Structural Characterization

of Potential Cancer Biomarker Proteins

by

Samita Rai

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2012 by the
Graduate Supervisory Committee:

Randall Nelson, Co-Chair
Mark Hayes, Co-Chair
Chad Borges
Alexandra Ros

ARIZONA STATE UNIVERSITY

August 2012

ABSTRACT

Cancer claims hundreds of thousands of lives every year in US alone. Finding ways for early detection of cancer onset is crucial for better management and treatment of cancer. Thus, biomarkers especially protein biomarkers, being the functional units which reflect dynamic physiological changes, need to be discovered. Though important, there are only a few approved protein cancer biomarkers till date. To accelerate this process, fast, comprehensive and affordable assays are required which can be applied to large population studies. For this, these assays should be able to comprehensively characterize and explore the molecular diversity of nominally "single" proteins across populations. This information is usually unavailable with commonly used immunoassays such as ELISA (enzyme linked immunosorbent assay) which either ignore protein microheterogeneity, or are confounded by it. To this end, mass spectrometric immuno assays (MSIA) for three different human plasma proteins have been developed. These proteins viz. IGF-1, hemopexin and tetranectin have been found in reported literature to show correlations with many diseases along with several carcinomas. Developed assays were used to extract entire proteins from plasma samples and subsequently analyzed on mass spectrometric platforms. Matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) mass spectrometric techniques where used due to their availability and suitability for the analysis. This resulted in visibility of different structural forms of these proteins showing their structural micro-heterogeneity which is invisible to commonly used immunoassays. These assays are fast, comprehensive and can be applied in large sample studies to analyze proteins for biomarker discovery.

DEDICATION

I would like to dedicate this thesis to my teacher and divine mother Shri Mataji Nirmala Devi. Her guidance and love is a constant source of strength in my life.

ACKNOWLEDGEMENT

I would like to sincerely thank my advisor Dr. Randall Nelson for supporting, guiding and encouraging me all through my graduate studies. His insistence on understanding ideas on a bigger scale has every time improved my vision towards the problem at hand and otherwise.

I would also like to thank Dr. Chad Borges for being my mentor and teaching me skills that has improved me as an analytical biochemist. He has been a great support during all my work and a go to person for all discussions big and small.

I would also like to thank members of the molecular biomarkers lab, Drs. Matthew Schaab, Paul Oran and Jason Jarvis, Nisha Sherma, and Douglas Rehder for their constant support and friendship.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

1.1 General overview:

Biomarker discovery and characterization is emerging as a useful tool in early diagnosis and subsequent treatment of many diseases. These biomarkers include both genetic and proteomic markers. Genetic markers are mainly DNA or RNA based markers. DNA-based markers include Single-Nucleotide Polymorphisms (SNPs) which can be used to map disease associated genes, chromosomal aberrations (such as the well-known Philadelphia translocation associated with chronic myelogenous leukemia), changes in DNA copy number and differential promoter-region methylation. RNA-based biomarkers include over or under expressed transcripts, and regulatory RNAs (for example, the microRNAs). HapMap is an international collaborative project to provide researchers world-wide, information about these genetic markers [6, 7]. Protein markers include cell-surface receptors, tumour antigens such as prostate-specific antigen (PSA) and proteins/peptides released by cells into serum, urine, sputum or other body fluids. The popular genomics approach, while useful in many ways, has its limitations. Firstly, the analysis requires sufficient amount of viable RNA. Secondly, techniques commonly applied in genomics provide information about

the levels of RNA present but do not indicate protein levels because many events occur during translation and post-translational modification.

Proteins themselves are macromolecules which are long chains of amino acids. These amino acid chains are constructed in cells when cellular machinery of the ribosomes translates RNA transcripts from DNA in the cell's nucleus. Within a given human proteome, the number of proteins can be as large as 2 million [8]. Proteins can be organized in four structural levels viz. primary, secondary, tertiary and quaternary. The primary structure refers to the sequence of amino acids in the polypeptide chain. Local folding of amino acid sequence into α helices and β sheets is referred to as the secondary structure of a protein and 3D conformation of the entire amino acid sequence is the tertiary structure. Interaction between multiple protein subunits folded in 3D gives rise to quaternary structure of a protein (Figure 1). Each level of protein structure is essential for the proper functioning of the protein molecule. Primary sequence of the amino acid chain determines where secondary structures will form, as well as the overall shape of the final 3D conformation. The 3D conformation of each small peptide or subunit determines the final structure and function of a protein conglomerate. Serving as important components of the physiological pathways in cells, proteins play critical roles in vital functions of the body such as, catalyzing various biochemical reactions as enzymes; acting as messengers, e.g. neurotransmitters; acting as control elements that regulate cell reproduction; influencing growth and development of various tissues, e.g. growth factors; transporting oxygen in the

blood, e.g. hemoglobin; and defending the body against disease, e.g. antibodies. When accurately characterized, proteins represent the dynamic state of cells, reflecting pathophysiological changes arrising due to certain diseases in a more timely and accurate manner than genomic and epigenetic alterations.

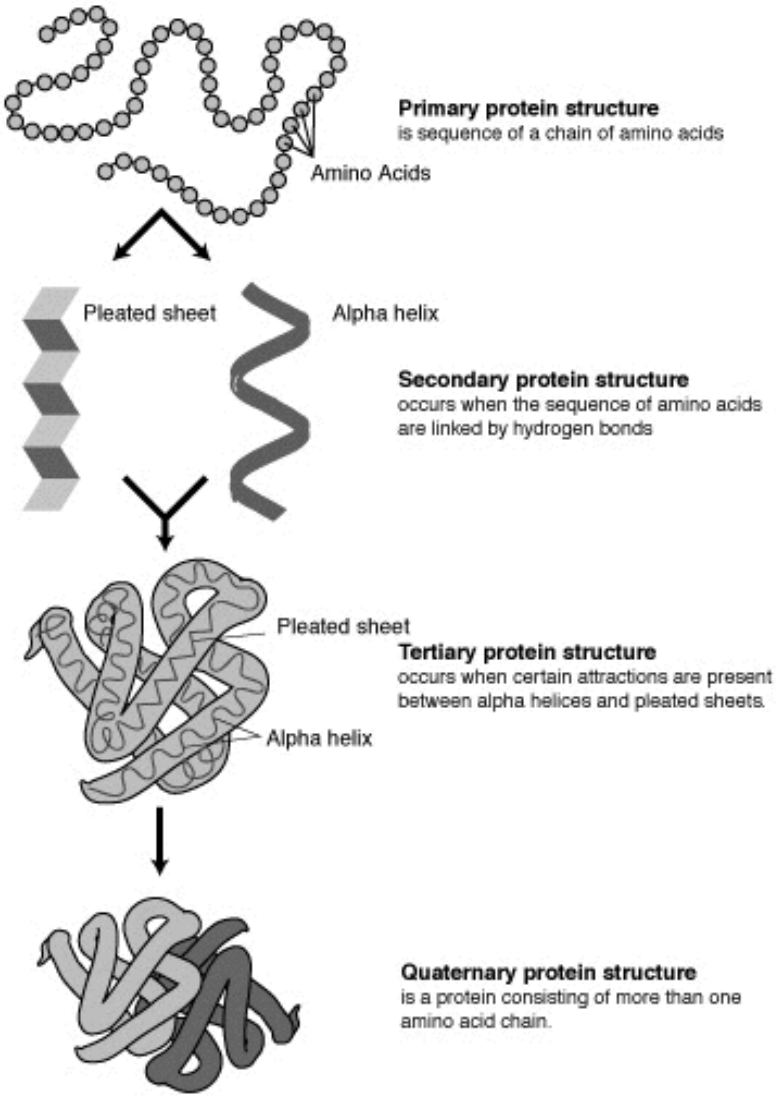

Figure 1: Characterization of protein structure. Courtesy: National Human Genome Research Institute

1.2 Proteins as biomarkers:

According to National Institute of health, biomarkers of disease are "Characteristics that are objectively measured and evaluated as indicators of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [9]. In case of proteins these characteristics can represent either change in concentration, amino acid sequence, and/or posttranslational modifications. Ideally, biomarkers of diseases are blood-based molecules that indicate state of the disease, allowing for minimally invasive testing. Such analysis not only helps in early diagnosis of a disease but also provides a reliable measure to monitor the course of the treatment and help doctors tailor their medication type and doses. Over the past decade, due to rapid development and improvement of proteomic technologies, hundreds of potential protein biomarkers have been identified [10]. But, to be approved as a clinically useful biomarker, a protein molecule must meet several conditions. Firstly, in order to distinguish normal from diseased samples the biomarker should have tissue specific expression. Expression in additional tissues means a high level of expression in normal healthy individuals. Secondly, to be effective for early diagnosis, the potential biomarker protein must be released into circulation at a detectable concentration at an early disease state. Biomarkers not expressed in the early stages are useful only for disease staging and monitoring. Thirdly, there must be sufficient biomarker released from the diseased tissue to be detectable over other highly expressed background proteins and lastly a biomarker should be

predictive which means it should have sufficient life in the blood and should also be released in proportional amount to the disease progression.

1.3 Cancer Biomarkers

According to American Cancer Society Cancer is a group of diseases characterized by uncontrolled growth and spread of abnormal cells. If the spread is not controlled, it can result in death. Cancer is the second most common cause of death in the US, exceeded only by heart disease. It is the cause of one in every four deaths in the United States. In 2011 alone, about 571,950 Americans are expected to die of cancer which is more than 1,500 people a day [11]. Even on a global scale cancer is becoming an alarmingly higher cause of death. Worldwide, cancer was estimated to be responsible for 13% of all deaths in 2007 [12]. Given these numbers, prevention of cancer is very important for societies on a global level. It is predicted that approximately one-third of all cancer cases could be prevented if detected at an early enough stage [12]. This means that appropriate biomarkers could play a big role in allowing clinicians to detect cancer onset at right time and save patient's lives. Cancer is a complex heterogeneous disease which evolves via multiple pathways and is a result of a variety of genetic, molecular, and clinical events which are not very well understood even till date[13]. Given that there is significant variation in the risk of developing cancer and that early detection often results in increased survival, finding biomarkers capable of identifying patients at highest risk and detecting tumors in the earliest

stages of development is a pressing need. In addition, suitable and high throughput techniques to screen a larger population in a cost effective manner is required.

In recent years, the focus of the scientific community on cancer and the rapid technological advances in the field of proteomics has led to identification of hundreds of potential protein biomarkers for cancer. Even though many potential biomarkers have been identified, very few have successfully advanced beyond the discovery phase to become clinically useful diagnostic biomarkers. Food and Drug Administration (FDA), has only approved less than two dozen cancer biomarkers [13]. of which only nine are protein biomarkers identifiable in the blood [14]. These protein markers include cancer antigen-125 (CA-125); and human epididymis protein 4 (HE4) for monitoring of ovarian cancer; a-fetoprotein (AFP) for staging of non-seminomatous testicular cancer and monitoring of hepatocellular carcinoma; prostate specific antigen (PSA) for screening and monitoring of prostate cancer, thyroglobulin (Tg) for monitoring of thyroid cancer; carcinogenic embryonic antigen (CEA) for monitoring of pancreatic cancer; and CA15-3/CA27-29 and HER2/neu for monitoring of breast cancer [14]. Other approved protein biomarkers are identified in urine, such as nuclear matrix protein 22, fibrin-/fibrinogen-degradation products, and bladder-tumor antigen for the monitoring of bladder cancer, or by immunohistochemistry using tumor tissues, such as the estrogen receptor for breast cancer [14]. Additional

approved cancer biomarkers are DNA based, such as human epidermal growth factor receptor 2, or HER2/neu, for breast cancer [14].

1.4 Microheterogeneity: The unheard noise

Protein structure plays an important role in its function. Even slight modifications in amino acid sequence or changes in small molecules attached to the main sequence due to post translational modifications can result in several forms of a protein which can be functionally very distinct. These alterations thus may have important role to play in disease physiology. Defining and quantifying these events is therefore critical to understanding the pathobiological mechanisms underlying disease. Currently, protein microheterogeneity within the general population which is not distinguishable, simply adds biological “noise” in conventional assays and sometimes contributes to misleading data. If these different forms can be distinctly seen then the knowledge of their qualitative differences, their relative abundances, and their frequency within distinct populations will markedly advance our understanding of the pathobiological mechanisms underlying disease and significantly improve our chances of finding clinically useful biomarkers. Unfortunately, due to speed and simultaneous analysis of all proteins being the priorities in earlier proteomics research, not much attention has been paid to the fine structure of proteins (microheterogeneity) which has the potential to completely alter protein activity. Thus, defining protein microheterogeneity in sizable populations of healthy and diseased subjects is of

great importance in order to better understand disease physiology specially in case of cancer. In order to establish a protein biomarker, high throughput techniques are required that are precise, cost effective and sensitive. For this purpose currently used techniques have their limitations as far as studying protein's microheterogeneity is concerned. Immunological assays such as enzyme linked

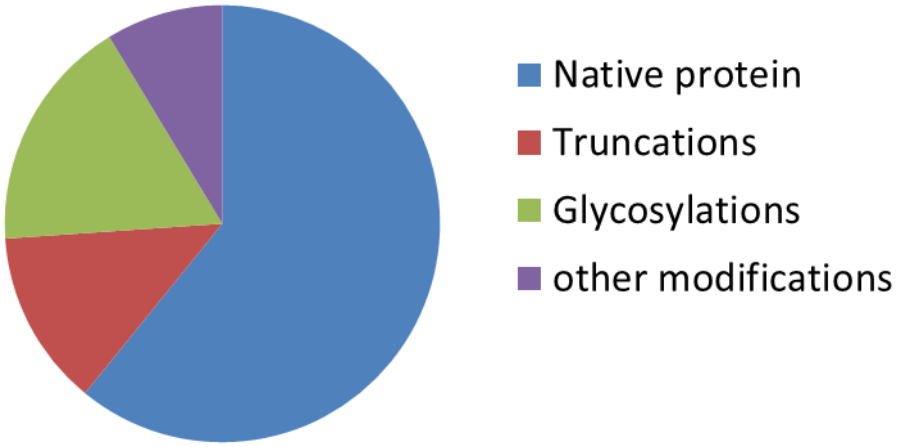

Figure 2: Illustration of real qualitative variation that exists within what is considered to be one protein using common immunoassays. MSIA enables us to look at the real picture where a protein exists in several different forms. This information about the heterogeneous structures opens up an unexplored space where potential biomarkers may exist.

immunosorbent assay (ELISA) have been used to study protein concentrations in biological samples. But, these assays only provide us information regarding the quantitative amount of the total protein in the sample. This brings us to a blind spot in the analysis because multiple modifications (arising from genetic diversity, transcriptional and post-translational events) leading to protein diversity (microheterogeneity) in the human population (Figure 2) remain indistinguishable from each other. In global proteomics, proteins are fragmented. Few of these

proteolytic fragments are then identified using mass spectrometry. The results are then mappd to a parental protein using database search engines. But in this approach, critical information about the native protein is irrecoverably lost because the approach assumes that one or a few proteolytic fragments unambiguously define a single precursor which may not be true considering protein microheterogeneity. It is only when all the variants in their intact forms are separately isolated and characterized with total sequence coverage that this diversity is evident and can be correctly assigned. With immunological methods, each variant requires its own highly specific antibody system (usually two) to yield an accurate evaluation of its contribution to the entire cluster. For the plethora of human proteins identified, only a handful of such immunoassays currently exist. Consequently, human protein heterogeneity is imperceptible by these techniques.

Chapter 2

MASS SPCETROMETRIC IMMUNO ASSAY

2.1 Technique

Mass Spectrometric Immunoassay (MSIA) invented by my advisor Dr. Nelson is a rational combination of micro-scale immunoaffinity capture and mass spectrometry [15]. In MSIA intact proteins are captured using specific antibodies as is done in traditional immunoassays. But, the advantageous strategy used in case of MSIA is that instead of immobilizing antibodies on a flat surface or beads the antibodies for protein or peptide capture are immobilized in small, porous microcolumns that are fitted at the entrance of a pipettor tip (Figure 3). These tips are termed affinity pipettes. Immobilizing of antibodies onto these pipettor fitted microcolumns allows for higher efficiency of protein capture because the sample can readily be aspired and dispensed (multiple times) through the tip to expose the immobilized antibody to the antigen present in the sample. Typically a sample is pipetted for about 1500 times with the use of pipettor robot. Once the pipetting is over the antibodies have the antigen in all the different variant forms available in the sample captured for analysis. Along with the protein of interest some other sample components and salts are also nonspecifically retained. These nonspecifically retained components are removed by rinsing the tip with appropriate buffer and water.

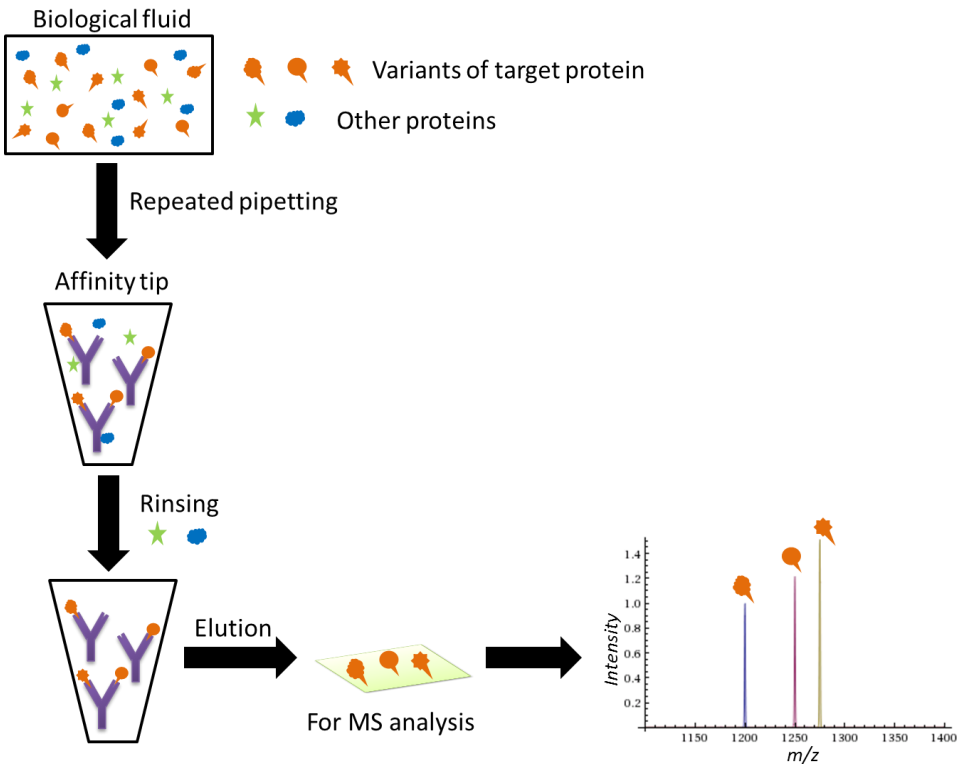

Figure 3: Diagram of the MSIA process. A biological sample is repetitively passed through the MSIA-Tip, performing immunoaffinity capture of the targeted protein. While nonspecific components within the biological fluid are rinsed away, the concentrated target analyte is eluted for analysis by either MALDI or ESI-based mass spectrometry. Mass spectral peak areas are then integrated to determine the relative percent abundance of each unique variant that, in the conventional vernacular, together constitute (nominally) "one" protein.

The protein of interest with all its captured variants is now ready for analysis using any suitable mass-spectrometric (MS) platform. The platform of choice depends on the target of interest which could be a full length protein or a peptide fragment. The size of the target and abundance in the sample also dictates the choice for a suitable MS platform. Then of course the availability and ease of use of the instrument plays a role in choosing the MS platform especially during the development of a qualitative MSIA. Next step after the buffer wash of the pipettor

tip depends on the chosen MS platform. During the course of work presented in this thesis I have worked with two platforms viz Matrix Assisted Laser Desorption Ionization (MALDI)-MS and electrospray ionization (ESI-MS). For MALDI based analyses, a small volume of MALDI matrix is aspirated into the affinity pipette, and the antigen-containing eluate is quickly deposited directly onto a MALDI target for ensuing MS analysis. For ESI based configurations, an acidic buffer is used and the eluate is deposited onto titer plates for downstream MS analysis.

2.2 Advantages

MISA offers several advantages over conventional techniques of protein analysis. Firstly, MSIA avoids protein fragmentation and enables us to capture intact proteins using affinity chromatography and analyze its complete structure using mass spectrometry. This unique feature of MSIA allows us to examin structural microheterogeneity of target proteins in different physiological conditions. Secondly, it offers the possibility of simultaneous capture and analysis of multiple proteins in the same sample. This can be done by immobilizing multiple antibodies in the same affinity pipette (i.e. multiplexing) as long as the antibody affinities and protein concentrations in the sample are not several orders of magnitude different, and also the masses of the proteins to be analyzed are relatively distinct so that their signals do not overlap in the mass spectra.After the capture of multiple proteins the following MS analysis results in a mass spectrum

that contains distinct signals from all proteins. Multiplexing is of great clinical significance in scenarios were one protein biomarker is not sufficient to provide information that can be clinically useful and needs a range of proteins to be analyzed. Also for screening purposes multiplexing is very helpful. This reduces both the cost and time associated with large population screenings. Thirdly, MSIA is high throughput. Samples are prepared onto a 96 well titer plate in parallel for the affinity capture. The eluates are then deposited onto 96 well formatted MALDI targets or titer plates for ESI. This entire process of affinity capture and elution of the target onto the MS platform after the assay development is fully automated and can be carried out in less than an hour, which means less than a minute for one sample to be extracted and eluted for analysis. Given this capability of high throughput analysis population studies of some biomarkers on a large scale involving thousands of samples have been performed [16]. Fourthly, MSIA can accommodate a range of sample volumes from as small as 10 µl to as large as 50 ml. The advantage comes from the process of repeated sample rinsing through the pipette until enough antigens are bound to the antibody. Given all these advantages of MSIA technique over conventional immunoassays it has been used to study a wide range of proteins with concentrations ranging from ~ 10 pg/ml to ~ 50 mg/ml [14] and promises to be a useful tool to study human plasma proteome which spans a wide dynamic range.

Chapter 3

PROTEINS OF INTEREST

In order to develop assays for cancer associated biomarkers we need to choose the candidate proteins. This choice is made by mining into research article/abstract database provided by PubMed and prioritizing the candidates based on their relevance to the disease of interest. This thesis includes work on development of assays for three different proteins viz. Insulin like growth factor-1 (IGF-1), Hemopexin and Tetranectin.

3.1 IGF-1

IGF-1 is a member of insulin growth factors (IGF-1 and IGF-2). It is produced mainly by the liver but is also produced in several other tissues [17]. The levels of IGF-1 in tissues and serum are variable and are regulated ontogenically as well as by multiple hormones and nutritional factors. Growth hormone is the most prominent stimulus for IGF-1 production. Normal IGF-1 levels in plasma range from 20-600 μg/L [18]. IGF-I levels in human fetal serum are relatively low, growth hormone-independent, and are positively correlated with gestational age and birth weight. During childhood IGF-1 concentrations in serum shows a slow and gradual rise and during puberty, IGF-I concentrations rise to 2–3 times above the levels found in adults which is related to the action of sex steroids. In

adulthood, serum IGF-I levels show a gradual and progressive age-associated decline [19]. As the name suggests these are involved in regulation of cell proliferation, differentiation and apoptosis.

3.1.1 Structure:

IGF-1 is a single polypeptide composed of 70 amino acids, with a calculated molecular weight of 7649 Da. It has three intramolecular disulfide bridges as shown in Figure 4 [20] These disulfide bridges between cysteine residues 6-48, 18-61 and 47-52 help to determine the shape of the protein by playing

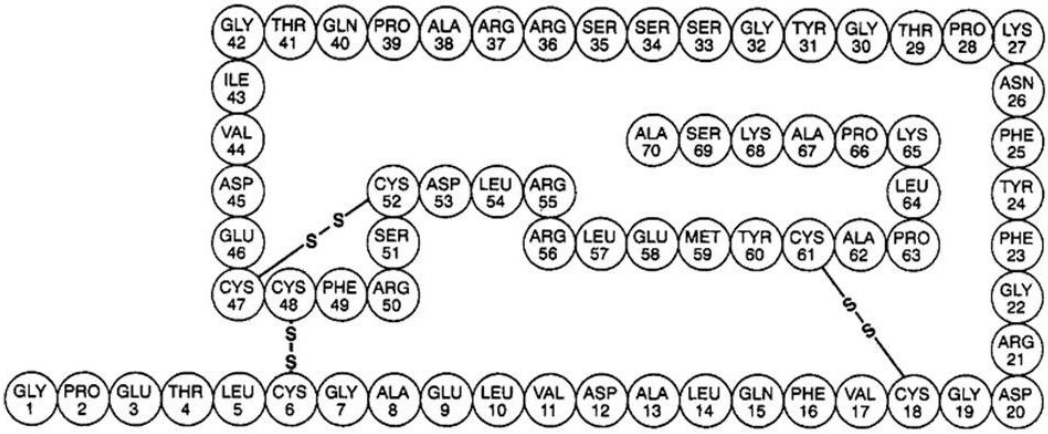

Figure 4: Sequence of human IGF-1. Disulfide bonds are indicated by solid lines [2].

important role in its folding and also play important role in its function. Without any of the disulfide bridges, significant decrease in its binding to its receptor is observed which indicates that IGF conformation undergoes change and its activity is suppressed if these disulfide bonds are disrupted [2]. IGF-1 as the name

suggests is insulin like in its structure. About 50% sequence similarity exists between insulin and IGFs. IGF molecules have four domains which are designated as A, B, C, and D (Figure 5). Domains A and B are homologous with the A and B chains of insulin, respectively. Even with this high degree of similarity many difference exist in their functions.

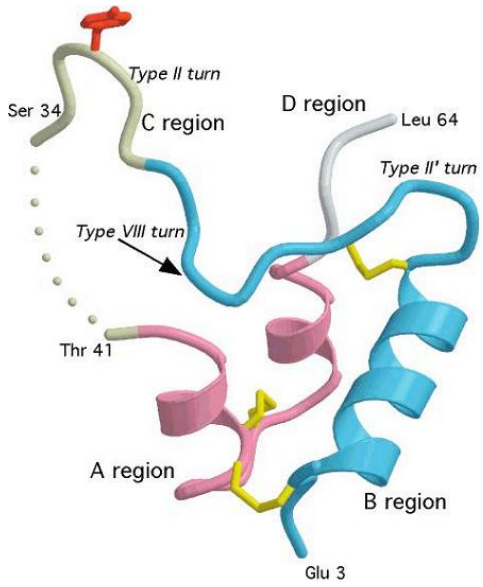

Figure 5: Tertiary structure of IGF-1 [1].

3.1.2 Functions:

The functions of IGFs are mediated through their interaction with two transmembrane receptors (IGFR-1 and IGFR-2). The type 1 IGF receptor mediates many diverse and tissue specific IGF actions on all cell types. In general, it is believed that all of the effects of IGF receptor activation are mediated by tyrosine kinase activation and phosphorylation of substrates, which

activate specific cellular pathways leading to various biological actions. Among these effects are induction of cell growth, maintenance of cell survival (prevention of apoptosis) and induction of cellular differentiation [21]. The type 2 IGF-receptor bears no structural homology with the type 1 IGF receptor and binds both IGF-II and Mannose-6-P-containing lysosomal enzymes, as well as retinoic acid, while IGF-I binds to it with substantially lower affinity [22]. This receptor acts as a growth inhibitory component of the IGF system, responding to and mediating multiple anti-mitogenic systems.

In plasma, IGF-1 is complexed with a family of proteins called IGF binding proteins (IGFBPs). About 99% of the IGFs circulate in the serum in the form of complex with IGFBPs and less than 1% only is available in the free unassociated form. The complex increases the half-life of IGF-1 in the plasma, help in transportation of IGF-1 to target cells and play a role in the interaction of the IGF-1 with its surface membrane receptors [23]. These IGFBPs are six distinct proteins termed IGFBP-I through IGFBP-6 having high binding affinity to IGF-1. IGFBP-3 is the major IGFBP in normal human plasma and circulates in adult plasma as part of a ternary complex; consisting of IGFBP-3, an IGF peptide, and an acid-labile subunit [24] IGFBPs both inhibit and enhance the action of IGF-1. Inhibition presumably is caused because of IGFBPs competing with IGF receptors for IGF peptides[25] while several of the IGFBPs enhance IGF action, by facilitating IGF delivery to target receptors.[26] IGFBPs in themselves are also bioactive and have a variety of IGF independent functions. These include growth

inhibition, direct induction of apoptosis, and modulation of the effects of other non-IGF growth factors. These effects of IGFBPs are mediated by binding to their own receptors. IGF action which is regulated by IGFBPs is indirectly modulated by IGFBP proteases, which have capability of IGFBP degradation[24] and demonstrates clear growth hormone dependence.

3.1.3 Clinical significance:

Clinical significance of IGF-1 is considerable. Apart from several carcinomas it has been associated with a number of other diseases including cardiovascular diseases (myocardial infarction (MI), heart failure) [27], acromegaly and diabetes [27].

Clinical studies of IGF-I in patients with type 2 diabetes has shown improvement in their insulin sensitivity [28] In type 1 diabetes, multiple clinical studies have shown improved glycemic control and decreased insulin requirement through IGF-I therapy[29] Apart from this, studies on several animal models have proved the role of IGF-1 in carbohydrate metabolism [30][31].

In relation to cardiovascular diseases, studies have shown a negative correlation between IGF-1 and total HDL-cholesterol ratio and IGF-1: IGFPB-3 molar ratio has also been inversely correlated with total HDL-cholesterol ratio and triglycerides [32]. IGF-1 meditated beta cell proliferation is found to be inhibited by free fatty acids suggesting direct link between obesity, IGF-1 and glucose regulation [33]. Several studies have associated low total or free IGF-1 levels with

an increased risk for cardiovascular disease. In particular, low IGF-1 levels have been correlated with increased risk of MI, angina and ischemic stroke [34] [35] [36].

IGF-1 has also been associated with different cancer types including prostate [37], colon[38] [39], [40], lung[41], [42] and premenopausal breast cancers [43], [44]. Specially, in the case of prostate cancer several epidemiological studies conducted in distinct populations have demonstrated that elevated serum IGF-I level is associated with an increased risk of prostate cancer[45], [46]. The risk of cancer is higher among people with raised concentrations of IGF-I, and it is lower among those with high concentrations of IGFBP-3 which is the main binding protein complexed with IGF-1. In many studies anthropometric markers of the activity of insulin-like growth factor-I, such as height and leg length, are associated with cancer incidence, particularly with the cancers for which risk increases with rising concentrations of IGF-1. [47]

3.2 Hemopexin:

Hemopexin is a plasma glycoprotein with a high affinity of binding to heme (K_d < 1 pM) at an equimolar ratio.[40] It transports heme to hepatocytes for salvage of the iron. It is mainly expressed in liver as an acute protein which is expressed as a result of an inflammatory event. [41]

3.2.1 Structure:

Human hemopexin is composed of a single polypeptide chain containing 439 amino acid residues. It has six intrachain disulfide bridges. The molecular weight of hemopexin as calculated from the amino acid sequence of the unmodified polypeptide chain is 49, 295 Da. This molecular weight is much less that what has been reported in the literature which ranges from 57,000 Da to 80,000 Da.[48], [49] The extra mass observed comes from additional modifications of the protein. The amino-terminal threonine residue has an O-linked galactosamine oligosaccharide, and the protein has five N-linked glycans attached to the acceptor sequence Asn-X-Ser/Thr (X represents any amino acid) and also binds heme. This increases the overall mass of hemopexin. Hemopexin exhibits internal homology

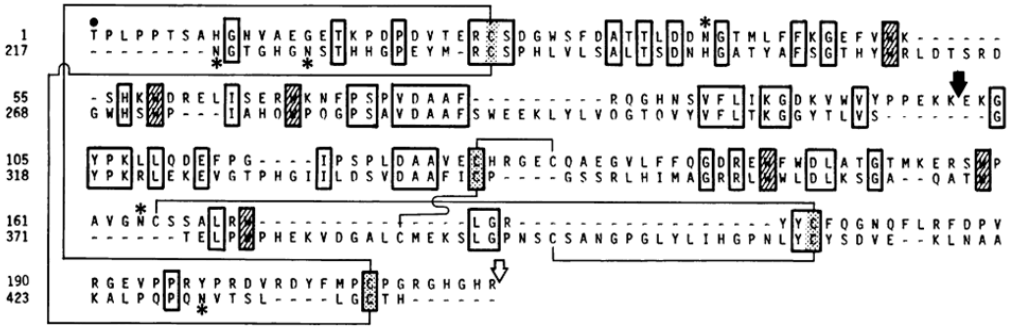

Figure 6: Internal homology of human hemopexin. Residues that appear in equivalent positions in each segment of the protein are enclosed in solid boxes. Gaps have been inserted to maximize the homology. Conserved cysteines and tryptophans are shaded, and the disulphide bonds are shown. The attachment sites for galactosamine and glucosamine oligosaccharides are indicated by . and * respectively [4].

in amino acid sequence. Figure 6 shows that hemopexin exhibits internal duplication. Hemopexin comprises of two homologous domains of about 200 residues each, joined by a 20-residue linker.

Crystal structure of isolated hemopexin C-terminal domain shows that it has a unique four-bladed β-propeller fold which is similar to those found in many other proteins like collagenases, clathrin and integrins. [50] Such domains are of importance in many biological processes especially as mediators of protein-protein interaction. [51] The N-terminal and C-terminal domains (residues 1–208 and 228–435) both have the same four bladed β-propeller fold, resembling two disks that lock together at a 90°angle, with the face of the N-domain packing against one edge of the C-domain. The connecting loop between the two domains provides structural features that form the heme binding site and mediate domain

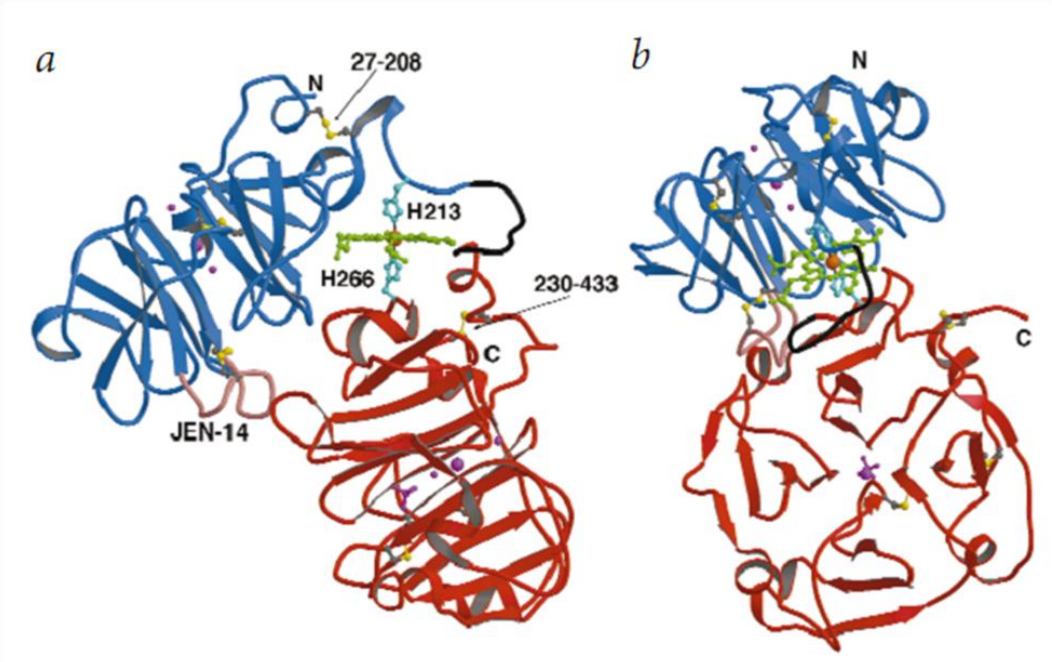

Figure 7: Structure of the heme–hemopexin complex. (a) and (b) are orthogonal views. The β-propeller domains are colored blue (N-domain) and red (C-domain), the heme group green, coordinating histidine residues cyan and disulfide bonds yellow [3].

association. This linker plays a critical part in creation of the binding site, forming its outer boundary and interacting with the heme to help stabilize the complex.

The linker has no significant secondary structure and its conformation is determined by the way it wraps around the heme. Two histidines coordinate the heme iron: His213, from the linker peptide, and His 266, from a loop of the C-terminal domain, giving a stable bis-histidyl Fe(III) complex, with planar heme and regular geometry Figure 7 shows the heme-hemopexin complex.

3.2.2 Functions:

Turnover of heme proteins, like hemoglobin, results in release of heme into extracellular fluids. Free heme has severe consequences for health because free heme is a major source of iron for invading bacterial pathogens and is highly toxic because of its ability to catalyze free radical formation. Heme mediated oxidative stress thus contributes to aging and various pathological states, including stroke and Alzheimer's disease.[52] Protection against free heme is provided by hemopexin. The high affinity of hemopexin for heme (K_d < 1pM) and the presence of specific receptors for the heme–hemopexin complex on hepatocytes and other cell types make hemopexin a candidate for the transport of heme to specific sites of catabolism. Once bound to hemopexin, heme is only released upon interaction with a specific surface receptor on liver cells where hemopexin is internalized by receptor mediated endocytosis. Internalization of heme–hemopexin complex further stimulates intracellular protective mechanisms, including the induction of heme oxygenase (HO), and other signaling pathways. Heme is catabolized by HO that opens the porphyrin ring producing biliverdin and carbon monoxide, and releasing iron that can be bound by ferritin. Ferritin is

the major intracellular depot of non metabolic iron. It is a multimeric protein composed of 24 subunits of two types (heavy chain and light chain), and has a very high capacity for storing iron (up to 4500 mol of iron per mol of ferritin). Together, HO and ferritin allow rapid shifting of iron from heme into the ferritin core where iron is less available to catalyze deleterious reactions. Thus heme binding and transport by hemopexin provides extracellular and intracellular protection against free heme oxidative damage, limits access by pathogens to heme and conserves iron by recycling the heme iron.[53]

3.2.3. Clinical significance:

In adults, serum hemopexin concentration range between 0.40-1.50 g/l. Hemopexin is synthesized in the fetus and the serum concentrations in newborn babies are ~20% of the adult values and start rising immediately after birth.[54]

Changes in hemopexin concentration have been correlated with many clinical conditions. Following intravascular hemolysis, free hemoglobin is complexed by circulating hemopexin which has a hemoglobin binding capacity between 0.7-1 g/l depending on the phenotype[55] Plasma hemopexin concentrations decrease in chronic or severe hemolysis. Mechanical trauma resulting from physical training and long distance jogging may also lead to enough hemolysis to deplete circulating hemopexin even in trained athletes. Lower serum concentrations of hemopexin during pregnancy and with menopause have been reported

in conditions of pre-eclampsia [56], [57] and in postmenopausal women treated with estrogens.[58] In the absence of hemolysis, decreased hemopexin concentrations are observed in chronic liver diseases due to decrease in synthesis of the protein [59], in active porphyria cutanea tarda due to massive overproduction of porphyrins [60], and in severe malabsorption. Low hemopexin concentrations can also be found in hemorrhagic pancreatitis which is attributable to the breakdown of erythrocytes in the peritoneal cavity followed by the release of hemin into the circulation.[54]

Hemopexin is a glycoprotein. Glycoproteins have been shown to correlate with progression of several cancers including prostate cancer [61], colorectal cancer [62], and breast cancer. [63] Hemopexin also has been shown to change in its structure specially glycosylation in several cancers. Glycosylation changes also correlate with liver diseases, so serum glycoprotein such as hemopexin promises to be a potential biomarker of hepatocellular carcinoma (HCC) which often develops on the background of cirrhotic liver. Studies have shown that branching α-1,3-fucosylated multiantennary glycans on hemopexin were increased in case of HCC patients when compared to another group of cirrhosis patients without HCC, fibrosis patients and healthy individuals. [64] In case of pancreatic cancer it was shown that both sialylation and fucosylation increased specifically in case of cancer patients compared to patients with chronic pancreatitis without cancer and healthy individuals. [65] Glycosylation changes in tumor-secreted proteins may also reflect fundamental changes in enzyme levels (or enzyme activities) involved

in the glycosylation pathway. Thus, study of hemopexin glycosylation variation may also lead to the identification of disease associated glycan alterations and new diagnostic markers in pancreatic cancer as well as in other types of cancer.

3.3 Tetranectin:

Tetranectin (TN) is a plasminogen binding protein which binds specifically to its kringle 4 domain (Kd = 0.5μM). [66] It also interacts with sulfated polysaccharides, fibrin and apolipoproteins. [67], [68], [69] Tetranectin is found is a many tissue types in human body. It was first isolated from human plasma, and is abundant in various cell types including monocytes, neutrophils, fibroblasts and osteoblasts. It is also found in the extracellular matrix of several carcinomas.[70]

3.3.1 Structure:

Human TN is a homotrimeric protein. Each of the monomer polypeptide chains consist of 181 amino acid residues. Three disulfide bonds connect Cys-50 to Cys -60, Cys-77 to Cys-176, and Cys-152 to Cys-168. Each monomer of TN consists of a carbohydrate recognition domain (CRD) (residues 50-181) connected to a long alpha helix. Each monomer has two distinct regions, one consists of six β-strands and two α-helices and the other region is composed of four loops having two calcium ions. The calcium ion at site 1 forms an eight fold coordinated complex and has Asp 116, Gluc 120, Gly147, Glu150, Asn151, and one water molecule as ligands. The calcium ion at site 2 is believed to be involved in

recognition and binding of oligosaccharides. This calcium ion forms a seven fold coordinated complex with Gln143, Asp145, Glu150, Asp165, and two water molecules as ligands. [5] The overall structure of a TN monomer is shown in Figure 8. Sequence homology studies and TN's calcium dependent binding to specific sugars, which is also a feature of a class of proteins called C-type (calcium dependent) lectins, groups it into the same class. Other C-type lectins include pancreatic stone protein, sea raven antifreeze protein and snake venom botrocetin.

3.3.2 Function:

Researchers have been studying tetranectin over years and these studies have led to our understanding of its amino-acid sequence, 3-D structure and corresponding gene sequence, characterization of its CRD, interaction with certain macromolecules and its expression in different cells and tissues but still its exact physiological role is not elucidated. TN resembles mannose binding proteins (MBP) from rat serum and also human MBP in its trimeric structure. As shown in Figure 8, the monomer is characterized by a long α-helix which forms the neck region. Trimerization of TN and these MBPs are formed with coiling of the α-helical coils of the monomers forming an α-helical coiled coil. The orientation of the CRDs with respect to the coiled coil stem determines the spatial separation of the carbohydrate-binding sites. This spatial separation differs in different proteins with CRD and holds importance in determining the specificity of the proteins. Since TN has a CRD that binds specific carbohydrates and also has a domain

within residues 1-49 which is responsible for its binding to kirngle 4 domian of plasminogen, one of its functions is to target plasminogen to specific carbohydrate

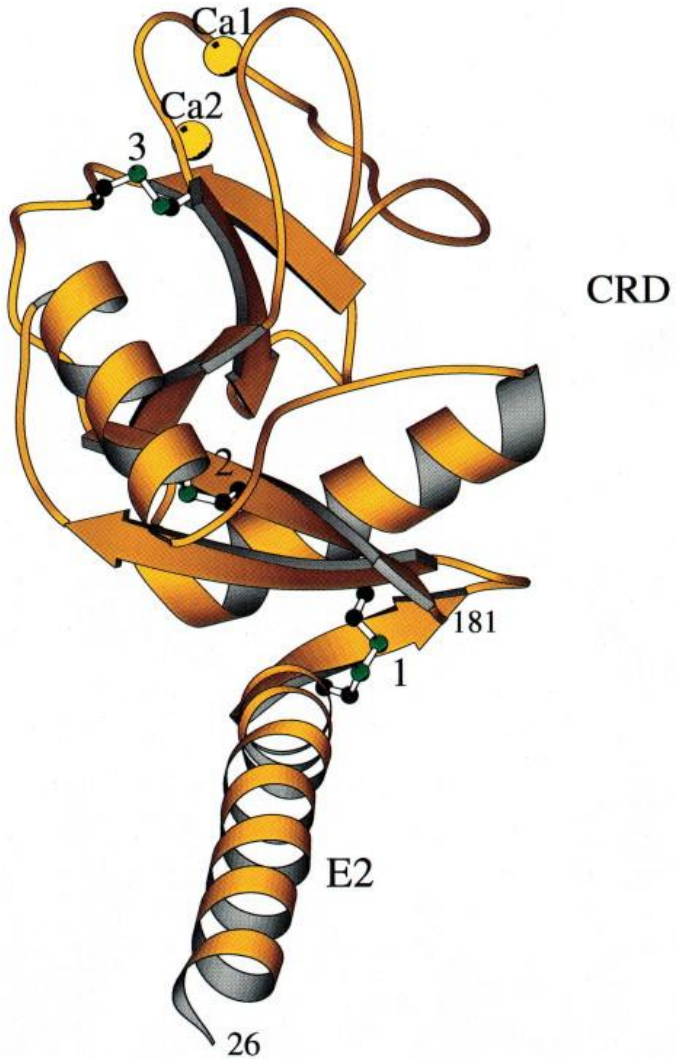

Figure 8: Overall structure of TN monomer. Two portions consisting of a long α-helix (represented as E2) and a carbohydrate binding domain (represented as CRD) are shown. The two calcium ions are shown as yellow spheres. The three disulfide bonds between residues 50-60, 77-176 and 152-168 are shown by ball and stick representation and labled 1, 2,3 [5].

ligands on cell surfaces in extracellular matrix or to fibrin. The fact that TN is a part of fibrinolytic system and also found abundantly in plasma and in various

tissues implies that it serves an important functional role in fundamental biological processes like tissue degradation, cell migration, and the spreads of cancer by metastasis and invasion. TN may also play an important role in mineralization during osteogenesis. [71]

3.3.3 Clinical Significance

Plasma level of tetranectin in health adults is ~100 nM. [72] This concentration has been found to change in case of various carcinomas. [73] TN levels have been found to decrease in case of ovarian cancer [74], colonic cancer [75] and also in rheumatoid arthritis. [76] TN is also present in stroma of various cancers including breast, colon and ovary whereas it is absent in the normal non cancerous tissues. [77] TN is also found to distribute along with plasminogen in a similar fashion at the invasive front of the cutaneous melanoma lesions. [78] Though the mechanism of how TN affects cancer progression is not very clear these are enough evidence to show it being an important protein which may be more generally associated with a variety of cancers.

Chapter 4

EXPERIMENTAL METHODS

4.1 Affinity tip preparation:

Pipette tips with carboxylic acid functionalized porous glass frits were ordered from Intrinsic Bioprobes. Inc. The obtained tips were treated with 0.2M hydrochloric acid to remove sodium/potassium salts and generate free carboxyl groups. Subsequently the tips were washed thoroughly with acetone and vacuum dried. The dried tips were activated with CDI (1,1'- Carbonyldiimidazole) by repeated rinse with 50g/l solution of CDI in 1-methyl-2-pyrrolidinone (NMP) for approximately 45min with the aid of a Beckman Multimek 96 channel automated pipettor robot. After CDI activation, affinity tips were blotted to remove excess NMP followed by two NMP rinses and blotting to remove excess CDI.

For IGF-1 analysis, these CDI activated tips were then coupled with rabbit monoclonal IGF-1 antibody (0.25mg/ml) obtained from R&D systems (Catalogue# MAB291) in MES buffered saline, pH 4.7 by repeated rinsing for a total exposure time of 50min with the help of pipettor robot.

For Hemopexin analysis, the CDI activated tips were coupled with Goat anti-Human Hemopexin antibody from Mybiosource, LLC (Catalogue# MBS686023) diluted in MES buffered saline to a final concentration of 0.05mg/ml, pH 4.7 by

repeated rinsing for a total exposure time of 50 min with the help of pipettor robot.

For Tetranectin analysis, the CDI activated tips were coupled with mouse anti-Tetranectin antibody from Thermo Scientific (Catalogue# 5B7) diluted in MES buffered saline to a final concentration of 0.05mg/ml, pH 4.7 by repeated rinsing for a total exposure time of 50 min with the help of pipettor robot.

After antibody coupling the affinity tips were rinsed with 1M ethanolamine pH 8.5 to block any vacant CDI-preactivated sites followed by two rinses with 10mM HBS-N. Affinity tips were then stored at 4^0C to be used within 3 months time.

4.2 Mass Spectrometric Immunoassay:

4.2.1 IGF-1:

Prior to extraction, 60µl of HEPES buffered saline (HBS-EP) and 60µl of 0.5% SDS (w/v) were added to 40µl plasma followed by 10 minutes of incubation at 32°C. After incubation an additional 840µl of HBS-EP was added to the solution for a total analytical volume of 1 ml. IGF-1 extraction was done with the aid of Beckman Multimek robot by repeatedly (50 repetitions) drawing and expelling (back into the analytical volume) 200µl aliquots of the analytical volume through an anti-IGF affinity pipette tip. After extraction, pipette tip was rinsed in the respective order using HBS-EP, H_2O, Tris HCl (100mM, pH 4.6), H_2O, acetonitrile:water (20:80; v/v), H_2O. Each rinse was accompanied with 5

repetitions of 200µl. Then the pipette tip was vacuum dried at room temperature and IGF-1 was eluted by drawing 4µl of MALDI matrix (saturated aqueous solution of α-cyano-4-hydroxycinnamic acid (ACCA), in 33% (v/v) acetonitrile, 0.45% (v/v) Trifluoroacetic acid (TFA) into the pipette and depositing on to the MALDI target.

The target plate was inserted in Bruker autoflex instrument for MALDI-TOF MS operating in the positive ion, delayed extraction mode; linear mode with 'ion source 1' at 20KV, 'ion source 2' at 18.5KV, lens at 7.00KV, 100ns after laser pulse delay extraction, deflection signal suppression up to m/z 2000. Eight thousand laser shots were signal averaged for each mass spectrum. Spectra was calibrated with mixture of 4 proteins supplied by Bruker (Cat No: 208241) which consisted of, in average mass, *m/z* 5743.52 (Insulin $[M+H]^+$), 6181.05 (Cytochrome-C $[M+2H]^{2+}$), 8565.76 (Ubiquitin-I $[M+H]^+$), and 12360.97 (Cytochrome-C $[M+H]^+$). Individual mass spectra were baseline subtracted and smoothed.

4.2.2 Hemopexin:

One hundred microliters of human plasma was mixed with 100µl of HEPES buffered saline (HBS-N) and Hemopexin was extracted by repeatedly (50 repetitions) drawing and expelling (back into the analytical volume) 80µl aliquots of the analytical volume through an anti-Hemopexin pipette tip. HBS-N, H_2O, 2M ammonium acetate:acetonitrile (3:1), H_2O rinse (in this order, each rinse was 10

repetition of 150µl) was followed by the extraction. After which the Hemopexin was eluted with the assistance of MALDI matrix (saturated aqueous of sinapinic acid, in 33% (v/v) acetonitrile, 0.45% (v/v) Trifluoroacetic acid (TFA)) on the MALDI target. The samples were analyzed with MALDI-TOF MS. The best obtained results were further analyzed with ESI-TOF MS.

ESI-TOF MS analysis: After hemopexin extraction and rinsing of the tip, elution was accomplished by air drying of the pipette and then 5µl of a mixture containing 0.4% Triflouroacetic acid in water and acetonitrile solution (90:10 v/v) was drawn and mixed over the pipette affinity capture frit for 30 sec and then dispensed into a 96-conical well polypropylene autosampler tray. Eluted sample was injected into the LC-TOF-MS within 10 min of elution.

Samples were injected by a Spark Holland Endurance autosampler in a microliter pick up mode and loaded by an Eksigent nano LC*1D at 10µl/min (90/10 water/acetonitrile containing 0.1% formic acid , solvent A) onto a protein captrap. After 2 min, the divert valve position was automatically toggled and flow over the cartridge changed to 1µl/min solvent thus running directly to ESI inlet. Over 5min the solvent gradient was ramped from 10% acetonitrile to 90% acetonitrile to elute the intact protein into the mass spectrometer.

4.2.3 Tetranectin

Forty microliters human plasma sample was incubated with 120µl of HEPES buffered saline (HBS-EP) containing 0.2% SDS and 10mM NOG for 15 minutes

at 36°C. After this incubation HBS-EP was added to the sample for a total analytical volume of 1ml. Tetranectin was extracted from the diluted sample with the aid of Beckman Multimek robot by repeatedly (50 repetitions) drawing and expelling (back into the analytical volume) 200µl aliquots through an anti-tetranectin affinity pipette tip. The affinity tip loaded with the protein was then rinsed in sequence with HBS-EP and water using 200µl each time for ten repetitions. MALDI matrix (saturated aqueous solution of sinapinic acid, in 33% (v/v) acetonitrile, 0.45% (v/v) Trifluoroacetic acid (TFA)) was used for the elution of tetranectin on to the MALDI target.

The target plate was inserted in Bruker ultraflex instrument for MALDI-TOF-MS operating in positive ion, delayed extraction mode; linear mode with 'ion source 1' at 25KV, 'ion source 2' at 23.5KV, lens at 9.00KV, 90ns after laser pulse delay extraction, deflection signal suppression up to *m/z* 4000. Ten thousand laser shots were signal averaged for each mass spectrum. Spectrum was calibrated with mixture of eight proteins supplied by Bruker. Individual mass spectra were baseline subtracted and smoothed.

Chapter 5

RESULTS AND DISCUSSIONS

5.1 MSIA development for IGF-1:

IGF-1 associates with IGF binding proteins and circulates in the plasma in the form of a complex. One of the challenges in developing assay for IGF-1 is to be able to disrupt the complex and extract IGF-1. SDS was used to disrupt this complex at a concentration 0.2% which was further diluted to 0.03% so as not to interfere with the extraction of the IGFs. Figure 9A shows the mass spectra where the IGF complex was not disrupted. It was obtained with a plasma sample diluted with HBS-EP and untreated with SDS. In Figure 9B the spectra is obtained from a

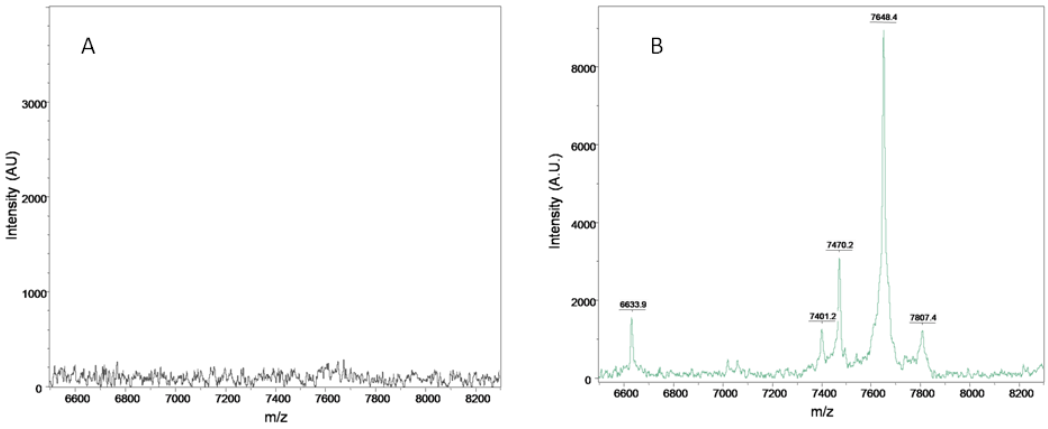

Figure 9: Result of IGF-IGFBP complex disruption. (A) MSIA analysis spectra of HBS-EP diluted plasma samples without any complex disruption. (B) Spectra obtained after SDS treatment of the plasma sample for complex disruption.

sample treated with SDS to break the IGF complex.

The untreated sample does not show any peaks corresponding to IGF. In the SDS treated sample various peaks are visible. Peaks at *m/z* = 7648.4 and 7470.2 correspond to IGF-1 and IGF-2 respectively. The peak obtained at *m/z* = 7401.2 corresponds to a variant of IGF-2 (des(Ala)-IGF-2). The calculated molecular weights of these proteins are 7648.7, 7469.4 and 7398.3 respectively. The peak observed at *m/z* = 6633.9 is nonspecifically retained apolipoprotein protein C-1. In order to remove non-specifically retained protein, the affinity tip was rinsed with a detergent solution consisting of 4.5% Tween 20, 150mM Octyl-β-pyranoside, 1.5M ammonium acetate in 10x concentrated PBS. Though this rinse removes nonspecifically retained apolipoprotein C-1 as seen in Figure 10, it increases the noise level in the spectra which diminishes the overall quality. So, it was decided not to incubate the plasma sample with the detergent mixture along

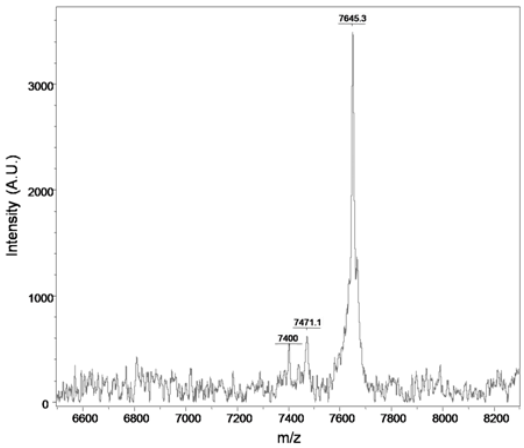

Figure 10: MSIA spectra obtained after detergent rinse of the affinity tip following extraction.

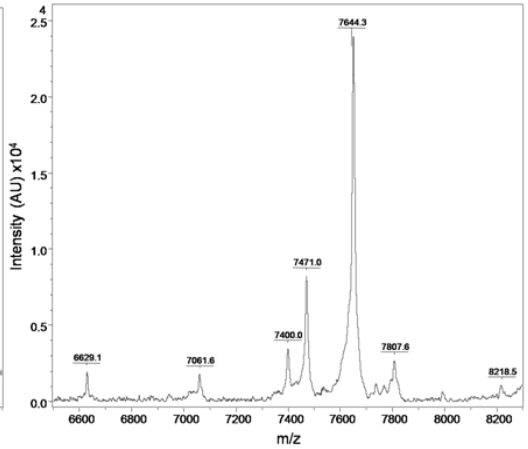

Figure 11: MSIA analysis spectra obtained after TRIS-HCl rinse of the affinity tip after protein capture.

with SDS and rather use Tris-HCl buffer as a rinsing agent to improve the spectra and explore the possibility of removing the nonspecific binding. Figure 11 shows the spectrum obtained after Tris-HCl rinse of detergent incubated sample extraction. Though nonspecifically retained apolipo-protein is still present the signal to noise ratio of the spectrum is considerably improved and a peak at m/z = 7061.6 corresponding to another apolipoprotein C-1' is visible.

This MSIA for IGF-1 was used to analyze sample from type 2 diabetic (T2D) patients in hope of being able to observe any variants of the protein. The sample that was used clearly showed bifurcation of the the IGF-1 peak at m/z 7648 (Figure 12). The mass difference between the bifurcated peaks is 16Da which is attributed to oxidation of methionine residue as it forms methionine sulfoxide with additional oxygen in the molecule. Though this result is not a conclusive evidence which can attribute methionine oxidation to T2D but it clearly

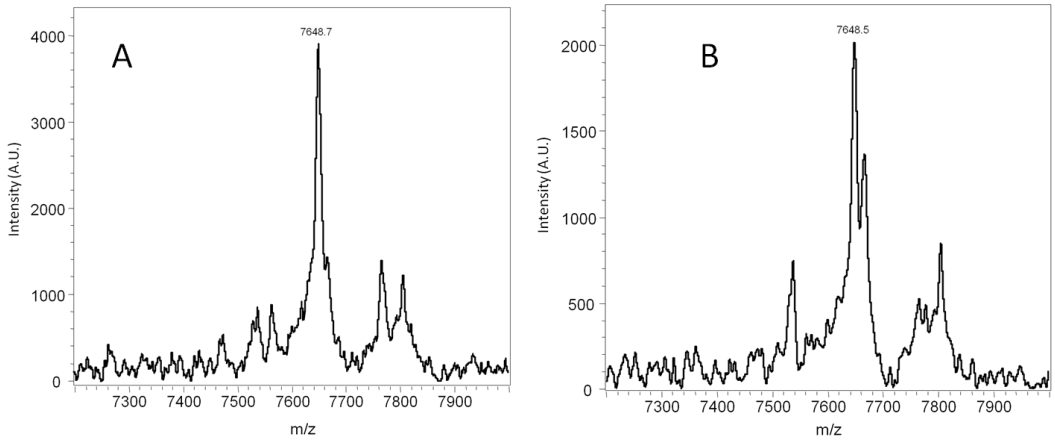

Figure 12: Analysis of type 2 diabetic (T2D) sample. (A) MSIA analysis spectra of healthy sample showing the peak corresponding to IGF-1. (B) Spectra obtained from T2D sample analysis which show a split peak of IGF-1 indicating oxidation of methionine residue.

demonstrates that usefulness of the present MSIA developed in capturing variations in the protein structure and promises to be useful in analysis of large cohorts for statistically meaningful studies.

5.2 Development of a MSIA for Hemopexin:

Conventional MALDI-TOF-MS lacks in its capability to resolve smaller but biologically meaningful mass shifts at *m/z* ratios greater than about 30,000. Hemopexin is 60kDa in size which is beyond this optimal range for analysis by MALDI. Electrospray ionization (ESI) is the preferred platform as it increases the charge state of intact proteins and provides higher resolution. In order to analyze samples by ESI-TOF-MS the primary assay was first developed on MALDI platform because of its ease of use and accessibility. The initial assay was optimized for plasma dilution and antibody concentration. One hundred micro-liters of plasma diluted in 300µL of HBS-N buffer was extracted with an affinity tip incubated in 0.1g/L of anti-Hemopexin antibody and rinsed in order with HBS-N, water, a 3:1 2 M ammonium acetate to acetonitrile mixture, then waterFigure 13 shows the MALDI-TOF spectrum obtained from this initial assay. A peak at *m/z* = 60441 corresponding to hemopexin is visible along with peak at m/z = 66552 which corresponds to nonspecifically retained serum albumin. Several attempts were made to eliminate this nonspecific binding of albumin which affects the sensitivity of the assay towards hemopexin. Plasma sample was

incubated in the previously reported detergent mix (150mM NOG, 4.5% tween20, 1.5M ammonium acetate, PBS)..

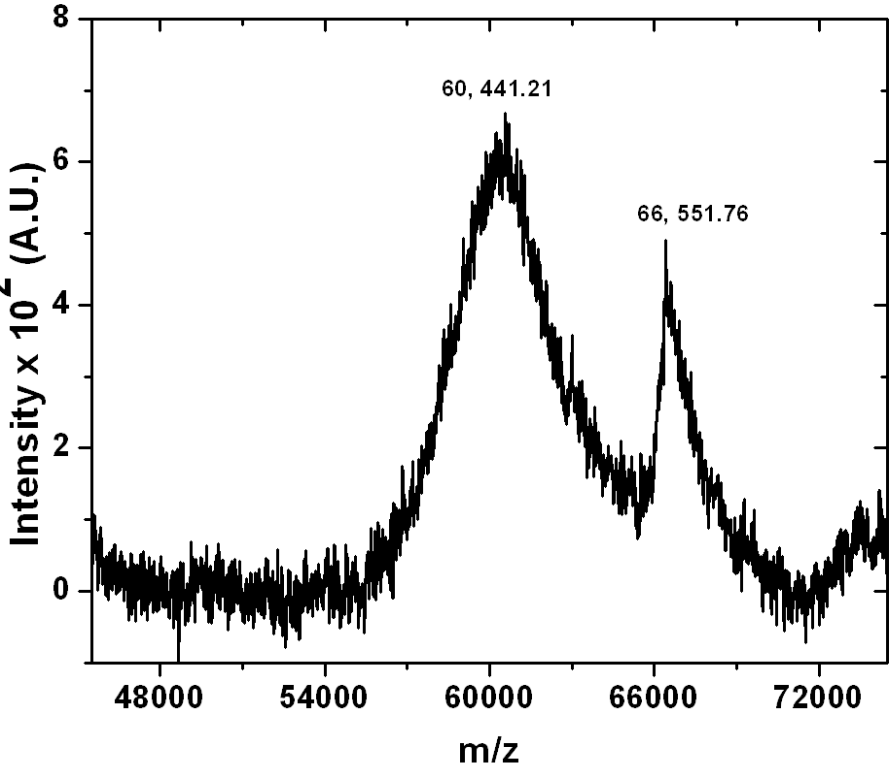

Figure 13: MSIA spectra obtained from HBS-N diluted plasma sample with anti-hemopexin antibody.

Thus, in a different approach the affinity tips were pre-incubated overnight at 4^0C in 60mM HCl in an effort to remove any non-specific binding. The spectrum obtained after this pre-incubation in 60mM HCl with a sample not-treated with the detergent did not show any peak for hemopexin. Further, detergent treated plasma sample was extracted with HCl pre-incubated tips but was not successful in avoiding nonspecifically retained albumin. So it was decided to use the initially

optimize the assay for analysis with ESI-MS. This was found to be the optimal condition to generate a clean spectrum of hemopexin. Samples prepared with these conditions were then analyzed by the ESI-MS as shown in Figure 14. The ESI-MS spectrum clearly shows different glcosylated forms of the hemopexin.

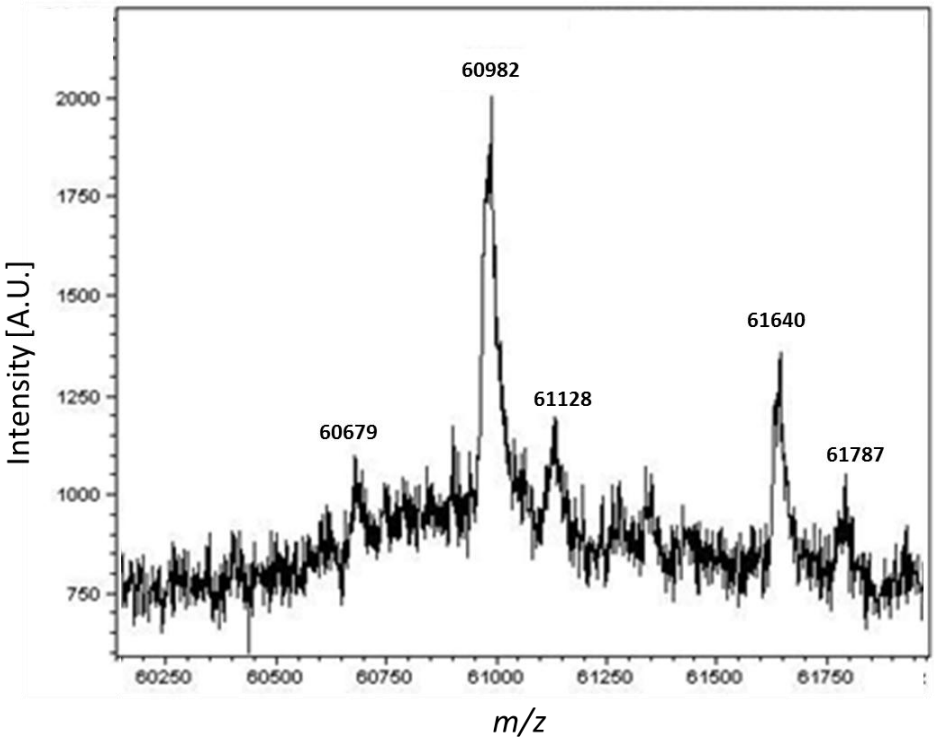

Figure 14: MSIA spectra obtained with ESI-MS platform shows distinct peaks for several glycosylated forms of hemopexin.

Table 1: Observed ESI mass spectrum peaks and their interpretations for MSIA analysis of Hemopexin

Measured Mass (Da)	Calculated Mass (Da)	Mass difference (Da)	Peak separation (Da)	Possible Modification	Calculated mass difference (Da)

60679	49294	11385		$(HexNac)_{21}$, $(Hex)_{26,}$ $(NeuAc)_{10}$	11394.63
60982		11688	303	+ NeuAc	11685.88
61128		11834	146	+ Deoxyhexose	11848.02
61640		12346	512	+ HexNac +NeuAc	12342.46
61787		12493	146	+ Deoxyhexose	12504.60

5.3 Development of a MSIA for Tetranectin:

Since tetranectin is a plasminogen binding protein several attempts were made to free TN from PLG. Figure 15 shows results from the use of several detergents for this purpose. Initially the plasma sample was incubated in 30mM NOG before extraction of TN with anti-TN functionalized affinity tip. This however did not lead to any detectable peak in the mass spectrum (Figure 15 A). After this, two different detergent mixes were prepared Mix1 (150mM NOG, 1.5 M ammonium acetate, Concentrated PBS consisting of 0.67 M sodium phosphate and 1M NaCl) and Mix2 (150mM NOG, 1.5 M ammonium acetate, PBS and 4.5% Tween) successively increasing the detergent strength. After incubation of the sample in these detergent mixes and MALDI analysis after subsequent extraction the MSIA spectrum as shown in Figure 14 B,C show visible peaks around 20504 Da and 20795 Da which are reported modified peaks for TN [79]. Further, detergent mix 3 (HBS-EP with 0.2% SDS and 10mM NOG) was prepare used `for sample

incubation followed by subsequent extraction and analysis. The MSIA spectrum obtained as shown in Figure 14 D clearly shows four different variations of the protein with peaks at m/z 20, 538.79 Da, 20, 825.29 Da, 21, 027.97 Da and 21, 118.3 Da. These peaks are consistent with variations of TN reported earlier in the literature with a new peak at m/z 21, 027.97 which has not been reported before. The mass shift in the peaks from the unmodified mass of the protein indicates modifications with an oligosaccharide prosthetic group containing terminal sialic acid residue. Table 1 summarizes the peaks observed and their possible

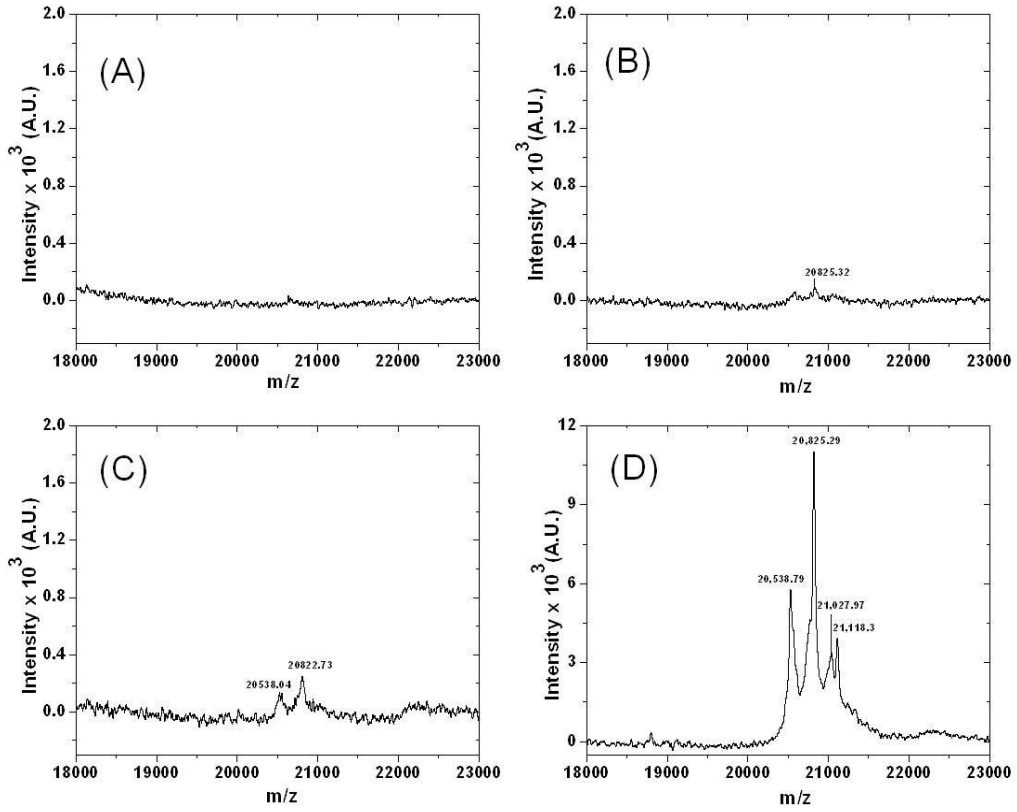

Figure 15: Use of different detergent in order to free TN from PLG. Plasma sample was incubated with indicated concentration of the detergent (A) 30mM NOG (B) detergent mix 1 (C) detergent mix 2 (D) 0.25% SDS

interpretations.

Table 2: Observed MALDI mass spectrum peaks and their interpretations for MSIA analysis of TN

Measured Mass (Da)	Calculated Mass (Da)	Mass difference (Da)	Peak separation (Da)	Possible Modification	Calculated mass difference (Da)
20, 538.79	20, 162.8	375.99		HexNac, Hex	365.33
20, 825.29		662.49	286.5	+ NeuAc	656.58
21, 027.97		865.17	202.68	+ HexNac	859.77
21, 118.3		955.5	90.33	- HexNac, + NeuAc	947.83

Chapter **6**

CONCLUSIONS AND FUTURE DIRECTIONS

In conclusion, mass spectrometric immuno assays for three different proteins have been developed. These proteins viz. IGF-1, hemopexin and tetranectin have been found in reported literature to show correlations with many diseases along with several carcinomas. Though structural characterization of these proteins has already been reported but still cohort studies involving large population has not been done. To gauge the role of these proteins in diseases these studies hold the key. The MSIAs developed during this thesis work provide easy, efficient and affordable method for structural characterization of these proteins enabling high throughput analysis in large number of samples which is not conveniently doable with other protein extraction techniques. Developed assays have shown their capability to capture and reveal different structural forms of target proteins divulging their structural micro-heterogeneity invisible to commonly used immunoassays. With well designed large population studies one or more of these structural variants alone or in combinations could be established as biomarkers of certain diseases.

In future these assays will be used to study qualitative variations in structure of these proteins in plasma samples from donors with diseases of interest in a small cohort of about hundred subjects. Contingent upon promising statistical

differences in small cohort study bigger and more diverse cohort study will follow with an aim to further characterize the structural variations based in gender, age, ethnicity etc. These calcifications will result in better understanding of protein diversity in population. Further, quantitative assays will also be developed with the help of suitable internal reference standards. Multiplex assay using more than one protein involved in same cancer in order to obtain a more robust analysis to avoid false positives can also be developed. These assays hopefully will reveal some properties either structural or concentration differences which can be established and used as biomarkers for disease diagnosis and/or monitoring.

REFERENCES

1. Vajdos, F.F., et al., *Crystal structure of human insulin-like growth factor-1: detergent binding inhibits binding protein interactions.* Biochemistry, 2001. **40**(37): p. 11022-11029.

2. Narhi, L.O., et al., *Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: Genetic models of protein-folding intermediates.* Biochemistry, 1993. **32**(19): p. 5214-5221.

3. Paoli, M., et al., *Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two β-propeller domains.* Nature Structural & Molecular Biology, 1999. **6**(10): p. 926-931.

4. Takahashi, N., Y. Takahashi, and F.W. Putnam, *Complete amino acid sequence of human hemopexin, the heme-binding protein of serum.* Proceedings of the National Academy of Sciences, 1985. **82**(1): p. 73.

5. Nielsen, B.B., et al., *Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an [alpha]-helical coiled coil.* FEBS letters, 1997. **412**(2): p. 388-396.

6. Available from: http://hapmap.ncbi.nlm.nih.gov/.

7. Nedelkov, D., et al., *Investigating diversity in human plasma proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(31): p. 10852.

8. *AMA.* Available from: http://www.ama-assn.org/ama/pub/physician-resources/medical-science/genetics-molecular-medicine/current-topics/proteomics.page.

9. Atkinson, A.J., et al., *Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework*.* Clinical Pharmacology & Therapeutics, 2001. **69**(3): p. 89-95.

10. Polanski, M. and N.L. Anderson, *A list of candidate cancer biomarkers for targeted proteomics.* Biomarker Insights, 2006. **1**: p. 1.

11. *Cancer Society*. Available from: http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-029771.pdf.

12. Wong, S.C.C., et al., *Advanced proteomic technologies for cancer biomarker discovery.* Expert review of proteomics, 2009. **6**(2): p. 123-134.

13. Dunn, B.K., et al. *Molecular markers for early detection*. 2010: Elsevier.

14. Ludwig, J.A. and J.N. Weinstein, *Biomarkers in cancer staging, prognosis and treatment selection.* Nature Reviews Cancer, 2005. **5**(11): p. 845-856.

15. Nelson, R.W., et al., *Mass spectrometric immunoassay.* Analytical chemistry, 1995. **67**(7): p. 1153-1158.

16. Anderson, N.L., et al., *The human plasma proteome: a nonredundant list developed by combination of four separate sources.* Molecular & cellular proteomics: MCP, 2004. **3**(4): p. 311.

17. LeRoith, D., et al., *Insulin-like growth factors in health and disease.* Annals of internal medicine, 1992. **116**(10): p. 854-862.

18. Strasburger, C.J., et al., *Normal values of insulin-like growth factor I and their clinical utility in adults.* Hormone Research in Paediatrics, 2001. **55**(2): p. 100-105.

19. Park, P. and P. Cohen, *Insulin-like growth factor I (IGF-I) measurements in growth hormone (GH) therapy of idiopathic short stature (ISS).* Growth hormone & IGF research, 2005. **15**: p. 13-20.

20. Rinderknecht, E. and R.E. Humbel, *The amino acid sequence of human insulin-like growth factor I and its structural homology with proinsulin.* Journal of Biological Chemistry, 1978. **253**(8): p. 2769-2776.

21. Holzenberger, *IGF-I signaling and aging.* Hormone and metabolic research, 2004. **39**.

22. Scott, C., *The Role of the M6P/IGF-II Receptor in Cancer: Tumor Suppression or Garbage Disposal?* Hormone and metabolic research, 2004. **36**(5): p. 261.

23. Monzavi, R. and P. Cohen, *IGFs and IGFBPs: role in health and disease.* Best Practice & Research Clinical Endocrinology & Metabolism, 2002. **16**(3): p. 433-447.

24. Nunn, S.E., et al., *Regulation of prostate cell growth by the insulin-like growth factor binding proteins and their proteases.* Endocrine, 1997. **7**(1): p. 115-118.

25. Collett-Solberg, P.F. and P. Cohen, *Genetics, chemistry, and function of the IGF/IGFBP system.* Endocrine, 2000. **12**(2): p. 121-136.

26. Wetterau, L.A., et al., *Novel aspects of the insulin-like growth factor binding proteins.* Molecular genetics and metabolism, 1999. **68**(2): p. 161-181.

27. Conti, E., et al., *IGF-1 and atherothrombosis: relevance to pathophysiology and therapy.* Clinical Science, 2011. **120**: p. 377-402.

28. Moses, A., et al., *Recombinant human insulin-like growth factor I increases insulin sensitivity and improves glycemic control in type II diabetes.* Diabetes, 1996. **45**(1): p. 91.

29. Bach, M., E. Chin, and C. Bondy, *The effects of subcutaneous insulin-like growth factor-I infusion in insulin-dependent diabetes mellitus.* Journal of Clinical Endocrinology & Metabolism, 1994. **79**(4): p. 1040.

30. Asada, T., et al., *Recombinant insulin-like growth factor I normalizes expression of renal glucose transporters in diabetic rats.* American Journal of Physiology-Renal Physiology, 1997. **273**(1): p. F27-F37.

31. Kaino, Y., et al., *Insulin-like growth factor I (IGF-I) delays the onset of diabetes in non-obese diabetic (NOD) mice.* Diabetes research and clinical practice, 1996. **34**(1): p. 7-11.

32. Colao, *Circulation insulin like growth facor-1 level is independently associated with coronary artery disease progression in young male survivors of myocardial infraction.* J. Am. Coll. Cardiol, 2000. **35**(3).

33. Rhodes, C., *IGF-I and GH post-receptor signaling mechanisms for pancreatic beta-cell replication.* Journal of molecular endocrinology, 2000. **24**(3): p. 303-311.

34. Conti, E., et al., *Markedly reduced insulin-like growth factor-1 in the acute phase of myocardial infarction.* Journal of the American College of Cardiology, 2001. **38**(1): p. 26-32.

35. Conti, E., *Reduced levels of insulin-like growth factor-1 in patients with angina pectoris, positive exercise stress test, and angiographically normal epicardial coronary arteries.* The American journal of cardiology, 2002. **89**(8): p. 973.

36. Van den Beld, A., et al., *Endogenous hormones and carotid atherosclerosis in elderly men.* American journal of epidemiology, 2003. **157**(1): p. 25.

37. Rowlands, M.A., et al., *Circulating insulin-like growth factor peptides and prostate cancer risk: A systematic review and meta-analysis.* International Journal of Cancer, 2009. **124**(10): p. 2416-2429.

38. Ma, J., et al., *Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3.* Journal of the National Cancer Institute, 1999. **91**(7): p. 620-625.

39. Renehan, A.G., et al., *Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis.* The Lancet, 2004. **363**(9418): p. 1346-1353.

40. Larsson, S.C., N. Orsini, and A. Wolk, *Diabetes mellitus and risk of colorectal cancer: a meta-analysis.* Journal of the National Cancer Institute, 2005. **97**(22): p. 1679-1687.

41. Yu, H., et al., *Plasma levels of insulin-like growth factor-I and lung cancer risk: a case-control analysis.* Journal of the National Cancer Institute, 1999. **91**(2): p. 151-156.

42. Wakai, K., et al., *Serum Insulin-like Growth Factors, Insulin-like Growth Factor-binding Protein-3, and Risk of Lung Cancer Death: A Case-control Study Nested in the Japan Collaborative Cohort (JACC) Study.* Cancer Science, 2002. **93**(12): p. 1279-1286.

43. Hankinson, S.E., et al., *Circulating concentrations of insulin-like growth factor I and risk of breast cancer.* The Lancet, 1998. **351**(9113): p. 1393-1396.

44. Muti, P., et al., *Fasting glucose is a risk factor for breast cancer.* Cancer Epidemiology Biomarkers & Prevention, 2002. **11**(11): p. 1361-1368.

45. Mantzoros, C., et al., *Insulin-like growth factor 1 in relation to prostate cancer and benign prostatic hyperplasia.* British journal of cancer, 1997. **76**(9): p. 1115.

46. Wolk, A., et al., *Insulin-like growth factor 1 and prostate cancer risk: a population-based, casecontrol study.* Journal of the National Cancer Institute, 1998. **90**(12): p. 911-915.

47. Gunnell, *insulin-like growth factors and cancer risk.* Growth hormone & IGF research, 2000.

48. Schultze, H.E., K. Heide, and H. Haupt, *Charakterisierung von hochgereinigtem Hämopexin.* Naturwissenschaften, 1961. **48**(22): p. 696-697.

49. Hayem-Lévy, A. and R. Havez, *Isolement et étude de l'hémopexine humaine.* Clinica Chimica Acta, 1973. **47**(2): p. 113-122.

50. Faber, H.R., et al., *1.8 Å crystal structure of the C-terminal domain of rabbit serum haemopexin.* Structure, 1995. **3**(6): p. 551-559.

51. Smith, T.F., et al., *The WD repeat: a common architecture for diverse functions.* Trends in biochemical sciences, 1999. **24**(5): p. 181-185.

52. Smith, M.A., et al., *Amyloid-β Deposition in Alzheimer Transgenic Mice Is Associated with Oxidative Stress.* Journal of neurochemistry, 1998. **70**(5): p. 2212-2215.

53. Juckett, M., *Heme and the Endothelium. EFFECTS OF NITRIC OXIDE ON CATALYTIC IRON AND HEME DEGRADATION BY HEME OXYGENASE.* The Journal of biological chemistry, 1998. **273**(36): p. 23388.

54. Thomas, L., *Clinical laboratory diagnostics: use and assessment of clinical laboratory results*. 1998: TH-books.

55. Wuyts, B., et al., *A new method for fast haptoglobin phenotyping and hemoglobin binding capacity calculation based on capillary zone electrophoresis.* Clinical chemistry and laboratory medicine, 2000. **38**(8): p. 715-720.

56. Minchin Clarke, H., T. Freeman, and W. Pryse-Phillips, *SERUM PROTEINS IN NORMAL PREGNANCY AND MILD PRE-ECLAMPSIA.* BJOG: An International Journal of Obstetrics & Gynaecology, 1971. **78**(2): p. 105-109.

57. Studd, J., J. Blainey, and D. Bailey, *SERUM PROTEIN CHANGES IN THE PRE-ECLAMPSIA-ECLAMPSIA SYNDROME.* BJOG: An International Journal of Obstetrics & Gynaecology, 1970. **77**(9): p. 796-801.

58. Hashimoto, S., et al., *Changes in 40 serum proteins of post-menopausal women.* Maturitas, 1991. **13**(1): p. 23-33.

59. Vladutiu, A.O. and J.S. Kim, *Absence of beta-globulin band in the serum protein electropherogram of a patient with liver disease.* Clinical chemistry, 1981. **27**(2): p. 334.

60. Moran, M., et al., *Correlation between levels of free and protein-bound plasma porphyrin and urinary porphyrins in porphyria cutanea tarda.* The International Journal of Biochemistry & Cell Biology, 1995. **27**(6): p. 585-588.

61. Peracaula, R., *Glycobiology*. 2003.

62. Kasbaoui, L., et al., *Differences in glycosylation state of fibronectin from two rat colon carcinoma cell lines in relation to tumoral progressiveness.* Cancer research, 1989. **49**(19): p. 5317.

63. Ng, R., et al., *Analyses of protein extracts of human breast cancers: changes in glycoprotein content linked to the malignant phenotype.* British journal of cancer, 1987. **55**(3): p. 249.

64. Debruyne, E.N., et al., *Diagnostic value of the hemopexin N-glycan profile in hepatocellular carcinoma patients.* Clinical chemistry, 2010. **56**(5): p. 823.

65. Zhao, J., et al., *Glycoprotein microarrays with multi-lectin detection: unique lectin binding patterns as a tool for classifying normal, chronic pancreatitis and pancreatic cancer sera.* Journal of proteome research, 2007. **6**(5): p. 1864-1874.

66. CLEMMENSEN, I., L.C. PETERSEN, and C. KLUFT, *Purification and characterization of a novel, oligomeric, plasminogen kringle 4 binding protein from human plasma: tetranectin.* European Journal of Biochemistry, 1986. **156**(2): p. 327-333.

67. Clemmensen, I., *Interaction of tetranectin with sulphated polysaccharides and trypan blue.* Scandinavian journal of clinical & laboratory investigation, 1989. **49**(8): p. 719-725.

68. Kluft, C., P. Los, and I. Clemmensen, *Calcium-dependent binding of tetranectin to fibrin.* Thrombosis research, 1989. **55**(2): p. 233-238.

69. Kluft, C., et al., *Functional analogy between lipoprotein (a) and plasminogen in the binding to the kringle 4 binding protein, tetranectin.* Biochemical and biophysical research communications, 1989. **161**(2): p. 427-433.

70. Christensen, L. and I. Clemmensen, *Tetranectin immunoreactivity in normal human tissues.* Histochemistry and Cell Biology, 1989. **92**(1): p. 29-35.

71. Wewer, U.M., et al., *A potential role for tetranectin in mineralization during osteogenesis.* The Journal of cell biology, 1994. **127**(6): p. 1767-1775.

72. Jensen, B. and I. Clemmensen, *Plasma tetranectin is reduced in cancer and related to metastasia.* Cancer, 1988. **62**(5): p. 869-872.

73. Nielsen, H., et al., *Decreased tetranectin in multiple myeloma.* American Journal of Hematology, 1990. **33**(2): p. 142-144.

74. Høgdall, C., L. Christensen, and I. Clemmensen, *The prognostic value of tetranectin immunoreactivity and plasma tetranectin in patients with ovarian cancer.* Cancer, 1993. **72**(8): p. 2415.

75. Verspaget, H., et al., *Tetranectin expression in human colonic neoplasia.* Histopathology, 1994. **25**(5): p. 463-467.

76. Kamper, E., et al., *Plasma tetranectin levels and disease activity in patients with rheumatoid arthritis.* The Journal of rheumatology, 1997. **24**(2): p. 262.

77. Christensen, L. and I. Clemmensen, *Differences in tetranectin immunoreactivity between benign and malignant breast tissue.* Histochemistry, 1991. **95**(5): p. 427-33.

78. De Vries, T.J., et al., *Tetranectin and plasmin/plasminogen are similarly distributed at the invasive front of cutaneous melanoma lesions.* J Pathol, 1996. **179**(3): p. 260-5.

79. Jaquinod, M., et al., *Mass spectrometric characterisation of post-translational modification and genetic variation in human tetranectin.* Biological chemistry, 1999. **380**(11): p. 1307-1314.