNA. Some of the differentially expressed proteins are discussed in the following sections.

Fibrinogens Fibrinogens are heavily sialylated *N*-linked glycoproteins and the fibrinogen γ chain is also core fucosylated [18]. Therefore, they should exhibit affinity towards SNA and AAL columns. Also, fibrinogen α, β and γ chains contain Le^x glycan types [14] which make them interact with the LTA column. In fact, fibrinogen α, β and γ chains were found to be over-expressed in all the three lectin columns. In a study done by Cho *et al*. [14], to identify the Le^x containing glycoproteins in breast serum plasma, it was observed that the fibrinogen α and β chains were altered in their concentration by more than 3-fold in breast cancer plasma as compared to the disease-free plasma. An elevated level of fibrinogen was found in 44% of the early stage and 22% of the advanced stage breast cancer patients [19]. Also, it has been suggested that plasma fibrinogen can be a useful marker for gastric cancer progression [20]. According to a recent article that has a compiled list of proteins [21] that are differentially

expressed in human cancer, the protein fibrinogen α chain has been listed as a cancer biomarker candidate that has more than 500 citations.

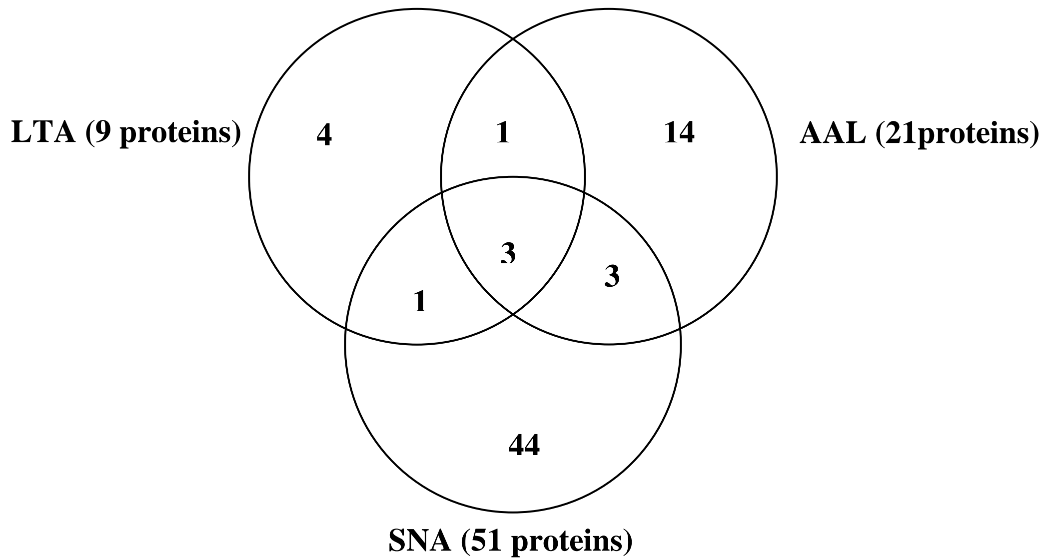


Figure 3. *Venn diagram for the differentially expressed proteins found in the lectin fractions.*

<u>Differentially expressed proteins captured by the LTA column</u> There were 19 and 12 unique proteins captured by the LTA column from the normal and the cancer serum, respectively (see Table 1). Some of these unique proteins were reported as cancer biomarker candidates, namely apolipoprotein C-I, neutrophil defensin 1 and serum paraoxonase/arylesterase 1 [21]. Apart from some of the fucosylated proteins, some of the Le^x determinant containing proteins such as plasminogen, kininogen-1, Ig gamma-2 chain C region, Ig gamma-3 chain C region, Ig mu chain C region, apolipoprotein E, vitronectin, clusterin and the fibrinogen chains were also identified in the LTA columns.

There were 9 differentially expressed proteins in cancer serum as compared to the disease-free serum in the LTA column fraction (see Table 2). Fibrinogen α, β and γ chain, hemoglobin subunit β, antithrombin-III, apolipoprotein E and Ig kappa chain V-III region VG

(Fragment) were over-expressed while afamin and serum amyloid P-component were down-expressed in the LTA fraction of the cancer serum. Antithrombin-III, which is an important serine protease inhibitor in the plasma, was found to be over-expressed by more than two-fold. According to a recent report [22], it was observed that thrombin-antithrombin complex was significantly over-expressed in breast cancer plasma and the levels of this complex significantly correlated with the levels of CA 15-3, which is an FDA approved cancer biomarker.

Hemoglobin subunit β, which was over-expressed in both LTA and the SNA column fractions, is known for its involvement in oxygen transport from the lung to the various peripheral tissues [23]. The high level of β subunit of hemoglobin found in the current study is in good agreement with the results reported in ref. [24], where SELDI-TOF analysis was done to compare the ductal carcinoma *in situ* (DCIS) type breast cancer serum with the disease-free serum.

TABLE 2

PROTEINS THAT WERE DIFFERENTIALLY EXPRESSED IN LTA FRACTION

Identified Proteins	Average spectral count		Ratio	
	Disease-free	Cancer	Cancer/ Disease-free	Disease-free /Cancer
Afamin	4	0	L	H
Antithrombin-III	4	8	2.27	0.44
Apolipoprotein E	0	6	H	L
Fibrinogen alpha chain	0	14	H	L
Fibrinogen beta chain	0	22	H	L
Fibrinogen gamma chain	0	25	H	L
Hemoglobin subunit beta	5	39	8.29	0.12
Ig kappa chain V-III region VG (Fragment)	0	4	H	L
Serum amyloid P-component	13	5	0.35	2.86

H and L represents that the protein is either present in high or low amount, either in disease-free or cancer serum.

Differentially expressed proteins captured by the AAL column There were 23 and 30 unique proteins (see Table 1) captured by the AAL column from the normal and the cancer serum, respectively. 15 of these unique proteins were found to be common to both normal and the cancer serum. Some of these unique proteins that were identified in both control and cancer such as attractin, complement C1r subcomponent-like protein, Ig alpha-2 chain C region, insulin like growth factor-binding protein 3, kallistatin, sulfhydryl oxidase 1, vitronectin, IgGFc-binding protein, von Willebrand factor and scavenger receptor cysteine-rich type 1 protein M130 are known to have core fucosylation [18].

There were 16 and 5 proteins that were found to be over-expressed and down-expressed, respectively, in the cancer serum relative to the disease-free serum in the AAL fraction (see Table 3). According to the findings reported in a recent study [14], some of the core fucosylated glycoproteins that include attractin, complement factor H and scavenger receptor cysteine-rich type 1 protein M130 were identified only in the hepatocellular carcinoma (HCC) serum, indicating that these proteins were present at elevated level in the HCC serum. In our current study also, these core fucosylated proteins were found at elevated levels except for attractin. The protein complement C3, which was elevated in both AAL and SNA fractions may indicate that there were changes in both fucosylation and sialylation of the protein. This finding corroborates with that observed in another study [25] as far as elevated levels of sialylation and fucosylation of complement C3 are concerned. Also, complement C3 was identified as a potential marker of colorectal cancer [25].

TABLE 3

PROTEINS THAT WERE DIFFERENTIALLY EXPRESSED IN AAL FRACTION

Identified Proteins	Average spectral count		Ratio	
	Disease-free	Cancer	Cancer/ Disease-free	Disease-free /Cancer
Alpha-1-acid glycoprotein 1	26	10	0.40	2.48
Alpha-1-acid glycoprotein 2	21	8	0.39	2.56
Apolipoprotein L1	0	4	H	L
Attractin	6	0	L	H
Complement C3	12	64	5.36	0.19
Complement C4-B	13	41	3.13	0.32
Complement factor H	5	13	2.53	0.39
Fibrinogen alpha chain	0	9	H	L
Fibrinogen beta chain	0	14	H	L
Fibrinogen gamma chain	0	11	H	L
Haptoglobin	89	30	0.34	2.98
Haptoglobin-related protein*	41	11	0.28	3.62
Ig gamma-4 chain C region	0	17	H	L
Keratin, type I cytoskeletal 14*	9	44	5.08	0.20
Keratin, type I cytoskeletal 16*	11	73	6.47	0.15
Keratin, type I cytoskeletal 17*	0	30	H	L
Keratin, type II cytoskeletal 5*	8	35	4.42	0.23
Keratin, type II cytoskeletal 6B*	0	73	H	L
Keratin, type II cytoskeletal 6C*	0	50	H	L
Scavenger receptor cysteine-rich type 1 protein M130	0	7	H	L
Serum amyloid P-component	20	0	L	H

*Non-glycoproteins
H and L represents that the protein is either present in high or low amount, either in disease-free or cancer serum.

Differentially expressed proteins captured by the SNA column Fifty one differentially expressed proteins were captured by the SNA column. Some of the sialylated Le^x glycan containing proteins such as Ig gamma-2 chain C region, apolipoprotein A-I, fibrinogen α, β and γ chains, inter-alpha-trypsin inhibitor heavy chain H4 and vitronectin [14] were differentially

expressed in the fraction obtained from the SNA column. With the exception of fibrinogen α and β chains, all these proteins have been shown previously to contain elevated levels of sialylated Le^x glycan in breast cancer plasma [14]. Also, a recent report states that higher levels of sialylated Le^x glycans were observed in the breast cancer sera, which might be a indicator of metastasis of cancer [26]. Some of the over-expressed proteins such as fibrinogen α chain, desmin and α-2-macroglobulin, have been listed as candidate cancer biomarkers with more than 500 citations in a list that has compiled the differentially expressed proteins in human cancer [21]. Also, it has been reported that kininogen-1 and complement C3 show changes in sialylation in breast cancer serum [27], which is in agreement with the findings reported in this study where they were found to be elevated in the SNA column fraction indicating the changes in sialylation. Also, in the current study, the protein plasma kallikrein was elevated in the cancer serum which agreed with the report that plasma kallikrein was at elevated levels in 25 out of 28 lung adenocarcinoma patients [28]. The protein haptoglobin did not show any significant differences in the spectral count in either the disease-free or the cancer serum of the SNA fraction. It should be noted that the number of unique peptides for haptoglobin in both disease-free and cancer serum was 20. This indicates that the α2→6 sialic acid was unchanged in its composition in breast cancer serum. It has been reported that a glycoform of haptoglobin that contains both the α1→3 and α2→6 fucosylated was at elevated level in lung cancer [29], but it is not clear whether the alteration is specifically in the α1→3 or α2→6 fucosylation. The protein hemopexin is a fucosylated and sialylated protein [30]. Although hemopexin was captured by all three columns, it was found to be in elevated levels (3.2 times) only in the SNA column. This may indicate that there was an increase in sialylation of the protein but there was no increase in the fucosylation of the protein. Pyruvate kinase isozymes M1/M2 (tumor P2-MK) is an enzyme

which is known to be in high levels in tumor tissue and body fluids [31]. It has been reported to be at high levels in ductal invasive breast cancer sera [10]. In the current study also, it was found at elevated levels in the cancer serum.

TABLE 4

PROTEINS THAT WERE DIFFERENTIALLY EXPRESSED IN SNA FRACTION

Identified Proteins	Average spectral count		Ratio	
	Disease-free	Cancer	Cancer/ Disease-free	Disease-free /Cancer
Actin, alpha skeletal muscle*	0	61	H	L
Alpha-2-macroglobulin	16	40	2.50	0.40
Alpha-actinin-2*	0	11	H	L
Alpha-actinin-3*	0	14	H	L
Apolipoprotein A-I	0	9	H	L
Apolipoprotein B-100	10	65	6.29	0.16
Beta-enolase*	0	13	H	L
Ceruloplasmin	5	21	4.43	0.23
Complement C3	0	35	H	L
Complement C4-A	14	39	2.72	0.37
Complement factor B	7	14	2.15	0.47
Complement factor H	19	66	3.46	0.29
Complement factor H-related protein 1	0	9	H	L
Creatine kinase M-type*	0	19	H	L
Desmin*	0	5	H	L
Fibrinogen alpha chain	0	16	H	L
Fibrinogen beta chain	0	16	H	L
Fibrinogen gamma chain	0	21	H	L
Filamin-C*	0	5	H	L
Fructose-bisphosphate aldolase A*	0	17	H	L
Glyceraldehyde-3-phosphate dehydrogenase*	0	6	H	L
Glycogen phosphorylase, muscle form*	0	11	H	L
Hemoglobin subunit alpha	0	6	H	L
Hemoglobin subunit beta	0	10	H	L

Hemopexin	6	20	3.21	0.31
Ig gamma-1 chain C region	5	20	3.69	0.27
Ig gamma-2 chain C region	6	15	2.32	0.43
Ig gamma-3 chain C region	5	16	3.27	0.31
Ig kappa chain C region*	7	19	2.85	0.35
Inter-alpha-trypsin inhibitor heavy chain H1	0	4	H	L
Inter-alpha-trypsin inhibitor heavy chain H2	0	14	H	L
Inter-alpha-trypsin inhibitor heavy chain H4	4	11	3.09	0.32
Keratin, type I cytoskeletal 14*	21	10	0.47	2.14
Kininogen-1	0	9	H	L
L-lactate dehydrogenase A chain*	0	5	H	L
Myosin light chain 1/3, skeletal muscle isoform*	0	11	H	L
Myosin regulatory light chain 2, skeletal muscle isoform*	0	12	H	L
Myosin-1*	0	148	H	L
Myosin-2*	0	118	H	L
Myosin-4*	0	89	H	L
Myosin-8*	0	86	H	L
Plasma kallikrein	5	13	2.44	0.41
Pyruvate kinase isozymes M1/M2*	0	11	H	L
Sarcoplasmic/endoplasmic reticulum calcium ATPase 1*	0	8	H	L
Serum albumin*	13	42	3.18	0.31
Titin*	0	9	H	L
Triosephosphate isomerase*	0	5	H	L
Tropomyosin alpha-1 chain*	0	21	H	L
Tropomyosin beta chain*	0	31	H	L
Vitronectin	0	5	H	L
Zinc-alpha-2-glycoprotein	0	5	H	L

*Non-glycoproteins
H and L represents that the protein is either present in high or low amount, either in disease-free or cancer serum.

Narrowing on Candidate Biomarkers – A Panel of Biomarkers

By comparing the results of the narrow selectivity lectins series (LTA→AAL→SNA or LAS series) to those of the broad selectivity lectins series (WGA→Con A→RCA or WCR series, see Chapter III), one can readily find out that 13 proteins were either up- or down-regulated in both of the studies (see Table 5). This perhaps establishes a "panel" of candidate biomarkers,

which in principle should provide greater sensitivity and accuracy than any of the markers used alone as pointed out by Xiao *et al.* [32] in their recent study of lung cancer related proteins. Out of these 13 proteins, 5 were oppositely regulated in both studies including afamin, complement factor B, haptoglobin, Ig kappa chain C region and Keratin, type I cytoskeletal 14. With the exception of albumin, keratin, type I cytoskeletal 14 and Ig kappa chain C region, which are not glycoproteins, the fact that different lectins with different glycan selectivity were involved in the capturing the other 10 glycoproteins would suggest that different glycoforms of the same glycoproteins underwent alteration in their glycosylation in breast cancer. More meaningful in this regard are the glycoproteins that were oppositely regulated in the two series. For illustration, alpha-1-acid glycoprotein which contains more of the sialylated tri- and tetraantennary than the bi-antennary *N*-glycans with outer arm fucosylation [33] in disease-free serum, was down regulated in the fraction of RCA and AAL may be due to a decrease in branching during breast cancer progression [34], an event that results from incompletion of the glycosylation process that normally leads to elevation in high mannose type glycans. In this decrease of branching, fucosylation in outer arms is lost, thus decreasing the affinity of AAL towards the altered glycoprotein. Another illustration of alteration in glycosylation can be provided by the case of haptoglobin, which is a sialylated, fucosylated glycoprotein having triantennary glycans in disease-free serum [35, 36]. In breast cancer serum, the glycosylation of this glycoprotein is also altered in the sense that branching is decreased resulting in more high mannose glycans, thus increasing its binding to Con A, and at the same time decreasing its binding to AAL, which correspond to up-regulation of the protein in the WCR series and down-regulation in the LAS series, respectively.

TABLE 5

LIST OF COMMON PROTEINS THAT WERE DIFFERENTIALLY EXPRESSED IN BOTH BROAD AND NARROW SPECIFICITY LECTINS

Identified proteins	Regulation in cancer serum (WGA→Con A→RCA series)	Regulation in cancer serum (LTA→AAL→SNA series)
Afamin#	Up (Con A)	Down (LTA)
Alpha-1-acid glycoprotein 1	Down (RCA)	Down (AAL)
Complement factor B#	Down (RCA)	Up (SNA)
Haptoglobin#	Up (Con A)	Down (AAL)
Hemoglobin subunit alpha	Up (WGA, Con A)	Up (SNA)
Hemoglobin subunit beta	Up (WGA)	Up (SNA, LTA)
Hemopexin	Up (RCA)	Up (SNA)
Ig gamma-2 chain C region	Up (Con A)	Up (SNA)
Ig kappa chain C region#*	Down (WGA)	Up (SNA)
Inter-alpha-trypsin inhibitor heavy chain H1#	Down (WGA)	Up (SNA)
Keratin, type I cytoskeletal 14*	Down (RCA)	Up (AAL), Down (SNA)
Serum albumin*	Up (Con A, RCA)	Up (SNA)
Zinc-alpha-2-glycoprotein	Up (RCA)	Up (SNA)

Oppositely regulated proteins
* Non-glycoproteins

In summary, the combination of the results from a lectins series of broad specificity with those obtained from a lectins series of narrow specificity narrow down the number of altered glycoproteins to a more representative panel of protein biomarkers. This may contribute to facilitating the development of future panels of candidate biomarkers that may find effective use in clinical laboratories for detecting and following the progression of breast cancer and other diseases.

Conclusions

Even though further studies have to be done to validate the differentially expressed proteins, the glycoproteomic approach provided here has been demonstrated as a useful method to reduce the complexity of the serum and to identify the fucosylated and sialylated glycoproteins that show alteration in their glycosylation pattern in breast cancer serum. The strategy investigated here would be a proof-of-concept for an efficient tool in analyzing targeted glycoproteins.

References

[1] www.breastcancer.org

[2] T. F. Orntoft and E. M. Vestergaard, *Electrophoresis* **1999**, *20*, 362-371.

[3] Z. Kyselova, Y. Mechref, P. Kang, J. A. Goetz, L. E. Dobrolecki, G. W. Sledge, L. Schnaper, R. J. Hickey, L. H. Malkas and M. V. Novotny, *Clin. Chem.* **2008**, *54*, 1166-1175.

[4] R. Saldova, M. R. Wormald, R. A. Dwek and P. M. Rudd, *Dis. Markers* **2008**, *25*, 219-232.

[5] G. Tabares, C. M. Radcliffe, S. Barrabes, M. Ramirez, R. N. Aleixandre, W. Hoesel, R. A. Dwek, P. M. Rudd, R. Peracaula and R. de Llorens, *Glycobiology* **2006**, *16*, 132-145.

[6] L. Warren, J. P. Fuhrer and C. A. Buck, *Proc. Natl. Acad. Sci. U S A* **1972**, *69*, 1838-1842.

[7] S. Hakomori and R. Kannagi, *J. Natl. Cancer Inst.* **1983**, *71*, 231-251.

[8] S. Thompson, B. Cantwell, K. Matta and G. A. Turner, *Glycoconj. J.* **1991**, *8*, 180.

[9] P. D. Rye and R. A. Walker, *Eur. J. Cancer Clin. Oncol.* **1989**, *25*, 65-72.

[10] K. L. Abbott, K. Aoki, J. M. Lim, M. Porterfield, R. Johnson, R. M. O'Regan, L. Wells, M. Tiemeyer and M. Pierce, *J. Proteome Res.* **2008**, *7*, 1470-1480.

[11] J. Zhao, D. M. Simeone, D. Heidt, M. A. Anderson and D. M. Lubman, *J. Proteome Res.* **2006**, *5*, 1792-1802.

[12] M. A. Comunale, M. Wang, J. Hafner, J. Krakover, L. Rodemich, B. Kopenhaver, R. E. Long, O. Junaidi, A. M. Bisceglie, T. M. Block and A. S. Mehta, *J. Proteome Res.* **2009**, *8*, 595-602.

[13] B. Mann, M. Madera, I. Klouckova, Y. Mechref, L. E. Dobrolecki, R. J. Hickey, Z. T. Hammoud and M. V. Novotny, *Electrophoresis* **2010**, *31*, 1833-1841.

[14] W. Cho, K. Jung and F. E. Regnier, *J. Proteome Res.* **2010**, *9*, 5960-5968.

[15] A. Kobata and K. Yamashita Glycobiology A Practical Approach, *Vol.* (Eds.): M. Fukuda and A. Kobata, IRL Press, Oxford, **1993**, pp. 103-125.

[16] L. Yan, P. P. Wilkins, G. Alvarez-Manilla, S. I. Do, D. F. Smith and R. D. Cummings, *Glycoconj. J.* **1997**, *14*, 45-55.

[17] M. Zhou, D. A. Lucas, K. C. Chan, H. J. Issaq, E. F. Petricoin, 3rd, L. A. Liotta, T. D. Veenstra and T. P. Conrads, *Electrophoresis* **2004**, *25*, 1289-1298.

[18] W. Jia, Z. Lu, Y. Fu, H. P. Wang, L. H. Wang, H. Chi, Z. F. Yuan, Z. B. Zheng, L. N. Song, H. H. Han, Y. M. Liang, J. L. Wang, Y. Cai, Y. K. Zhang, Y. L. Deng, W. T. Ying, S. M. He and X. H. Qian, *Mol. Cell. Proteomics* **2009**, *8*, 913-923.

[19] W. P. Mielicki, M. Tenderenda, P. Rutkowski and K. Chojnowski, *Cancer Lett.* **1999**, *146*, 61-66.

[20] J. H. Lee, K. W. Ryu, S. Kim and J. M. Bae, *Hepatogastroenterology* **2004**, *51*, 1860-1863.

[21] M. Polanski and N. L. Anderson, *Biomarker Insights* **2006**, *2*, 1-48.

[22] O. Ozyilkan, E. Baltali, O. Ozdemir, G. Tekuzman, S. Kirazli and D. Firat, *Tumori* **1998**, *84*, 364-367.

[23] D. Ianzer, K. Konno, C. H. Xavier, R. Stocklin, R. A. Santos, A. C. de Camargo and D. C. Pimenta, *Peptides* **2006**, *27*, 2957-2966.

[24] J. Solassol, P. Rouanet, P. J. Lamy, C. Allal, G. Favre, T. Maudelonde and A. Mange, *Oncogene* **2010**, *29*, 550-560.

[25] Y. Qiu, T. H. Patwa, L. Xu, K. Shedden, D. E. Misek, M. Tuck, G. Jin, M. T. Ruffin, D. K. Turgeon, S. Synal, R. Bresalier, N. Marcon, D. E. Brenner and D. M. Lubman, *J. Proteome Res.* **2008**, *7*, 1693-1703.

[26] T. Fujita, K. Murayama, T. Hanamura, T. Okada, T. Ito, M. Harada, A. Komatsu, H. Koyama, T. Kanai, K. Maeno, Y. Mochizuki, Y. Hama, K.-i. Ito, J. Amano and M. Fujimori, *Jpn. J. Clin. Oncol.* **2011**, *41*, 394-399.

[27] Z. Zeng, M. Hincapie, B. B. Haab, S. Hanash, S. J. Pitteri, S. Kluck, J. M. Hogan, J. Kennedy and W. S. Hancock, *J. Chromatogr. A* **2010**, *1217*, 3307-3315.

[28] S. H. Heo, S. J. Lee, H. M. Ryoo, J. Y. Park and J. Y. Cho, *Proteomics* **2007**, *7*, 4292-4302.

[29] H. Y. Tsai, K. Boonyapranai, S. Sriyam, C. J. Yu, S. W. Wu, K. H. Khoo, S. Phutrakul and S. T. Chen, *Proteomics* **2011**, *11*, 2162-2170.

[30] J. Zhao, T. H. Patwa, W. Qiu, K. Shedden, R. Hinderer, D. E. Misek, M. A. Anderson, D. M. Simeone and D. M. Lubman, *J Proteome Res* **2007**, *6*, 1864-1874.

[31] H. R. Hathurusinghe, K. S. Goonetilleke and A. K. Siriwardena, *Ann. Surg. Oncol.* **2007**, *14*, 2714-2720.

[32] T. Xiao, W. Ying, L. Li, Z. Hu, Y. Ma, L. Jiao, J. Ma, Y. Cai, D. Lin, S. Guo, N. Han, X. Di, M. Li, D. Zhang, K. Su, J. Yuan, H. Zheng, M. Gao, J. He, S. Shi, W. Li, N.

Xu, H. Zhang, Y. Liu, K. Zhang, Y. Gao, X. Qian and S. Cheng, *Mol. Cell. Proteomics* **2005**, *4*, 1480-1486.

[33] K. Higai, Y. Aoki, Y. Azuma and K. Matsumoto, *Biochim. Biophys. Acta* **2005**, *1725*, 128-135.

[34] M. L. A. de Leoz, L. J. T. Young, H. J. An, S. R. Kronewitter, J. Kim, S. Miyamoto, A. D. Borowsky, H. K. Chew and C. B. Lebrilla, *Molecular & Cellular Proteomics* **2011**, *10*.

[35] Z. He, L. P. Aristoteli, L. Kritharides and B. Garner, *Biochem. Biophys. Res. Commun.* **2006**, *343*, 496-503.

[36] S. Zhang, H. Shu, K. Luo, X. Kang, Y. Zhang, H. Lu and Y. Liu, *Mol. Biosyst.* **2011**, *7*, 1621-1628.

CHAPTER V

TANDEM LECTIN AFFINITY CHROMATOGRAPHY FOR IN-DEPTH GLYCOPROTEOMIC ANALYSIS BY COMBINING NARROW AND BROAD SPECIFICITY LECTINS

Introduction

The analysis of serum glycoproteome is a clinically important and necessary task. It is one of the major human serum sub-proteomes. Among more than 100 post-translational modifications (PTMs) known so far, glycosylation is one of the most important PTM [1]. In fact, about 50% of the plasma proteins are glycosylated [2]. Serum glycoproteome is a highly complex mixture. The complexity arises from the complex microheterogeneity of glycoproteins that result from different glycosylated variants or glycoforms of glycoproteins. Analysis of this largely complex glycoproteome requires separation approaches that can separate a wide range of glycoforms. To achieve this, six lectins namely, *Lotus tetragonolobus* agglutinin (LTA), *Aleuria aurantia* lectin (AAL), *Sambucus nigra* (SNA), wheat germ agglutinin (WGA), concanavalin A (Con A) and *Ricinus communis* agglutinin-I (RCA-I) were used in this study. LTA, AAL and SNA have narrow specificity towards glycoproteins (see Chapter IV Introduction) and WGA, Con A and RCA-I have relatively broader specificity towards glycoproteins (see

Chapter III Introduction) as compared to the former set of lectins.

Apart from the microheterogeneity of glycoproteins, there exists another complexity arising from the wide dynamic concentration range of serum proteins. Depletion of high abundance proteins using immunoaffinity methods has been a common approach to reduce the concentration range [3]. But it has been reported that by depleting the high abundance proteins many other valuable proteins are also co-depleted [4]. Thus, in this study serum was used as is without any prior depletion.

Although an initial study by Madera *et al* [5] has investigated the potentials of combining broad and narrow specificity lectins in glycoproteomics, the lectins studied were limited to only four immobilized lectins namely Con A, SNA, *Ulex europaaeus* agglutinin-I (UEA-I) and *Phaseolus vulgaris* agglutinin-L (PHA-L). Due to its excessive nonspecific interactions, the macroporous silica support material used in the immobilization of the four lectins led to enriching glycoproteins as well as non-glycosylated proteins. In fact, 54 proteins were non glycosylated out of the 108 proteins that were captured that is 50% were non glycosylated. This is a strong indication of nonspecific interactions with the support matrix. In order to alleviate this drawback and also provide a more in-depth glycoproteomics, the investigation in this Chapter is concerned with 6 immobilized lectins on a novel monolithic matrix with much less nonspecific interactions and whose specificity spans a much wider range.

Experimental

The experimental design was the same as in previous Chapters:

- The instruments used were the same as those described in Chapter III

- Reagents are the same as in Chapters III and IV.
- The monolithic columns were prepared as described in Chapter IV.
- The immobilization of lectins was the same as described in Chapters III and IV.
- The protein assay procedure was same as that described in Chapter III.
- The LC-MS/MS conditions were same as in Chapter IV.
- LC-MS/MS data analysis conditions were same as in Chapter III, except for the parent ion tolerance, which was 10 ppm.

- Fractionation of Human Serum Glycoproteins - Chromatographic Conditions

The LTA→AAL→SNA series involves the use of Ca^{2+} ions in the binding mobile phase which is 20 mM Tris containing 100 mM NaCl at pH 6.0, whereas the WGA→Con A→RCA series involves the use of Ca^{2+}, Mg^{2+} and Mn^{2+} ions in the same binding mobile phase. Since this study combines both lectin column tandem series, the binding mobile phase contained Ca^{2+}, Mg^{2+} and Mn^{2+} ions so that it can be compatible with both lectin series. The chromatographic fractionation conditions were same as in Chapter IV. The lectin columns were arranged in tandem series, and each column was removed from the series and eluted individually.

Results and Discussion

Overall Strategy

The primary objective of the present study was to enrich a wide range of glycoforms of the glycoproteins present in human serum using serial lectin affinity chromatography. To achieve this, six lectin columns were arranged in tandem series. Out of the six-lectin columns, three columns have narrow specificity and the other three

have broad specificity towards the glycoproteins. The narrow specificity columns (hereafter referred to as LAS columns) include the LTA, AAL and SNA columns, whereas the broad specificity columns (hereafter designated as WCR columns) included the WGA, Con A and RCA-I. The tandem columns were arranged in the following order: LTA→AAL→SNA→WGA→Con A→RCA-I. In this arrangement, the LAS columns were followed by the WCR columns so that the pass through glycoproteins from the narrow selectivity columns would be captured by the broad specificity columns. This arrangement should facilitate the selective capturing of (i) the fucosylated glycoproteins by LTA and AAL columns, (ii) the sialylated glycoproteins by SNA and WGA columns, (iii) the high-mannose glycoproteins by the Con A column, (iv) the biantennary bisected hybrid types by the WGA column and (iv) the tri- and tetraantennary branched glycoproteins that pass through all the preceding columns by the RCA-I column.

Analysis of the Proteins Captured by the Lectin Columns

As mentioned in the Experimental section, the fractions that were captured and subsequently eluted from each of the lectin columns were dialyzed against water, concentrated and an aliquot of each lectin fraction was analyzed by LC-MS/MS. Only the proteins that exhibited identification probability greater than 99% with peptide identification probability greater than 95% containing at least two unique peptides were considered and they are reported in Table 1. The number of proteins identified in the fractions from LTA, AAL, SNA columns was 79, 98 and 60, respectively, and that totaled 237 proteins. The number of proteins identified in WGA, Con A and RCA-I

TABLE 1

LIST OF PROTEINS IDENTIFIED IN THE LECTIN-BOUND FRACTIONS

Identified Proteins	Accession Number	Mol. Wt.	Spectral count						Lectin(s) column(s) that retained the protein[a]
			LTA	AAL	SNA	WGA	Con A	RCA-I	
Adiponectin	ADIPO_HUMAN	26 kDa	0	0	0	11	0	0	W
Afamin	AFAM_HUMAN	69 kDa	0	0	0	5	144	20	WCR
Alpha-1-acid glycoprotein 1	A1AG1_HUMAN	24 kDa	15	80	4	141	33	43	LASWCR
Alpha-1-acid glycoprotein 2	A1AG2_HUMAN	24 kDa	12	50	0	112	21	26	LAWCR
Alpha-1-antichymotrypsin	AACT_HUMAN	48 kDa	10	18	2	91	44	39	LASWCR
Alpha-1-antitrypsin	A1AT_HUMAN	47 kDa	57	31	11	62	190	79	LASWCR
Alpha-1B-glycoprotein	A1BG_HUMAN	54 kDa	23	0	3	0	115	23	LSCR
Alpha-2-antiplasmin	A2AP_HUMAN	55 kDa	0	0	0	14	30	14	WCR
Alpha-2-HS-glycoprotein	FETUA_HUMAN	39 kDa	12	7	0	66	8	14	LAWCR
Alpha-2-macroglobulin	A2MG_HUMAN	163 kDa	106	238	36	713	746	960	LASWCR
Aminopeptidase N	AMPN_HUMAN	110 kDa	0	0	0	5	0	0	W
Angiotensinogen	ANGT_HUMAN	53 kDa	10	6	0	19	61	26	LAWCR
Antithrombin-III	ANT3_HUMAN	53 kDa	8	0	0	0	59	0	LC
Apolipoprotein A-I	APOA1_HUMAN	31 kDa	44	29	7	81	5	99	LASWCR
Apolipoprotein A-II*	APOA2_HUMAN	11 kDa	7	5	2	11	0	15	LASWR
Apolipoprotein A-IV*	APOA4_HUMAN	45 kDa	9	0	0	0	0	0	L
Apolipoprotein B-100	APOB_HUMAN	516 kDa	80	40	56	161	0	242	LASWR
Apolipoprotein C-II*	APOC2_HUMAN	11 kDa	3	0	0	0	0	0	L
Apolipoprotein C-III	APOC3_HUMAN	11 kDa	3	5	0	0	0	0	LA
Apolipoprotein D	APOD_HUMAN	21 kDa	13	16	4	17	16	7	LASWCR

Apolipoprotein E	APOE_HUMAN	36 kDa	5	9	3	0	0	0	LAS
Apolipoprotein L1	APOL1_HUMAN	44 kDa	0	0	0	6	0	37	WR
Apolipoprotein(a)	APOA_HUMAN	501 kDa	6	2	3	21	0	0	LASW
Attractin	ATRN_HUMAN	159 kDa	0	36	0	166	176	41	AWCR
Beta-2-glycoprotein 1	APOH_HUMAN	38 kDa	18	25	0	9	110	54	LAWCR
Beta-Ala-His dipeptidase	CNDP1_HUMAN	57 kDa	0	0	0	14	0	0	W
Biotinidase	BTD_HUMAN	61 kDa	0	0	0	9	4	0	WC
C4b-binding protein alpha chain	C4BPA_HUMAN	67 kDa	3	3	0	0	0	0	LA
Carboxypeptidase B2	CBPB2_HUMAN	48 kDa	0	0	0	4	0	0	W
Carboxypeptidase N catalytic chain	CBPN_HUMAN	52 kDa	0	7	0	35	0	0	AW
Carboxypeptidase N subunit 2	CPN2_HUMAN	61 kDa	0	9	0	49	0	27	AWR
CD44 antigen	CD44_HUMAN	82 kDa	0	8	0	15	0	0	AW
CD5 antigen-like*	CD5L_HUMAN	38 kDa	13	31	5	29	15	129	LASWCR
Ceruloplasmin	CERU_HUMAN	122 kDa	39	75	13	47	467	55	LASWCR
Cholinesterase	CHLE_HUMAN	68 kDa	0	5	0	70	43	20	AWCR
Clusterin	CLUS_HUMAN	52 kDa	11	12	0	16	0	0	LAW
Coagulation factor V	FA5_HUMAN	252 kDa	0	3	0	0	0	0	A
Coagulation factor X	FA10_HUMAN	55 kDa	0	0	0	32	0	0	W
Coagulation factor XII	FA12_HUMAN	68 kDa	0	0	0	145	0	0	W
Coagulation factor XIII B chain	F13B_HUMAN	76 kDa	0	4	0	0	0	0	A
Complement C1q subcomponent subunit A	C1QA_HUMAN	26 kDa	0	0	0	15	0	0	W
Complement C1q subcomponent subunit B	C1QB_HUMAN	27 kDa	5	7	2	31	0	3	LASWR
Complement C1q subcomponent subunit C	C1QC_HUMAN	26 kDa	11	4	7	24	0	10	LASWR
Complement C1r	C1R_HUMAN	80 kDa	8	10	2	93	0	14	LASWR

subcomponent									
Complement C1r subcomponent-like protein	C1RL_HUMAN	53 kDa	0	0	0	16	0	8	WR
Complement C1s subcomponent	C1S_HUMAN	77 kDa	5	4	4	47	0	13	LASWR
Complement C2	CO2_HUMAN	83 kDa	0	0	0	0	66	10	CR
Complement C3	CO3_HUMAN	187 kDa	135	11	3	36	405	33	LASWCR
Complement C4-B	CO4B_HUMAN	193 kDa	46	16	20	89	50	141	LASWCR
Complement C5	CO5_HUMAN	188 kDa	4	0	0	0	0	5	LR
Complement component C6	CO6_HUMAN	105 kDa	0	0	4	0	0	0	S
Complement component C7	CO7_HUMAN	94 kDa	0	0	0	0	5	0	C
Complement factor B	CFAB_HUMAN	86 kDa	16	0	10	0	35	31	LSCR
Complement factor H	CFAH_HUMAN	139 kDa	24	22	44	173	212	327	LASWCR
Complement factor H-related protein 1	FHR1_HUMAN	38 kDa	0	0	9	0	52	0	SC
Complement factor H-related protein 3	FHR3_HUMAN	37 kDa	0	5	0	0	0	0	A
Complement factor I	CFAI_HUMAN	66 kDa	0	0	0	35	37	54	WCR
Corticosteroid-binding globulin	CBG_HUMAN	45 kDa	0	6	0	25	50	14	AWCR
Cysteine-rich secretory protein 3	CRIS3_HUMAN	28 kDa	0	0	0	0	11	0	C
Desmoplakin*	DESP_HUMAN	332 kDa	0	0	7	0	0	0	S
Dopamine beta-hydroxylase	DOPO_HUMAN	69 kDa	0	0	0	0	29	0	C
Extracellular matrix protein 1	ECM1_HUMAN	61 kDa	0	0	0	29	0	0	W
Fibrinogen alpha chain	FIBA_HUMAN	95 kDa	3	0	0	0	0	0	L
Fibronectin	FINC_HUMAN	263 kDa	5	0	0	0	0	0	L
Ficolin-3	FCN3_HUMAN	33 kDa	0	18	0	74	0	12	AWR
Gamma-glutamyl hydrolase	GGH_HUMAN	36 kDa	0	0	0	0	10	0	C
Haptoglobin	HPT_HUMAN	45 kDa	73	91	60	521	345	426	LASWCR

Haptoglobin-related protein*	HPTR_HUMAN	39 kDa	33	48	0	242	0	215	LAWR
Hemoglobin subunit alpha	HBA_HUMAN	15 kDa	0	9	0	63	59	74	AWCR
Hemoglobin subunit beta	HBB_HUMAN	16 kDa	3	4	0	93	46	70	LAWCR
Hemopexin	HEMO_HUMAN	52 kDa	39	32	9	512	345	176	LASWCR
Heparin cofactor 2	HEP2_HUMAN	57 kDa	0	0	2	0	0	26	SR
Hepatocyte growth factor activator	HGFA_HUMAN	71 kDa	0	0	0	50	0	0	W
Histidine-rich glycoprotein	HRG_HUMAN	60 kDa	17	15	6	63	10	17	LASWCR
Hornerin*	HORN_HUMAN	282 kDa	0	0	7	0	0	0	S
Ig alpha-1 chain C region	IGHA1_HUMAN	38 kDa	25	47	18	100	121	167	LASWCR
Ig alpha-2 chain C region	IGHA2_HUMAN	37 kDa	0	44	0	0	105	128	ACR
Ig delta chain C region	IGHD_HUMAN	42 kDa	0	0	0	31	0	0	W
Ig gamma-1 chain C region	IGHG1_HUMAN	36 kDa	51	33	20	32	99	153	LASWCR
Ig gamma-2 chain C region	IGHG2_HUMAN	36 kDa	40	23	16	13	54	102	LASWCR
Ig gamma-3 chain C region	IGHG3_HUMAN	41 kDa	34	17	14	0	46	94	LASCR
Ig gamma-4 chain C region	IGHG4_HUMAN	36 kDa	22	14	0	0	42	71	LACR
Ig heavy chain V-I region V35*	HV103_HUMAN	13 kDa	0	4	0	0	0	11	AR
Ig heavy chain V-II region ARH-77*	HV209_HUMAN	16 kDa	0	0	0	0	0	18	R
Ig heavy chain V-III region BUT*	HV306_HUMAN	12 kDa	0	10	0	7	8	47	AWCR
Ig heavy chain V-III region GA*	HV308_HUMAN	13 kDa	0	0	0	0	0	11	R
Ig heavy chain V-III region GAL*	HV320_HUMAN	13 kDa	0	0	0	0	7	17	CR
Ig heavy chain V-III region VH26*	HV303_HUMAN	13 kDa	5	5	3	10	10	46	LASWCR
Ig kappa chain C region*	IGKC_HUMAN	12 kDa	27	38	15	60	79	151	LASWCR
Ig kappa chain V-I region	KV109_HUMAN	13 kDa	0	0	0	0	0	8	R

HK101 (Fragment)*									
Ig kappa chain V-I region HK102 (Fragment)*	KV110_HUMAN	13 kDa	0	5	0	0	4	18	ACR
Ig kappa chain V-III region B6*	KV301_HUMAN	12 kDa	0	0	0	0	0	9	R
Ig kappa chain V-III region HAH*	KV312_HUMAN (+1)	14 kDa	9	10	0	7	18	43	LAWCR
Ig kappa chain V-III region VG (Fragment)*	KV309_HUMAN	13 kDa	0	3	0	0	5	26	SCR
Ig kappa chain V-III region VH (Fragment)*	KV310_HUMAN	13 kDa	0	0	2	0	0	13	SR
Ig kappa chain V-IV region (Fragment)*	KV401_HUMAN (+1)	13 kDa	7	7	0	4	13	26	LAWCR
Ig lambda chain V region 4A*	LV001_HUMAN	12 kDa	0	4	0	0	0	29	AR
Ig lambda chain V-I region BL2*	LV107_HUMAN	14 kDa	0	0	0	0	0	9	R
Ig lambda chain V-I region WAH*	LV106_HUMAN	12 kDa	0	3	0	0	10	11	ACR
Ig lambda chain V-II region BUR*	LV205_HUMAN	12 kDa	0	0	0	0	0	13	R
Ig lambda chain V-III region LOI*	LV302_HUMAN	12 kDa	5	4	4	18	18	16	LASWCR
Ig lambda chain V-III region SH*	LV301_HUMAN	11 kDa	4	0	0	0	0	20	LR
Ig lambda-1 chain C regions*	LAC1_HUMAN	11 kDa	8	17	0	19	0	60	LAWR
Ig lambda-2 chain C regions*	LAC2_HUMAN	11 kDa	11	16	8	32	45	72	LASWCR
Ig mu chain C region	IGHM_HUMAN	49 kDa	44	66	29	118	86	410	LASWCR
Ig mu heavy chain disease protein	MUCB_HUMAN	43 kDa	0	0	0	0	0	283	R
IgGFc-binding protein	FCGBP_HUMAN	572 kDa	0	12	0	0	0	0	A
Immunoglobulin J chain	IGJ_HUMAN	18 kDa	5	7	6	28	12	41	LASWCR

Insulin-like growth factor II	IGF2_HUMAN	20 kDa	0	3	0	0	0	0	A
Insulin-like growth factor-binding protein 3	IBP3_HUMAN	32 kDa	0	6	0	21	0	0	AW
Insulin-like growth factor-binding protein complex acid labile subunit	ALS_HUMAN	66 kDa	0	15	0	77	10	13	AWCR
Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1_HUMAN	101 kDa	4	0	5	149	0	20	LSWR
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2_HUMAN	106 kDa	10	0	6	213	0	9	LSWR
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3_HUMAN	100 kDa	0	0	4	5	0	7	SWR
Inter-alpha-trypsin inhibitor heavy chain H4	ITIH4_HUMAN	103 kDa	10	17	31	357	6	32	LASWCR
Intercellular adhesion molecule 2	ICAM2_HUMAN	31 kDa	0	3	0	0	0	0	A
Junction plakoglobin*	PLAK_HUMAN	82 kDa	0	0	2	0	0	0	S
Kallistatin	KAIN_HUMAN	49 kDa	0	5	0	27	0	0	AW
Keratin, type I cuticular Ha1*	K1H1_HUMAN (+1)	47 kDa	0	4	0	0	0	0	A
Keratin, type I cytoskeletal 10*	K1C10_HUMAN	59 kDa	75	74	54	14	60	10	LASWCR
Keratin, type I cytoskeletal 14*	K1C14_HUMAN	52 kDa	20	23	39	0	0	0	LAS
Keratin, type I cytoskeletal 16*	K1C16_HUMAN	51 kDa	0	21	28	0	0	0	AS
Keratin, type I cytoskeletal 9*	K1C9_HUMAN	62 kDa	27	31	64	24	23	15	LASWCR
Keratin, type II cuticular Hb3*	KRT83_HUMAN	54 kDa	0	10	0	0	0	0	A
Keratin, type II cytoskeletal 1*	K2C1_HUMAN	66 kDa	57	72	83	28	31	0	LASWC
Keratin, type II cytoskeletal 2 epidermal*	K22E_HUMAN	65 kDa	56	50	58	18	58	0	LASWC
Keratin, type II cytoskeletal 5*	K2C5_HUMAN	62 kDa	16	23	34	0	0	0	LAS
Keratin, type II cytoskeletal 6A*	K2C6A_HUMAN (+1)	60 kDa	0	23	22	0	0	0	AS

Keratinocyte proline-rich protein*	KPRP_HUMAN	64 kDa	2	6	0	0	0	0	LA
Kininogen-1	KNG1_HUMAN	72 kDa	12	51	11	121	10	82	LASWCR
Leucine-rich alpha-2-glycoprotein	A2GL_HUMAN	38 kDa	0	0	0	38	60	0	WC
Low affinity immunoglobulin gamma Fc region receptor III-B	FCG3B_HUMAN	26 kDa	0	5	0	20	0	0	AW
L-selectin	LYAM1_HUMAN	42 kDa	0	11	0	19	0	0	AW
Lumican	LUM_HUMAN	38 kDa	0	2	0	52	0	0	AW
Lymphatic vessel endothelial hyaluronic acid receptor 1	LYVE1_HUMAN	35 kDa	0	0	0	13	0	0	W
Lysosome-associated membrane glycoprotein 1	LAMP1_HUMAN	45 kDa	0	0	0	3	0	0	W
N-acetylmuramoyl-L-alanine amidase	PGRP2_HUMAN	62 kDa	0	0	0	21	0	0	W
Peptidase inhibitor 16	PI16_HUMAN	49 kDa	0	0	0	9	0	0	W
Phosphatidylinositol-glycan-specific phospholipase D	PHLD_HUMAN	92 kDa	0	0	0	43	0	0	W
Plasma kallikrein	KLKB1_HUMAN	71 kDa	3	0	16	34	45	56	LSWCR
Plasma protease C1 inhibitor	IC1_HUMAN	55 kDa	8	30	0	234	0	0	LAW
Plasminogen	PLMN_HUMAN	91 kDa	17	0	0	0	0	0	L
Platelet glycoprotein Ib alpha chain	GP1BA_HUMAN	69 kDa	0	0	0	6	0	0	W
Plexin domain-containing protein 2	PXDC2_HUMAN	60 kDa	0	0	0	8	0	0	W
Polymeric immunoglobulin receptor	PIGR_HUMAN	83 kDa	0	0	0	0	0	7	R
Pregnancy zone protein	PZP_HUMAN	164 kDa	34	29	0	138	159	246	LAWCR
Protein AMBP	AMBP_HUMAN	39 kDa	2	11	3	43	31	27	LASWCR
Protein Z-dependent protease	ZPI_HUMAN	51 kDa	0	0	0	14	0	0	W

inhibitor									
Proteoglycan 4	PRG4_HUMAN	151 kDa	0	0	0	26	0	0	W
Prothrombin	THRB_HUMAN	70 kDa	8	5	0	0	53	23	LACR
Putative V-set and immunoglobulin domain-containing protein 6*	VSIG6_HUMAN	14 kDa	0	0	0	0	0	6	R
Scavenger receptor cysteine-rich type 1 protein M130	C163A_HUMAN	125 kDa	0	7	0	7	4	0	AWC
Selenoprotein P	SEPP1_HUMAN	43 kDa	0	6	0	16	0	0	AW
Serotransferrin	TRFE_HUMAN	77 kDa	152	41	51	13	639	146	LASWCR
Serum albumin*	ALBU_HUMAN	69 kDa	221	70	26	181	177	177	LASWCR
Serum amyloid P-component	SAMP_HUMAN	25 kDa	20	9	14	0	26	5	LASCR
Serum paraoxonase/arylesterase 1	PON1_HUMAN	40 kDa	6	0	0	0	0	0	L
Sex hormone-binding globulin	SHBG_HUMAN	44 kDa	0	0	0	8	0	0	W
Sulfhydryl oxidase 1	QSOX1_HUMAN	83 kDa	0	2	0	15	0	0	AW
Thyroxine-binding globulin	THBG_HUMAN	46 kDa	0	0	0	0	54	0	C
Transthyretin	TTHY_HUMAN	16 kDa	16	6	0	2	0	5	LAWR
Vasorin	VASN_HUMAN	72 kDa	0	0	0	9	0	0	W
Vitamin D-binding protein	VTDB_HUMAN	53 kDa	37	0	0	0	0	0	L
Vitronectin	VTNC_HUMAN	54 kDa	3	13	5	18	0	15	LASWR
Zinc-alpha-2-glycoprotein	ZA2G_HUMAN	34 kDa	5	13	0	0	130	9	LACR

* Non-glycoproteins

[a] L-LTA; A- AAL; S-SNA; W-WGA; C- CON A; R-RCA-I

column fractions were 103, 73 and 93, respectively, and that totaled 269 proteins. The identified proteins captured by all the six-lectin columns totaled 506 proteins. The number of non-redundant proteins that were identified in all six column fractions was 165. There were 31 proteins common to all the six columns, indicating that only 19% [(31/165)*100 = 19%] of the proteins showed overlap between the six columns, a result that is comparable to the result obtained by others [5] where 15% overlap was seen with 4 different lectin columns. The number of proteins that were unique to LTA, AAL, SNA, WGA, Con A and RCA-I were 7, 8, 4, 21, 5 and 9, respectively. The number of common and unique proteins captured by the LTA, AAL and SNA columns are shown in the Venn diagram in Fig. 1. The number of common and unique proteins captured by the WGA, Con A and RCA-I are shown in the Venn diagram in Fig. 2. The percentage of glycoproteins identified from LTA, AAL, SNA, WGA, Con A and RCA-I were 73%, 71%, 68%, 84%, 77% and 69%, respectively, which correspond to 27%, 29%, 32%, 16%, 23% and 31% of non-glycosylated proteins, respectively. As was shown and discussed in Chapters III and IV most of the non-glycoproteins were keratin and some of the immunoglobulin chains. The overall percentage of glycoproteins obtained from all six columns was 73%.

<u>Benefits of Combining Broad and Narrow Specificity Lectins in a Six Tandem Columns Format</u>

As shown in the Venn diagram of Fig. 3, there were 94 common and non-redundant proteins captured by the LAS and WCR columns (operated in the combined LAS-WCR

format) as well as 27 and 44 unique and non-redundant proteins were found in the LAS and WCR column fractions, respectively. When compared to each of 3 lectin column

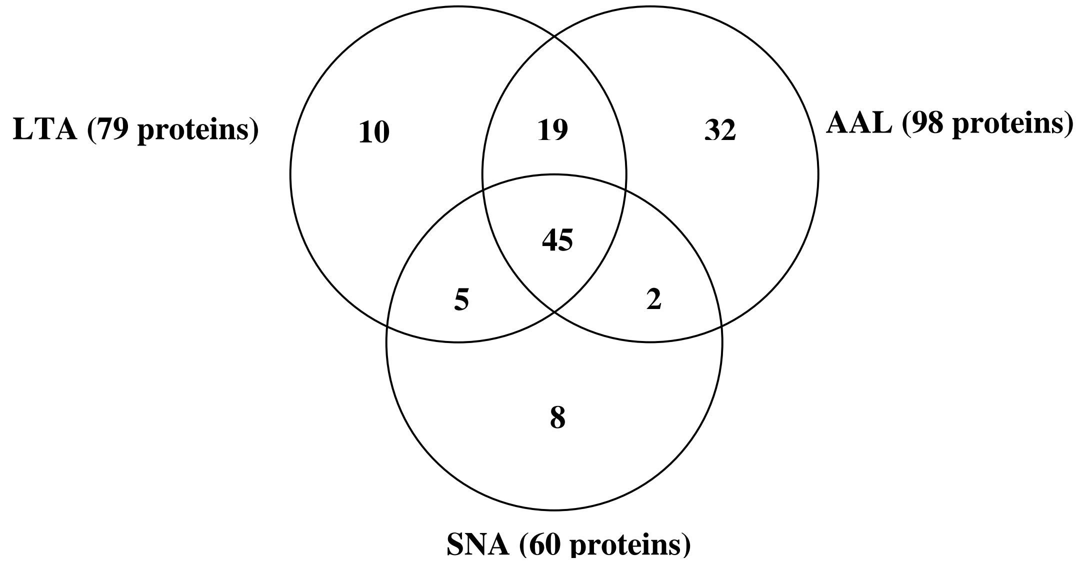


Figure 1. *Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the narrow specificity lectin columns.*

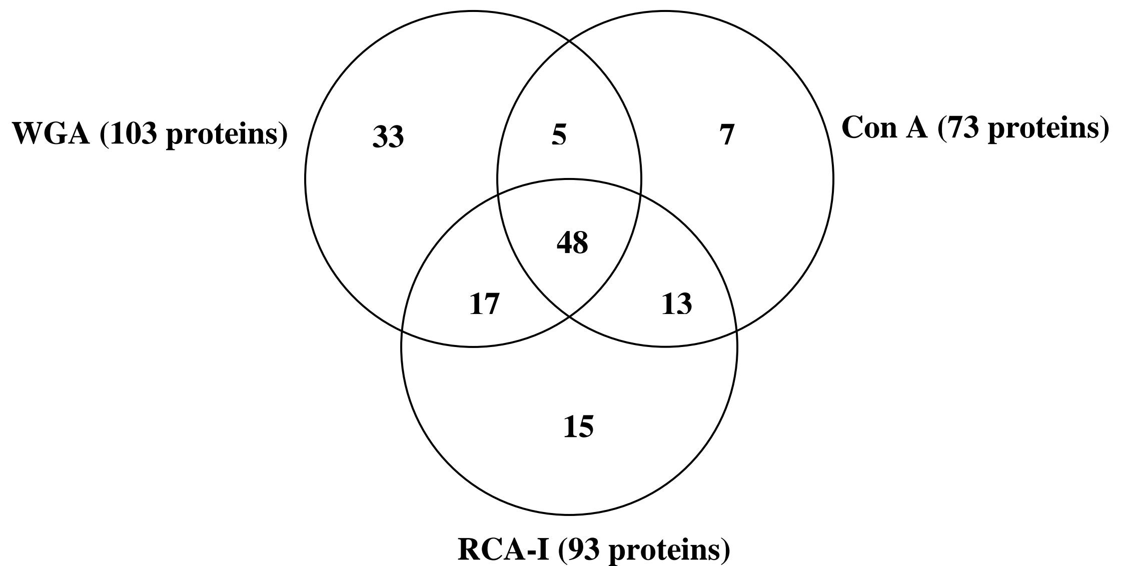


Figure 2. *Venn diagram showing the number of proteins in common and number of proteins unique to each lectin captured from the broad specificity lectin columns.*

series operated alone (discussed in Chapters III and IV), the total proteins as well as the unique and common proteins captured by the six tandem lectin columns series increased significantly. As shown in Fig. 4, the WCR series operated alone captured a total of 75 proteins whereas it captured 138 proteins when it was used in the combined LAS-WCR tandem column series. This represents an increase of 84%. This can be attributed to the

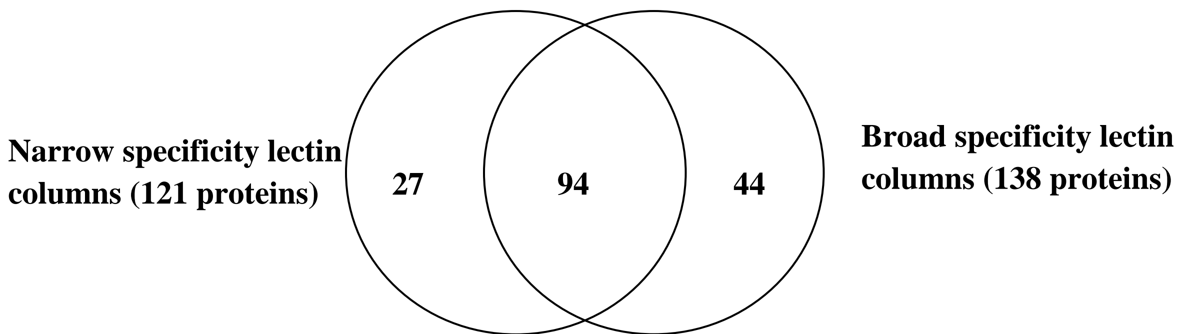


Figure 3. *Venn diagram showing the number of proteins in common and unique to narrow and broad specificity lectin columns.*

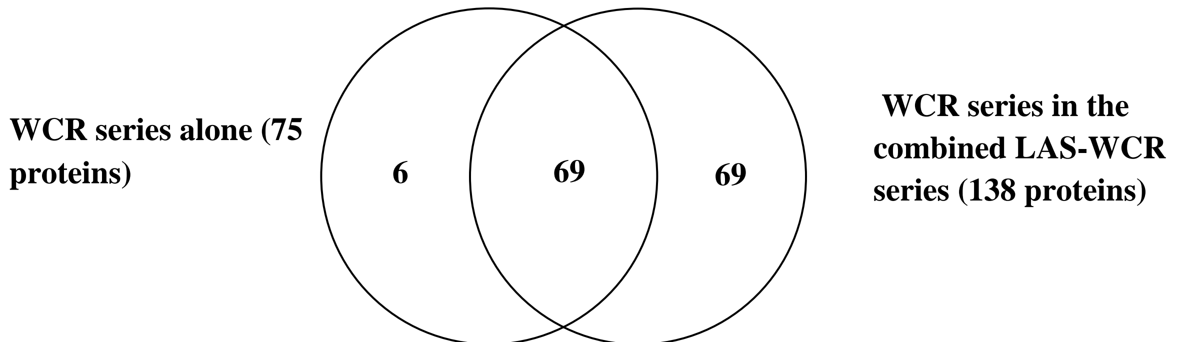


Figure 4. *Comparison of the WCR series alone to the WCR series in the combined LAS-WCR series.*

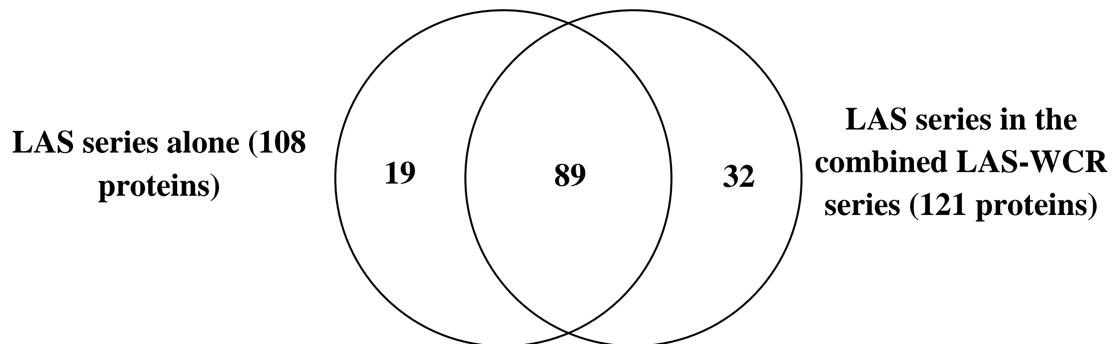


Figure 5. *Comparison of the LAS series alone to the LAS series in the combined LAS-WAR series.*

fact that the LAS series, which was placed first in the six tandem columns, has captured some of the proteins that would otherwise bind to the WCR series, thus freeing up some sites in the WCR series to capture additional proteins. This is further confirmed by the number of proteins captured by the LAS series operated alone when compared to the number of proteins captured by the LAS series operated in the combined LAS-WCR series, see Fig 5. As can be seen in Fig. 5, the LAS series alone captured 108 proteins, which was very close to the 121 proteins captured by the same three-column series operated in the tandem six columns series. Here, the combination is not very beneficial to the series that is placed first in the chain of the 6-column tandem series. The difference is only 121 – 108 = 13 proteins which represent an increase of only 12%.

Another observation for the WCR series can be made as follows. In the 75 proteins obtained from the WCR series operated alone as was found in Chapter III and the 138 proteins obtained from the WCR columns operated in the combined format (i.e., LAS-WCR) as shown in this Chapter, it was found that 69 proteins of the 138 proteins

were identified only in the WCR columns of the 6-column tandem series but not in the WCR series operated alone (refer to Chapter III). Out of these 69 proteins, 37 proteins were unique to the WCR columns in the LAS-WCR series. The remaining 32 proteins (69 - 37 = 32) were identified in the LAS columns. The gain of these new 37 proteins, which were obtained from the WCR columns only when combining the LAS and WCR columns series is a sizable gain of identified proteins. Out of these 37 proteins only six were non-glycoproteins (see Table 2). Even though both columns set ups (i.e., WCR series operated alone as in Chapter III and WCR series operated in the combined LAS-WCR format) had different mass spectrometric analysis conditions, it can be concluded (as just mentioned above) that the gain in the number of identified proteins was due to the fact that LAS column were placed before the WCR columns in the combined series. It is also worth mentioning that when a mass analysis was done under the same conditions for both columns set ups still 55 new proteins were identified by the WCR columns in the combined columns set up. Out of these 55 new proteins 30 proteins were specifically found only in the WCR columns and not in the LAS columns. This new gain of identified proteins indicates that placing the WCR columns behind the LAS columns allows for identification of many new proteins that would have not been identified by using the WCR series alone.

The gain of 63 proteins (138 – 75 = 63 proteins), which represents 84% increase, indicates very well that the combination of the LAS series with the WCR columns in the six tandem columns series allowed the identification of a largerer number of proteins. On this basis, it is interesting to take a closer look at the unique proteins that were captured by the WCR series alone and not common with the LAS series in the combined LAS-

WCR series. As can be seen in Fig. 6, 56 proteins were unique to the WCR of the LAS-WCR series, whereby WGA, Con A and RCA-I captured 28, 12 and 16 proteins, respectively. Some of these were found to be low abundance proteins and clinically important proteins as discussed in the next section.

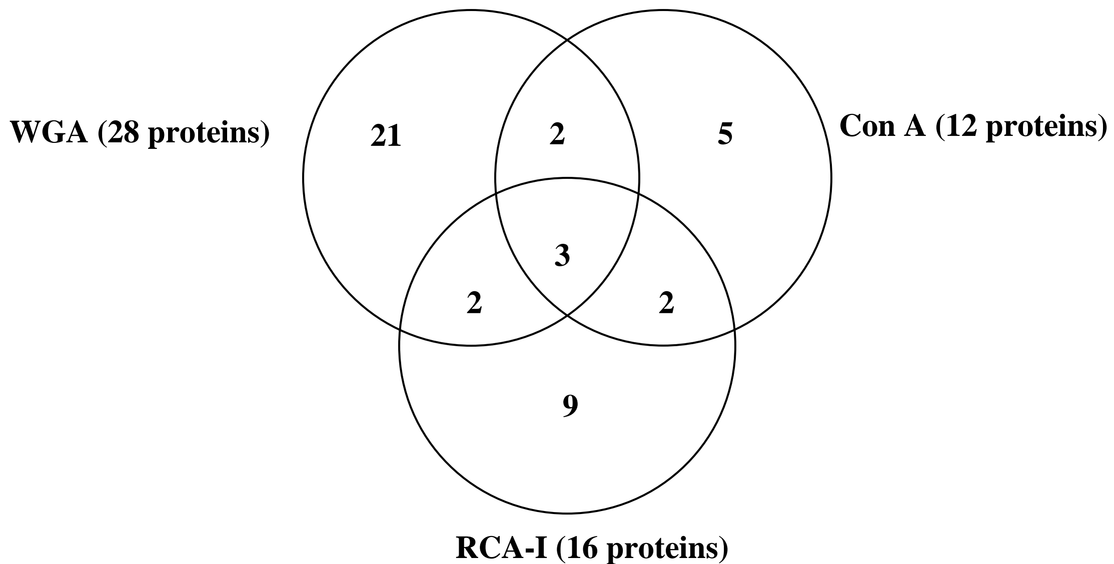


Figure 6. *Venn diagram showing the distribution of the unique proteins of the WCR series among the three lectins*

Returning to the LAS series, the following observation can be made. In the 108 proteins captured by the LAS series operated alone (refer to Chapter IV), it was found that about 89 out of 108 proteins identified in the LAS series (that is 82% of the proteins) were also identified in the LAS series operated in the six-column tandem series. The remaining 19 proteins (i.e., 18% of the total identified proteins) were not identified in the LAS series operated in the six-column tandem series. It needs to be mentioned that out of these 19 proteins, 13 proteins were non-glycoproteins. This mismatch among the LAS columns whether operated alone or within the combined LAS-WCR columns series could

be due to technical biases in the chromatographic and mass spectrometric analysis. It should also to be mentioned that the LC-MS/MS analysis of the LAS series operated alone was done in triplicates while the same analysis of the LAS columns operated within the combined format was done only once.

TABLE 2

PROTEINS THAT WERE IDENTIFIED IN WCR COLUMNS ONLY IN THE COMBINED LAS-WCR COLUMNS

Adiponectin
Aminopeptidase N
Apolipoprotein L1
Beta-Ala-His dipeptidase
Biotinidase
Carboxypeptidase B2
Coagulation factor X
Coagulation factor XII
Complement C1q subcomponent subunit A
Complement C1r subcomponent-like protein
Complement component C7
Cysteine-rich secretory protein 3
Dopamine beta-hydroxylase
Extracellular matrix protein 1
Gamma-glutamyl hydrolase
Hepatocyte growth factor activator
Ig delta chain C region
Ig heavy chain V-II region ARH-77
Ig heavy chain V-III region GA
Ig kappa chain V-I region HK101 (Fragment)
Ig kappa chain V-III region B6
Ig lambda chain V-I region BL2
Ig lambda chain V-II region BUR
Ig mu heavy chain disease protein
Lymphatic vessel endothelial hyaluronic acid receptor 1
Lysosome-associated membrane glycoprotein 1

Peptidase inhibitor 16
Phosphatidylinositol-glycan-specific phospholipase D
Platelet glycoprotein Ib alpha chain
Plexin domain-containing protein 2
Polymeric immunoglobulin receptor
Protein Z-dependent protease inhibitor
Proteoglycan 4
Putative V-set and immunoglobulin domain-containing protein 6
Sex hormone-binding globulin
Thyroxine-binding globulin
Vasorin

*Non-glycoproteins

Some of the Clinically Important Proteins Identified in the Fractions Captured by the Lectin Columns

In the current study, out of the 165 identified proteins many were found to be clinically important proteins. The proteins CD44 antigen, haptoglobin and serotransferrin were captured by both narrow and broad specificity columns. According to a recent compilation of proteins that are differentially expressed in human cancer [6], these proteins have been listed as candidate cancer biomarkers with more than 500 citations. Also, both LAS and WCR columns captured some of the other proteins such as α-1-acid glycoprotein, α-1-antitrypsin, α-2-HS-glycoprotein, α-2-macroglobulin, apolipoprotein A1, apolipoprotein A-II, apolipoprotein C-III, ceruloplasmin, coagulation factor XII B chain, complement C3, complement factor H related protein 1, fibronectin, L-selectin, pregnancy zone protein, transthyretin and vitronectin that have been listed as candidate cancer biomarkers [6].

Some of the proteins that were identified only in the LAS column fractions are discussed in this section. The protein apolipoprotein C-III (Apo C-III) was captured by LTA and AAL columns alone. It has been reported that Apo C-III can induce destruction of β-cell leading to insulin deficiency in type 1 diabetes. Thus, an increased level of Apo C-III has been correlated to type 1 diabetes in children [7, 8]. In the current study, serum paraoxonase/arylesterase and fibrinogen alpha chain were captured only by LTA column. In a recent study, it was reported that serum paraoxonase/arylesterase activity was at elevated levels in stage 2 esophageal cancer as compared to stages 3 and 4 [9]. In another report, it was found that the levels of the protein cancer antigen-125 in epithelial ovarian cancer are associated with higher lipid peroxidation, which in turn is associated with reduced paraoxonase activity [10]. Fibrinogen alpha chain has been listed as candidate cancer biomarker with more than 500 citations [6].

Clinical importance of some of the proteins that were unique to the WCR columns are discussed here. The protein complement component 7 (C7) was captured only by the Con A column. It has been found that there is a possible relationship between esophageal tumorigenesis and reduced expression C7 mRNAs and similar reduction of C7 mRNAs was also observed in kidney and colon cancers [11]. Aminopeptidase N (APN) has been reported to play key a role in tumor progression. From the glycan analysis of APN obtained from *Manduca sexta* (which is a type of insect), it was found that APN contains two core fucosylated *N*-glycans and 13 probable O-glycosylation sites [12]. As the binding sites present in the AAL column (which has affinity towards the core fucosylated glycans) were occupied by other proteins, APN passed through the AAL column and was captured by the WGA column, which has affinity towards *N*-linked

glycans. APN is a zinc-binding type 2 transmembrane ectopeptidase [13]. Differential expression of certain ectopeptidases in human malignancies has made them valuable clinical markers [14]. It has been reported that APN expression in tumor tissue has been associated with a poor prognosis for patients with pancreatic and colon cancer [15, 16]. Also, elevated levels of APN have been found in different solid tumors [17-19].

It is clear from the above discussion that the combined use of both narrow and broad specificity lectins is a useful method to capture a wide range of glycoproteins from serum. In the absence of either the narrow or broad specificity lectins, information about many clinically important proteins that were unique to either narrow or broad specificity lectins would have been lost.

Access to a Wider Protein Concentration Range

In the 165 identified proteins, there were proteins belonging to the high, medium and low abundance categories. High abundance glycoproteins such as serotransferrin, complement C3, α-2-macroglobulin, α-1-antitrypsin and haptoglobin and medium abundance proteins such as ceruloplasmin, complement factor B, complement C4, C1q, α-1-acid glycoprotein, apolipoprotein B and apolipoprotein A1 were also captured by the lectin columns. Except for some of the immunoglobulin chains, most of the other proteins were present at low concentration. Some of the representative proteins are β-2-glycoprotein and angiotensinogen (1×10^{-6} g/L serum), apolipoprotein (a), antithrombin III and vitronectin (1×10^{-4} g/L serum), plasminogen (1×10^{-3} g/L serum) and complement C2 and kininogen-1 (1×10^{-2} g/L serum).

Conclusions

In this research, serial lectin affinity chromatography was used to capture the glycoproteins and LC-MS/MS was used to identify the proteins. This strategy of combining both narrow and broad specificity lectins allowed capturing a wide range of glycoproteins. Since the strategy reported here could capture high, medium and low abundance proteins, the process of depletion of high abundance proteins can be avoided. Also, since the specificities of the lectin used in this study are well known, this strategy could be applied to diseased serum. By doing so proteins that have specific alteration in glycosylation in the diseased serum could be determined and thus can be used in discovery of biomarker candidates.

References

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

[2] R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta* **1999**, *1473*, 4-8.

[3] R. Pieper, Q. Su, C. L. Gatlin, S. T. Huang, N. L. Anderson and S. Steiner, *Proteomics* **2003**, *3*, 422-432.

[4] J. Granger, J. Siddiqui, S. Copeland and D. Remick, *Proteomics* **2005**, *5*, 4713-4718.

[5] M. Madera, Y. Mechref, I. Klouckova and M. V. Novotny, *J Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2007**, *845*, 121-137.

[6] M. Polanski and N. L. Anderson, *Biomarker Insights* **2006**, *2*, 1-48.

[7] P. Blackett, D. C. Sarale, J. Fesmire, J. Harmon, P. Weech and P. Alaupovic, *South Med. J.* **1988**, *81*, 469-473.

[8] J. E. Hokanson, G. L. Kinney, S. Cheng, H. A. Erlich, A. Kretowski and M. Rewers, *Diabetes* **2006**, *55*, 834-838.

[9] C. Kerim, B. Mehmet, B. Salim, Tekin., K. Fatih, T. Atilla and Y. Abdülkadir, *Eur. J. Gen. Med.* **2010**, *7*, 398-403.

[10] H. Camuzcuoglu, D. T. Arioz, H. Toy, S. Kurt, H. Celik and O. Erel, *Gynecol. Oncol.* **2009**, *112*, 481-485.

[11] R. Oka, T. Sasagawa, I. Ninomiya, K. Miwa, H. Tanii and K. Saijoh, *Eur. J. Cancer* **2001**, *37*, 1158-1165.

[12] P. J. Knight, J. Carroll and D. J. Ellar, *Insect Biochem. Mol. Biol.* **2004**, *34*, 101-112.

[13] H. Tsukamoto, K. Shibata, H. Kajiyama, M. Terauchi, A. Nawa and F. Kikkawa, *BMC Cancer* **2008**, *8*, 74.

[14] C. Antczak, I. De Meester and B. Bauvois, *Bioessays* **2001**, *23*, 251-260.

[15] H. Hashida, A. Takabayashi, M. Kanai, M. Adachi, K. Kondo, N. Kohno, Y. Yamaoka and M. Miyake, *Gastroenterology* **2002**, *122*, 376-386.

[16] N. Ikeda, Y. Nakajima, T. Tokuhara, N. Hattori, M. Sho, H. Kanehiro and M. Miyake, *Clin. Cancer Res.* **2003**, *9*, 1503-1508.

[17] G. Mechtersheimer and P. Moller, *Am. J. Pathol.* **1990**, *137*, 1215-1222.

[18] F. Alliot, J. Rutin, P. J. Leenen and B. Pessac, *J. Neurosci. Res.* **1999**, *58*, 367-378.

[19] J. Dixon, L. Kaklamanis, H. Turley, I. D. Hickson, R. D. Leek, A. L. Harris and K. C. Gatter, *J. Clin. Pathol.* **1994**, *47*, 43-47.

VITA

Subhashini Selvaraju

Candidate for the Degree of

Doctor of Philosophy

Thesis: IN-DEPTH PROTEOMICS AND GLYCOPROTEOMICS BY MASS SPECTROMETRY AND AFFINITY-BASED PLATFORMS FOR SELECTIVELY CAPTURING PROTEINS AND GLYCOPROTEINS FROM BREAST CANCER AND DISEASE FREE HUMAN SERUM

Major Field: Chemistry

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Chemistry at Oklahoma State University, Stillwater, Oklahoma in December, July, 2011.

Completed the requirements for the Master of Science in Applied Chemistry at Anna University, Chennai, India in 2003.

Completed the requirements for the Bachelor of Science in Chemistry at Madras University, Chennai, India in 2001.

Experience:
Graduate Teaching Assistant, Oklahoma State University, 2006 to present.
Research Scientist, Aurigene Discovery Technologies, Bangalore, India, 2005-2006.
Officer R&D, Nicholas Piramal India Limited, Chennai, India 2003-2005.

Name: Subhashini Selvaraju Date of Degree: July, 2011

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: IN-DEPTH PROTEOMICS AND GLYCOPROTEOMICS BY MASS SPECTROMETRY AND AFFINITY-BASED PLATFORMS FOR SELECTIVELY CAPTURING PROTEINS AND GLYCOPROTEINS FROM BREAST CANCER AND DISEASE FREE HUMAN SERUM

Pages in Study: 180 Candidate for the Degree of Doctor of Philosophy

Major Field: Chemistry

Scope and Method of Study:

The major purpose of the work presented in this dissertation is to contribute to the introduction and evaluation of new ways and strategies for reducing the complexity of human serum and in turn contribute to facilitating the identification of candidate biomarkers with the assistance of advanced mass spectrometry techniques. In one approach, immobilized metal affinity chromatography (IMAC) and protein equalizer techniques were combined to reduce the complexity of the human serum. In another approach, lectin affinity chromatography involving tandem lectin columns with narrow and broad specificity lectins were evaluated in the selective enrichment of the glycoproteins present in human serum.

Findings and Conclusions:

The investigation described in this dissertation has significantly contributed to the in-depth proteomics analysis of human serum. The equalization/IMAC strategy allowed the identification of 82 non-redundant proteins, which was facilitated by the IMAC post-fractionation process after equalization. The tandem-lectin affinity based platforms developed in this investigation selectively captured glycoproteins from breast cancer serum and disease-free serum and many proteins that were differentially expressed in the cancer serum were identified. In a platform where a combination of broad and narrow specificity lectins were evaluated in tandem series, 165 non-redundant proteins were identified. The platforms developed in this investigation are expected to be of general use and to facilitate the identification of additional candidate biomarkers for various diseases in the future.

ADVISER'S APPROVAL: Dr. ZIAD EL RASSI