CANCER BIOMARKER DISCOVERY: GLYCOMICS STUDY OF HUMAN SERUM

by

Wei Wang

(Under the Direction of Lance Wells)

ABSTRACT

A biomarker is a detectable characteristic change resulting from a certain stimulus. The stimuli can be tumor growth, virus invasion, drug usage, etc. Biomarker discovery is essential for cancer diagnostics as well as treatment. Glycosylation is an important post-translational modification. Aberrant glycosylation has been hallmark of malignant transformation. Recently efforts have been made to correlate glycosylation pattern changes with cancer progression. In this review, we make an extensive study of possible mass spectrometry associated glycomics methods in finding cancer biomarkers in human serum. Established methods, such as glycoprotein enrichment, glycosylation site mapping, and glycan analysis are discussed and compared. New mass spectrometry methods, such as Ion Mobility Mass Spectrometry, Multiple Reaction Monitoring, Electron Transfer Dissociation are introduced.

INDEX WORDS: cancer biomarker , serum, enrichment, glycosylation site mapping, Ion Mobility Mass Spectrometry, Multiple Reaction Monitoring

CANCER BIOMARKER DISCOVERY: GLYCOMICS STUDY OF HUMAN SERUM

By

Wei Wang

B.S. Wuhan University , P.R. China, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfullment of the Requirements for the Degree

Master of Science

ATHENS, GEORGIA

2008

CANCER BIOMARKER DISCOVERY: GLYCOMICS STUDY OF HUMAN SERUM

by

Wei Wang

Major Professor: Lance Wells

Committee: Michael Tiemeyer

Jonathan I Amster

Electronic Version Approved

Maureen Grasso

Dean of the Graduate School

The University of Georgia

December 2008

DEDICATION

To my wonderful parent

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my deepest thanks to my major professor Dr Tiemeyer, he is resourceful and helpful throughout the two years at his lab.

Also I like to thank lab members Kazuhiro ,Mindy and Mayumi for teaching me the sample preparation methods, they have been the most patient teachers I have ever met.

For Toshihiko, the software expert, who taught me using endnote and illustrator, without your help my thesis writing wouldn't be a pleasant one.

For Sarah Baas, who always keeps me company and helps with proofreading and giving a lot of suggestions for defense practice.

For Ashley, who keeps on inspiring me finishing up everything on time.

Finally, I would like to thank my parents, without your support I could not accomplish half what I did today.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Review: Cancer Biomarker Discovery: Glycomics Study of Human Serum Glycosylation Machinery

Protein post-translational modification occur in many kinds, phosphorylation, acetylation, glycosylation, methylation, and ubiquitylation. Some are reversible some occur in tandem, Glycosylation is an important type of post-translational modificaion. The most studied glycosylation forms are N-glycosylation and O-glycosylation.

In N-linked glycans biosynthesis, the branching glycan unit in Nglycoproteins is preassembled on a dolicol scaffold by a series of membrane-associated glycosyltransferases in the endoplasmic reticulum. The assembled N-glycan is flipped into the lumen of the endoplasmic reticulum and transferred to peptide by multisubunit oligosaccharyltransferase (OST). The biogenesis are the carboxamide nitrogen atoms of asparagine groups is almost always in the sequence Ser/Thr-X-Asn. The initial tetradecasaccharyl chain $Glc_3Man_9(GlcNAc)_2$ then undergoes a remarkable enzymatic hydrolytic trimming and branching in the ER and Golgi to form high mannose, hybrid and complex structures . The multiplicity of glycosyltransferases in the Golgi can create enormous diversity in the mature N-glycan chains. It has been estimated that about a third of all proteins that enter secretory pathways in eukaryotic cells may be N-glycosylated, and so tens of thousands of glycoprotein variants may coexist in eukaryotic cells.

O-glycosylation biosynthesis is less complex than N-glycosylation. Unlike N-glycosylation, O-glycosylation does not have a consensus sequence. The sugar residue is covalently linked to Ser/Thr residue on peptide chain. Many proteins contain the monosaccharide GalNAc that is put on by a specific O-GalNAc transferase, which helps to form Core structures.

These enzymes are localized in subcellular compartments including the ER and Golgi. So that is the initial sites of O-glycan formation.

Aberrant Glycosylation Associated with Cancer Progression.

Glycosylation requires a tightly regulated system of enzymes and transportors that mammalian cells employ to link to various types of oligosaccharide structures to proteins and lipids. It is important for cell signaling, autoimmunity, tumorogenesis, angiogenesis, etc. Over half of all proteins are glycosylated [1]. Current serum biomarkers, such as prostate-specific antigen (Prostate cancer biomarker), CA125(breast cancer biomarker) CA19-9(pancreatic cancer biomarker),are glycoproteins whose glycosylation state changes with tumor progression.

Glycans regulate a variety of cell functions, such as tumor proliferation, invasion, haematogenesis, metastasis and angiogenesis. Considering the functions of glycosylation, it is not surprising to see diseases associated with glycosylation abnormalities. Glycans on proteins often stick out on cell surfaces and are involved in cell-cell signaling via surface lectin sugar recognition. Changes in glycosylation may facilitate metastasis. Important glycosylation changes involved in tumor progression and invasion can be found this review. [2]. Among the most studied are expression levels of sialic acid and fucose, GnT-V product and cancer.

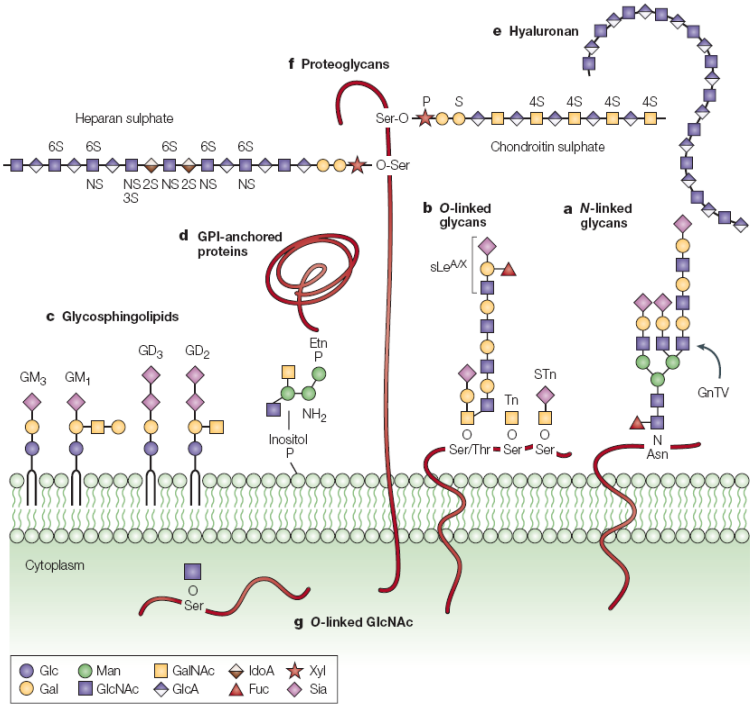


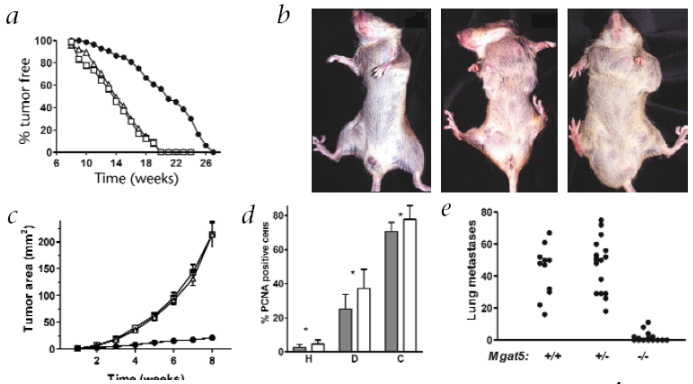
Figure 1 Reported glycoforms associated with cancer. In the graph are the possible glycoforms that has been reported to associate with cancer . For instance, they are reported upregulation of Lexis X and a structure, GNT-V product, etc (Mark M. Fuster, J. D. E., *Nature cancer review* 2005)

Increased expression of sialyated glycans are seen on many different cancer cell surfaces. This is due to elevated activity of sialyltransferases that are responsible for constructing, specific sialylated structures such as, TF antigen, Sialyl lewis antigen, sialyl α2-6 lactosaminyl structure, polysialic acid ,etc [3] Evidence shows that altered Salyltransferase (ST) expression has significant correlation with oncogenesis, tumor progression, and lymph node metastases . Detailed glycan analysis from human serum of cancer patients also show that there is increased expression of tri-sialylated tri-antennary glycan containing α 1-3 fucose, which forms part of sialyal Lewis x epitope [4].Since sialic acid is at the terminal position of the sugar chain in the molecule which is most exposure to the extracellular matrix. It is suggested that increased expression of sialic acid

causes charge repulsion among cells ,which facilitates cell motility and leads to cancer invasiveness.

Fucosylation is another important type of glycosylation. FucT3,4,5,6 and 9 are involved in the synthesis of Lewis antigens. Sialyl Lewis X or sialyl Lewis A is used as a tumor marker of certain types of cancer. Enhanced expression of Lewis antigens is a poor prognostics in colon cancer. Fut8, in particular, is a gene that encodes fucosyl transferase that puts on fucose on reducing terminal GlcNAc in α1-6 linkage. It is expressed in the liver, where abundant serum glycoproteins are made. Increased experession of Fut8 has been observed in malignant conditions. Thus, it is not surprising to notice that glycoproteins with increased core fucosylation such as GP73, AFP-L3, haptoglobin are found to be important biomarkers in various cancers. [5]

A perfect example of glycosylation association with cancer would be the GnT-V expression. GnT-V , or Mgat5 , is a gene that encodes β(1-6) GlcNAc transferase. Malignant cells often show increased expression of complex β(1-6) branched N-linked glycans, caused by increased expression of GnT-V. Experiments have shown that knocking down GnT-V expression by siRNA in mammalian cancer cells caused attenuation of cancer cell invasiveness .[6] Increased expression of GnT-V are found in various cancer types, including ovarian cancer and breast cancer. [7, 8]Dennis group made a knock out mouse model of GnT-V and induced mammary tumor growth by polyomavirus middle T oncogene. They found that tumor growth in mgat 5-/- mice was considerably less than control. This is a direct evidence that GnT-V promotes metastasis. [9]



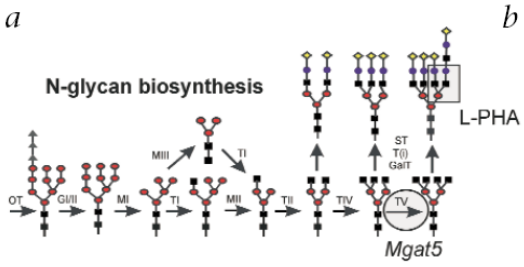


Figure 2 GNT-V knock out mouse model showing GNT-V associated with tumor genesis. Figure a is a count on tumor metastasis into the lung, more lung tumors are observed in wild type versus the GNT-V knock out mouse, showing that there is a direct correlation between GNT-V expression and tumor genesis (***Nature Medicine*** 2000, 6, 7.)

Profiling Glycans-Glycan analysis and Associated Techniques

Lectins and antibodies can be used to probe glycans general characteristics, but they can not reveal the intricacies of glycan structure; some chemical technologies for probing glycans are highlighted in the literature [10]

Common glycan analysis protocol use procedures like releasing, derivertization, mass spectrometry analysis, data interpretation are found in the literature [11]. The structural information obtained from tandem mass spectrometry is greatest when glycans are released from peptide or protein backbone. Robust release method is needed for N-linked and O-linked glycans. PNGase F works for releasing N-glycans. Since there is no available enzyme for complete release of O-glycans, people use chemical methods: β-elimination and ammonium based elimination. Both β-elimination and ammonium based elimination are base-catalyzed, β-elimination works with a strong base NaOH and $NaBH_4$. Ammonium based elimination does not need reducing agent and the intact aldehyde is ready for reductive amination. However, ammonium based elimination is a less effective releasing method, and results in about 10 fold less signals, but β-elimination can result in peeling reactions. Hydrazinolysis is another method to cleave both O-linked and N-linked, but it needs specialized instrument to handle hydrazine and cryo-drying of the sample[12]. It works well for O-linked release, but N-linked glycan requires harsher conditions, so it requires higher temperature which can lead to side reactions. Furthermore, there is extra chemical work up to regenerate intact reducing terminal.

Table 1 Glycan Releasing Methods Comparsion

	Principle	Release Type	Pros	Cons
PNGase F/A Endoglycosidases F1, F2 and F3	Enzymatic works to cleave sugar moiety from Protein and peptide	N-glycan	Efficient, easy to handle	Expensive, Enzymes are structural specific which leads to incomplete release
β-elimination	Strong base-catalyzed elimination	O-glycan	Complete Release	Results in Peeling
Ammonia-based Elimination	Weak Base-catalyzed elimination	O-glycan	Reducing terminal	Not complete release, 10 fold less recovery than β-elimination
Hydrazinolysis	Workwith hydrazine	N-glycan and O-glycan	Broad range of release	Not complete release Destruct protein component, need re-acetylation.Additional equipment to handle hydrazine. N-linked release requires higher temperature, results in side reactions

Glycans are derivertized to accommodate different detection methods. Commonly used fluorescent detection derivertization is reductive aminition. This introduces a fluorescent group such as 2-aminobenzamide (2-AB), 2-AP (2-aminopyridine), 2-aminobenzoic acid (2-AA) etc. Identification of labeled glycans is performed by comparing HPLC elution time with standard glucose units.

Another popular derivertization is permethylation, which turns free hydroxls on the sugar chain to -OMe. Permethylated glycans more easily ionized than underivertized by mass spectrometry .

New approach of glycan analysis, which does not dependent on isolation and analysis of proteins ,or the use of antibodies, holds particular promise in biomarker discovery. Our research center derived a solid glycan labeling technique derived from permethylation, permethylation

with $C^{13}H_3I$. Borrowing the idea from proteomics labeling, to reduce ionization differences between samples, C^{12} and C^{13} labeling are used for samples, and proves to be a robust isotope labeling method , [13, 14]. However, since this labeling occurs at each methlylation site, it gives a wide range of masses and analytes with similar m/z value usually overlap. Another more accurate labeling method for glycan derivertization is quantitative isobaric labeling, or QUIBL [15]. This method permethylates different aliquots of glycans with $^{13}CH_3I$ and CH_2DI. The two forms have the same nominal mass but differ in 0.002922 Da at each methylation site. Since glycans have many permethylation sites, the mass difference increases to an extent that can be nicely distinguished by high-resolution mass spectrometer. Shown in the figure 3 is differential labeling of a complex glycan resolved by FT-ICR. This labeling technique has been tested on Embryonic stem cells and embryonic body cell glycan analysis. This is a very robust glycan labeling method, which equals out the ionization efficiency difference, however, due to the design of the experiment, it calls for good resolution instrument.

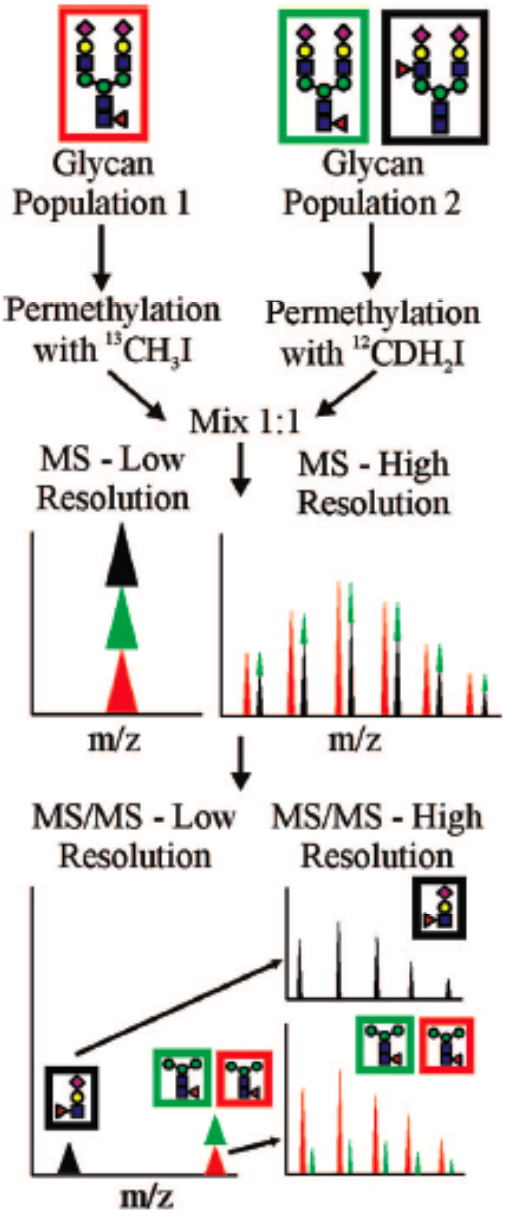


Figure 3

Flowchart for quantitative glycan analysis using isobaric labeling. Glycans from two biological samples are permethylated in either $^{13}CH_3I$ or $^{12}CH_2DI$ and mixed together prior to analysis. Peaks overlap in a low-resolution instrument, but in high resolution instrument the isotopic peaks can be distinguished

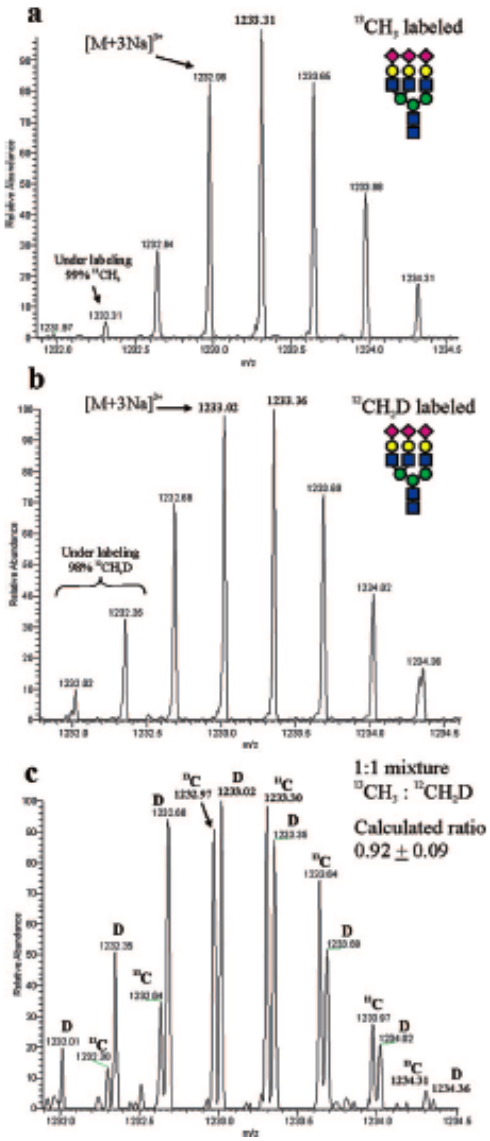


Figure 4

FT-ICR mass spectrum of $^{13}CH_3I$ (a) and CH_2DI (b), c is overlay of a and b , easily distinguishable by high resolution mass spectrometer, in a lesser instrument, they would overlap (*Journal of proteome research* 2007, 7, 7)

Unlike proteins and DNA, glycosylation is a non-template driven modification. Currently there is no complete database for carbohydrate sequence and linkage positions. Available search-by-molecular-weight search engine Glycomode can give possible composition combinations.

Glycobenchwork is another useful software which can show proposed fragmentation for glycan structures. To this point, data interpretation relies on skilled personnel piecing together full MS and MS/MS spectrum. The Reinhold group is at the forefront of developing carbohydrate database. They published a series of papers to illustrate sugar fragment mechanisms and proposed an algorithm for a database for carbohydrate sequencing [16-18].

Our lab, in collabration with Lance Wells lab, has improved many protocols by analyzing the total N-linked and O-linked glycans from Drosophila embryos and developed Total ion mapping (TIM) for identification and quantification[14, 19]. TIM works with XCalibur software package. Automated MS and MS/MS spectra were obtained in collection window 2.8 mass units in width with 5 scans, each 150ms in duration, were averaged for each collection window. The detection range is 500-2000 in 2.8 successive mass units with window-to window overlap of 0.8 mass units. TIM is extremely good for picking up minor components. We have used TIM to successfully characterize total glycans from pancreatic juice and stem cell samples.

Recent example in glycan analysis has been done on specific glycoproteins. For instance, The Mudd group focused on a specific glycoprotein (MUC1) from serum from advanced breast cancer patient.[20] Aberrant O-linked glycosylation has been implicated in disease progression. They used hydrazinolysis to release O-glycans, 2-AB labeling, HPLC profile against glucose unit, linkage position were determined by exoglycosidase. Core 1 structure dominated the profile. There are high levels of sialylation. Core 2 structure contributed 17% and TF antigen account for 14% of the total glycan. This is the first data of total glycan for MUC1. The data is somehow incomplete since there is only one advanced breast cancer sample, and there is no comparison

made between controls. As it is postulated that aberrant MUC1 glycosylation can be good biomarker for cancer.

Labrilla's group has done extensive glycan analysis on serum to find ovarian cancer biomarker. His idea is to focus on the sugar component and ignore the proteins it linked to. His group pioneered in using Infrared-multiphoton dissociation (IRMPD) as ionization method to fragment oligosaccharide and Fourier transform ion cyclotron resonance mass spectrometer for detection. IRMPD provides essentially same fragmentation ions as CID which can provide larger energy that not only isolate ions but also fragment down to the last residue. [21] They looked at O-linked glycans from cancer cells and ovarian cancer serum, glycoproteins were enriched by immunoblot and glycan were released , for each band of glycoprotein, they did glycan analysis as well as peptide fingerprinting to identify the peptide. They found out that there is a pattern difference between several glycoproteins between control and cancer samples and concluded that not only a single protein but several aberrantly glycosylated protein are prominent in ovarian cancer. [22]

A site specific analysis of N-glycans on haptoglobin in sera of pancreatic cancer, chronic pancreatitis, and control was done. Glycosylation pattern changed as the disease severes, There are more core fucosylated di-,tri-,and tetra-branched glycans and more Lewis X-type fucose increased at the Asn211 site. The conclusion was that site specific analysis of N-glycan of haptoglobin may be a useful as novel biomarker for pancreatic cancer. [23]

Haptoglobin is an acute-phase protein that has been associated with various cancer biomarker discovery. Earlier proteomic research (2-D gel) have found increased haptoglobin-I precursors in ovarian cancer patients, verified by Mass spectrometry identification, and antibody tests. [24]

Haptoglobin α-subunit has been characterized as potential serum biomarker in ovarian cancer.[25]Since heptoglobin is among the most abundant glycoproteins secreted by the liver, it is reasonable to hypothesize that the synthesis of the protein will increase as a general inflammatory response. So it is possible to assume such markers are not significant enough for a distinct type of cancer or for early stage.

Probing glycosylation changes with serum

It is common practice to do the investigation where cancer arises. So we would expect to find changes in glycosyltaion in the fluids or tissue near the inflamed or carcinoma area. However in a clinical setting, this tissue and fluid are hard to obtain. Just like it is too late to detect pancreatic cancer until patient's pancreas is cut open. An alternative way would be try to find diluted biomarkers that represents such malignant changes in clinical setting. Since serum comes from all parts of the body, and blood is most easily taken, it is the ideal non-invasive candidate fluid for biomarkers research. The glycoproteins secreted into the bloodstream compromise a major part of the serum proteome. Many clinical biomarkers and therapeutic targets are glycoproteins. Current antibody-based immunochemical test for cancer biomarkers of ovarian (CA125), breast(CA27.29 or CA15-3), pancreatic ,gastric, colonic (CA19-9) target highly glycosylated mucin protein. But existing biomarkers fall short with specificity and sensitivity, and has fairly high false positives. Besides biomarkers vary with different cancer staging, gender, race, location of the tumor, etc. Often times it is hard to tell between cancer and inflammatory disease since proteins differentially expressed in inflammatory disease are also involved in cancer.[26]

Human Serum is a highly complex biological fluid. In 2002, The Human Proteome Organization (HUPO) lunched a project to identify proteins in human serum. 51 institutions

participated and did the analysis with their home-based method. They identified 3020 proteins with more than 2 peptide matches.[27] The protein concentrations in serum ranges at least 10 orders of magnitude and the most abundant 20 proteins represents over 99 % of the total protein.[28] The abundant proteins often mask the presence of rare ones, making it extremely hard to identify the latter. I have taken graph from a recent news article in Science magazine which clearly shows the complexity of human serum proteome. The proteome are divided to three categories, classic plasma proteins , tissue leakage products, and interleukins ,cytokines. Only the last category are glycoproteins that can be possible biomarkers. That is why people from HUPO regard serum biomarkers as " Needle in a haystack ."

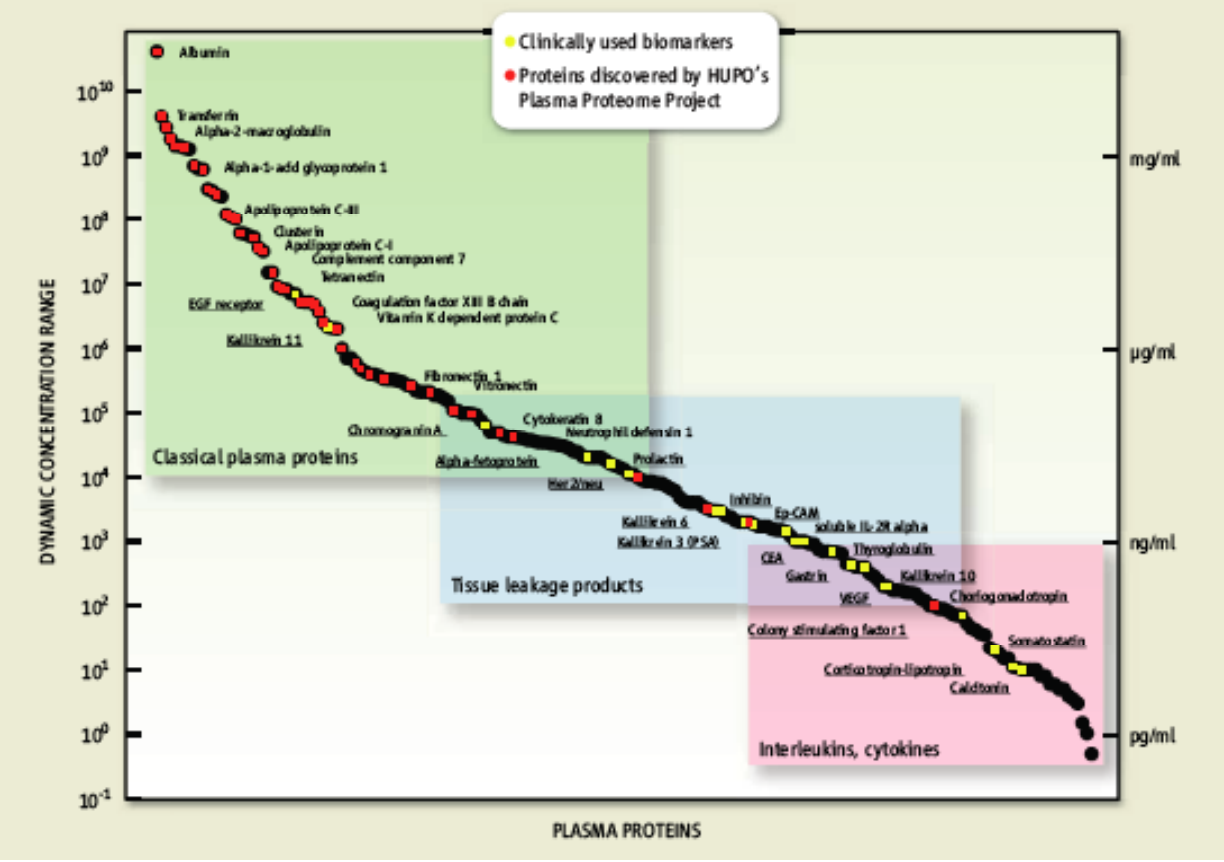


Figure 5 Needles in a haystack.

The abundance of different proteins in blood varies by more than 10 orders of magnitude.

Most commercially used biomarkers (yellow dots) are present in only in minute quantities in blood, below the level at which most proteins are detected (red dots). (Science, SEPTEMBER 2008, 321, Page 1760)

As many potential biomarkers are present in only minute quantities, a variety of methods are used to "amplify" the signals. Selected methods are removal of abundant proteins, glycoprotein enrichment, serum fractionation, etc. [29] There is commercially available serum protein removal kit contains immunodepletion columns, which can remove 6-12 most abundant serum proteins , such as albumin, α-transferin, This procedure removes hard rocks on our journey to find gold.

Serum fractionation is good way to simplify contents. Traditional proteomics methods of multidimensional chromatographic separation and SDS-PAGE gels gives in depth knowledge of the minor components, and sensitivity is greatly enhanced by implementing mass spectrometry as detection methods. [30]But as for easy and high-throughtput protein identification, fractionation into minute quantities is not the way to go.

Capturing Glycosylation with Advanced Techniques.

For reasons discussed earlier, it is in the best interest to focus biomarker candidates on specific glycoproteins types. Identification of glycoprotein relies on solid glycol-capture techniques. Glycoprotein enrichment is probably the best "amplification" method used. It takes advantage of the selectivity of immobilized lectins. Lectins are carbohydrate binding proteins isolated from plants and animals, they have sugar recognition domains which can specifically bind to a sugar motif. For instance, Concanavalin (Con A) selectively binds to high mannose type, Wheat Germ Agglutin (WGA) to complex type, LLA to core fucosylation, and *Phaseolus vulgaris* leukocytic phytohemagglutinin (LPHA), a lectin specific for ß1,6-branched

oligosaccharides, was used to assay GnT-V activity. Compared with antibodies, lectins have low affinity and broader specificity. They are widely used in glycomics as enrichment and detection methods. [31] Detailed procedures of glycoprotein enrichment through lectin affinity chromatography can be found in the literature.[32]

Hancock and co-workers developed a multilectin affinity column which combines ConA, WGA and Jacalin to capture the majority of glycoproteins present in human serum. [33] This unique combination of lectins not only allows you to selectively enrich glycoforms of interest but also it is very high-throughput. However, detailed procedure and chemical conditions of mixing are always desired to maximize recovery rate. Dr Lubman's group elaborated this method by applying two lectin columns to achieve maximum enrichment at two levels, glycoprotein and glycopeptide. The first column uses ConA, a high mannose binding column for global glycoprotein selection from serum proteome. The second ConA column extracts glycopeptides from trypsin digests. [34] They can also use a specific column to monitor glycosylation of interest , such as Wheat Germ Agglutinin, (WGA) Elderberry lectin,(SNA), Maackia amurensis lectin, (MAL) to extract sialoglycopeptide and monitor sialic acid expression and linkage positions.[35]

The lectin-sugar-recognition is also used as detection method as well. The Dubman group did a proof-of principle study with Glycoprotein enrichment and muti-lectin detection on serum from pancreatic cancer patients. Glycoproteins were first enriched by ConA&WGA columns. Specific patterns of lectin-sugar interaction were visualized by fluorescent tagging. Using the bioinformatic pattern analysis, they were able to identify the different glycan experession in cancer and controls. [36]

There are also available established chemical methods to pull out glycoproteins. A convenient method developed recently is labeling the glycans with a chemical reporter. Bertozzi

group incorporates a metabolic labeled azide sugar residue in cell culture and pulls out the sugar component by affinity separation. Labeling can be done on different sugar residues, such as N-azidoacetylmannosamine (ManNAz), N-azidoacetylgalactosamine(GalNAz), N-azidoacetylglucosamine(GlcNAz), 6-azidofucose(6AzFuc)[37]. Highlights of these chemical methods for glycosylation discovery can be found in this review.[38]However, this method is restricted to use in a cell culture medium.

Mapping glycosylation Sites with Mass Spectrometry

Mapping glycosylation site is a complicated task. First, the glycoprotein needs to be selectively enriched. This can be achieved either by lectin binding or chemical immobilization. Additional information can be obtained by using either enzymatic or chemoselective reaction to incorporate isotope tags at specific glycosylation sites. The isotope labeling facilitates mass spectrometry based confirmation of glycoprotein identification and quantification of glycosites. Three best used glycoprotein enrichment and N-glycosite mapping methods are: solid phase extraction of N-linked glycopeptide (SPEG), isotope-coded tagging and mass spectrometry (IGOT) and BEMAD(β-elimination and micheal addition)

SPEG is also known as Glycopeptide Capture and Release method and was initiated by the Aebersold group and first tested on C-elegans.[39].This method oxidizes the carbohydrate in glycopeptide from cis-diols into aldehydes, these aldehydes were treated chemoselectively with hydrazide moieties, forming a hydrazide linkage that anchors glycopeptide to a solid support. After trypsin digestion, the non-binding proteins are washed away. Incorporating an isotope tag (succinc acid d0 and d4) helps mass spectrometric identification, quantification and comparative study between samples. At last, a Peptide N-Glycosidase, PNGase F releases the peptide by cleaving the sugar residues on the peptide backbone, turning the Asparagine to Aspartic acid,

which gives a mass increase by 1Da to the peptide chain. This is a global method for glycoprotein enrichment and glycosite mapping and has been applied to human serum sample for biomaker discovery. Detailed protocol can be found in the literature. [40] Zhou etal further optimized the capture and release conditions and implemented this method with glycosylation pattern analysis of mouse serum in benign and induced skin cancer conditions. [41] It is suggested improved enrichment efficiency could be achieved by taking chemical immobilization step at the glycopeptide level rather than glycoprotein level, since more than one glycosylation sites occur in a single glycoprotein. [42]This is a universal method for enrichment since it is high-throughtput and does not distinguish between glycans like lectins do, but it has the drawback of destroying and losing the sugar component.

Another useful glycopeptide enrichment and site mapping method is brought about by Japanese scientists.[43] This method use lectin affinity chromatography to selectively bind glycoproteins, during the release of glycans, they incorporated O_{18} isotope tag on the glycosylation site by running the enzyme release in O18 water. Taking into account the 1 Dalton change from Asn to Asp, it altogether gives a 3 Dalton mass shift at the glycosylation site. This large scale method was used to the characterization of N-linked high- mannose and hybrid-type glycoprotein from Caenorhabditis elegans[43] and Drosophila. [44]

In the area of O-glycosylation quantification and site mapping, Dr Wells group initiated BEMAD for β-elimination and Micheal Addition. Base catalyzed elimination reaction creates an α,β- unsaturated ketone, followed by Micheal addition of light or heavy labeled dithiothreitol (DTT). The glyco- or phosphoryl proteins are enriched by a thiol chromatography and further analyzed by mass spectrometry. This method can be used for quantitatively labeling and site

mapping of O-linked glycosylation and o-phosphate modifications by coupling with enzymatic dephosphorylation or O-glycan hydrolysis to block either one. [45]

Table 1.2 Glycopeptide Enrichment and Glycosylation Site Mapping

	Capture Method	Enrichment Type	Productive Use	Application and Comments
Solid phase extraction of N-linked glycopeptide (SPEG)	Chemical reaction	N-linked	Caenorhabditis elegans, Mouse Plasma	Global site mapping method, destroys the glycan component
Isotope coded glycosylation site-specific tagging (IGOT)	Lectin Enrichment	N-linked	Caenorhabditis elegans Mouse liver	Lectin enrichment is specific, may loss certain type of glycans
β-Elimination/Michael addition with DTT (BEMAD)	Thiol column	O-phosphate or O-link modified peptides	Mouse brain	May map O-phosphate and O-glycan at the same time, need prior enzymatic release to block one or the other

All three methods involve the loss of glycan component by enzymatic release or chemical cleavage, they are robust methods to map sites but a challenging question to answer at this point would be which glycan links to which glycopeptide and at which site. To address this challenge researchers took advantage of different fragmentaion methods that results in bond breakage difference. There are two kinds of fragmentation: traditionally used Collision induced dissociation (CID) and the newly developed Electron transfer dissociation (ETD) or electron capture dissociation (ECD). Collision activation has been widely used in tandem mass spectrometry as a core tool for proteomics. It fragments peptides into b and y ions series. In the case of glycopeptide, since the activation energy for glycosic bond is lower than peptide backbone, it breaks the glycosic bonds first at lower activation energies. Electron transfer dissociation is a new method to fragment peptides and it utilizes ion/ion chemistry. ETD fragments peptides by transferring an electron from a radical anion to a protonated peptide. This fragments the peptide backbone, causing the cleavage at Cα-N bond, just like CID. This creates a serie of complimentary c and z type ions instead of b and y ions.[46]

CID fragmentation typically generates spectra showing limited peptide backbone fragmentation. However, when these peptides were fragmented by ETD, peptide backbone fragments and gives a complimentary series of ions and thus extensive peptide sequence information. In addition, labile PTMs remain intact. So we can easily map glycosylation site by comparing the z ions to find loss of the entire N-glycosylated asparagines side chain. [47]

A good experiment design with rapid alternating Electron Transfer dissociation (ETD) and Collision-induced dissociation (CID) in a single instrument provides detailed information on glycan structure, peptide sequence and precise glycosylation site. [48] This is done on both N-linked and O-linked synthetic glycopeptides models.[47, 49] The first biological sample to obtain N-

glycosylation profile using this method was done on immunoglobulin (IgG) subclasses from human serum[50].

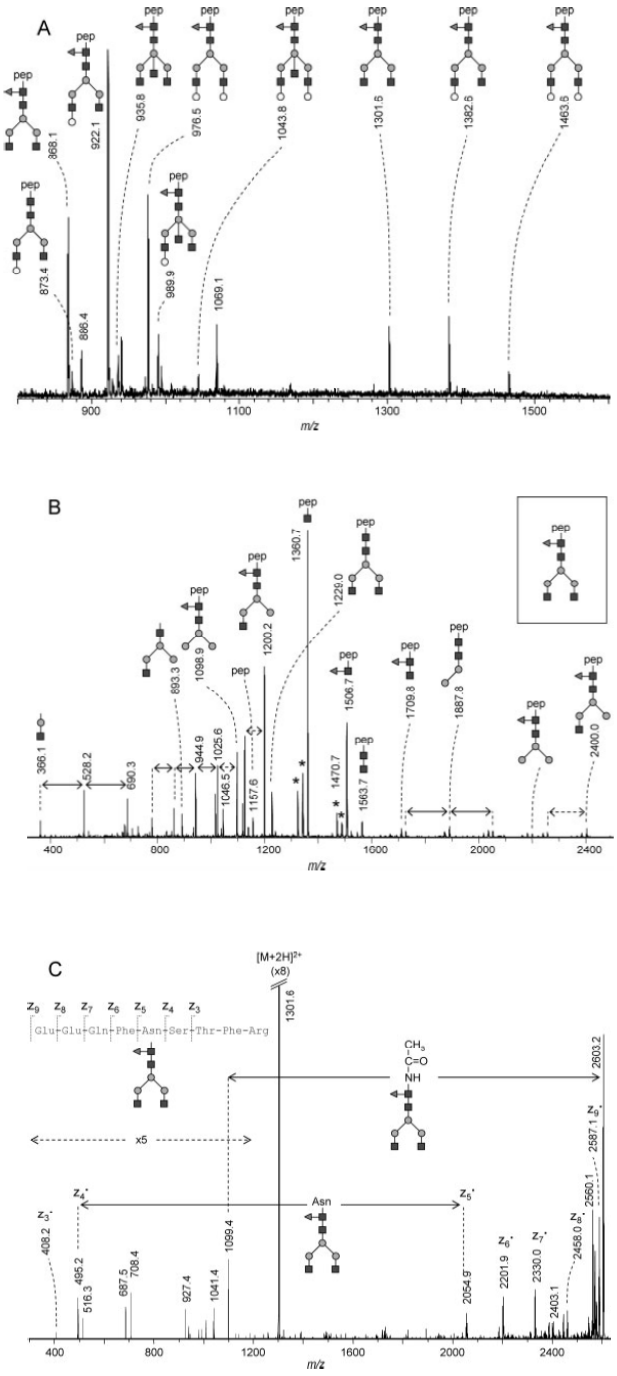


Figure 6 Analysis of IgG3 glycopeptides with neutral glycan chains Figure A is MS profile of

IgG3 glycoproteins with neutral glycan chains. They select ion m/z 1301.6 to do CID to get the glycan sequence in Figure B, in figure C they monitored ETD on 1301.6 ion to obtain peptide sequence, and observed loss of entire asparigine linked sugar chain , shown by Z_5 to Z_4 , this has been observed on synthetic standard peptides.[47] (Yehia Mechref, M. M., Milos V. Novotony, , *Methods in Molecular Biology* 424, 23)

A play with ETD-CID ionization is by far the most direct site mapping method, compared with MS/MS glycan analysis method, it robust and simple but it lacks sugar fragmentation information leading to deducing sugar isomeric constructs.

Monitoring Glycosylation with Ion Mobility Mass Spectrometry

A new area in mass spectrometry that has been used for glycosylation associated analysis is ion-mobility mass spectrometry (IMS-MS). The mass spectrometer has a drift tube, in which the charged ions forced to move under a low electric field, creating ion mobility separation. In this way, ions undergo two dimensions of separation: ion mobility and m/z value. This facilitates the detection of product of interest. [51] Ion mobility mass spectrometry gives you the advantage to look at isobars, same mass but structurally distinct moieties.

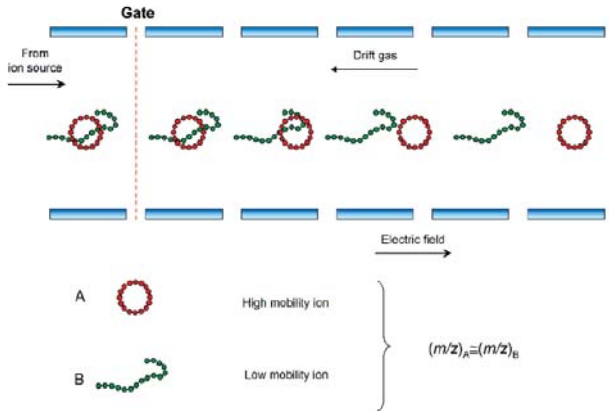


Figure 7 Schematic presentation of the principle of ion mobility separation in electric field across the drift gas. Letters A and B are conventional notations for the ions with similar or

equal m/z values but different mobility. (Yehia Mechref, M. M., Milos V. Novotony, , *Methods in Molecular Biology* 424, 23)

IMS-MS has been used to study human urine glycans. It is been proven that Ion Mobility separation greatly reduced ion overlapping of different charge states. [52] Shown in graph a is the overlapped doubly and triply charged ions with similar m/z value. b, c are the single component from extracted Ion Chromotogram. Using ion mobility separation, two ions are easily separated as shown in d: plot of 2-dimensional separation m/z over drift time. e : Selected areas indicate extracted ion current chromatogram A for the precursor ions at m/z 1008.827 and chromatogram B for the precursor ions at m/z 1007.354.

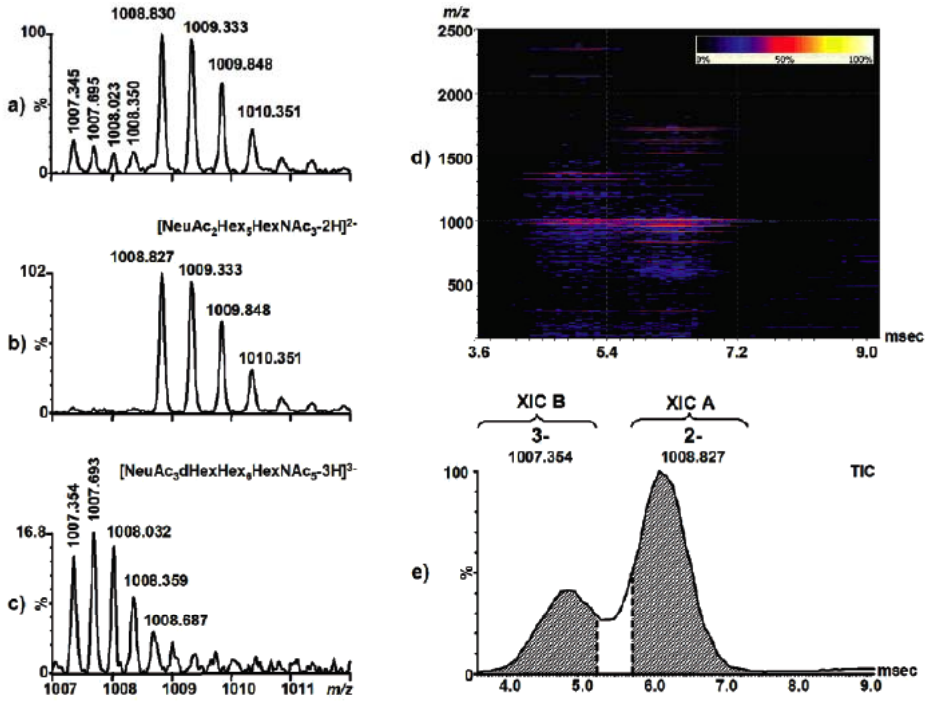


Figure 8 Ion mobility mass spectrometry has resolving to separate glycans of different charge state but with similar m/z value. Shown in figure b and c are two forms of glycans with similar m/z values. B is doubly charged , c is triply charged, figure a is the overlay of the two forms. By m/z value alone they can not be separated, but taking into account of ion mobility

differences , figure d, they are perfectly separated, and e is the extracted ion chromatogram of the two ions separated by drifting time. (*Methods in Molecular Biology* 424, 23)

Ion mobility has resolving power of separating glycan structural isomers due to carbohydrate conformation differences, glycan moieties with bigger cross sections have more chance of hitting buffer gas molecules in the drift tube leading to longer drifting time. The separation is achieved in milliseconds. Mobility separation greatly reduces spectral congestion of glycan mixtures and facilitates mass spectrometry detection [53] Ion mobility mass spectrometry gives you advantage to look at isobars, but structurally distinct moieties. Novotny's group first tested IMS-MS on glycans from liver cancer patients, Cirrhosis patients and healthy controls; they were able to find isomer distribution differences that could not be measured easily by MS techniques, unless MS/MS is used. They concluded from statistics finding that isomeric information maybe indicative of disease states. [54]

Ion mobility mass spectrometry can be used in site mapping as well. The preliminary work is done on a known monoclonal antibody. The idea is the same with rapidly alternating ECD and CID discussed earlier, the only advantage is it adds another dimension of ion mobility separation. In short, site mapping is done in two steps. In the first step, identify glycopeptide as well as elution time by monitoring glycopeptide reporter (oxonimum ions) in order to trace glycopeptide precursor ions .The second step determines the glycan and peptide sequence, as well as site mapping information, this is done by a two-step sequential fragmentation. The instrument has a dual-collision cell design that allows collision energies to be applied just enough to fragment either peptide or glycan moiety and scan the product ion. This method avoids the lengthy enrichment procedure by doing precursor and product ion scans, and greatly eases detection by giving another dimension of separation. [55]

Probing human serum with SELDI

Surface Enhanced Laser Desorption Ionization (SELDI) is a modification of MALDI. It is a protein chip array-based chromatographic retention technology illustrated below by a company note [56]. Namely, the serum is selectively bound on a chip. The selectivity comes from antibody, lectin, etc. When the chip dries out, forming a crystal matrix, a laser is fired, generating the ion signals. The signals are then recorded and sorted by different statistic methods. Statistical patterns are correlated with disease and severity. Researchers have used SELDI for early biomarker discoveries in cancer. [57]

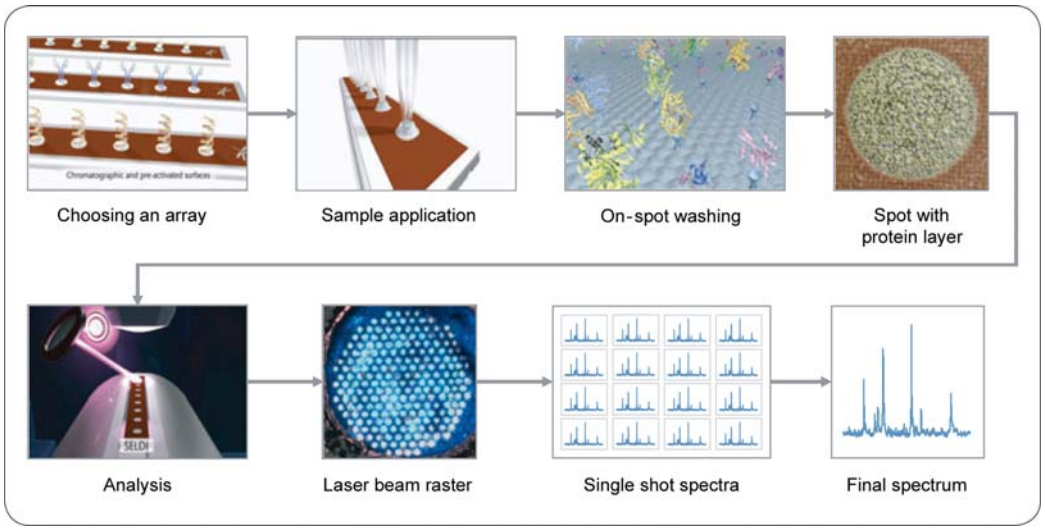


Figure 9 schimatic representation of SELDI sample preparation and data generation

SELDI- TOF instrument has relatively low resolution in which many different ions are packed within the same peak, so peaks cannot be properly identified. SELDI biomarker approach is the most high throughput method and also has been the most questioned method. What is bound on the chip and what is not? How much percentile of analyte is retained in terms of signal intensity? How about SELDI ionization efficiency? Significant loss of diagnostic information maybe lost in the wash. Recent publications have favored this method and published data were

almost 100% sensitivity but none of the peak was identified. Using different statistics methods may find different pattern result, in other words, there is huge room in the statistics aspect.[58]

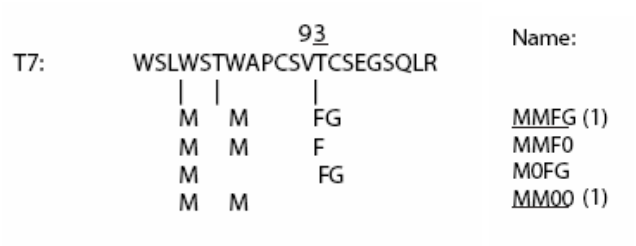
Multiple Reaction Monitoring as a Quantitative Verification Method

In analyzing biological samples of minute quantities by mass spectrometry, sensitivity and resolution are the key issues. A good way to address this problem is by improving selectivity. A new type of mass spectrometry technique, Multiple Reaction Monitoring is used. This is done on a triple –quadrupole instrument with three quadrupole mass spectrometers arranged in tandem. The three quads functions as: precursor ion scan, collision cell, product ion scan. This method records the formation of one or more specific product ions from a precursor with a selected m/z value, so called " transitions". This technique greatly increases sensitivity, because of high selectivity of precursor ions or precursor/product ion pairs. Tao He's group used immunoaffinity-MRM to select and quantify carcinoembryonic antigen (CEA), a low abundance tumor biomarker in lung cancer. Isotopic labeled peptides were spiked in as internal standard which renders MRM a quantitative method. A nice linear response from 15ng/ml to 250ng/ml is observed[59]. Researchers have also developed quantitative assays for low abundance plasma proteins. 6 plasma proteins were detected with linear response over concentration down to 1-10ng/ml range. [60]This method is high throughput, which allows monitoring several proteins of interest at one time. However, the only limitation is, the fragmentation pattern of transition must be known, in other words, you have to know what you are looking for.

In this sense, MRM is best used as a quantitative validation method for cancer biomarker discovery. Paulovich's group tested on mouse with induced breast cancer. In the experiment setting, they used proteomic based mass spectrometry for biomarker discovery, and confirmed several potential biomarkers, Osteopontin and Eilbulin-2, with quantitative MRM-mass

spectrometry. As they noted ,MRM overcomes the drawback of paucity of antibody and achieves sensitive detection in a quantitative manner.[61]

Jan Hofsteenge's group uses MRM to study Peters Plus Syndrome, a subtype of Congenital Disorder of Glycosylation (CDG), where patients suffer from defective O-glycosylation. The research is done on a family, where parents are normal while their sons are Peter Plus patients. They used Imunoaffinity-mass spectrometry to monitored a reporter protein properdin with possible defected O-fucosylated glycan and did comparison between samples. They observed several signiture peptides where they see transitions of aberrant glycosylation. Shown below is a signiture peptide T7 from properdin with different glycosylation expression. They were able to monitor transition from MMFG to MMF0 and M0FG in healthy parents in Figure A and B, common fragmentation of losing mannose or glucose. However, such transitions was not observed from their diseased son in Figure C and D, but there are 18 times the amount of MMF0 and its fragmentation product M0F0 not shown in the figure. This shows that patients lack glucose modification on Fucose, which confirms previous findings of genetic mutation of β1,3-glucosyltransferase in Peters Plus Syndrome.[62]



Naming of different glycosylation forms

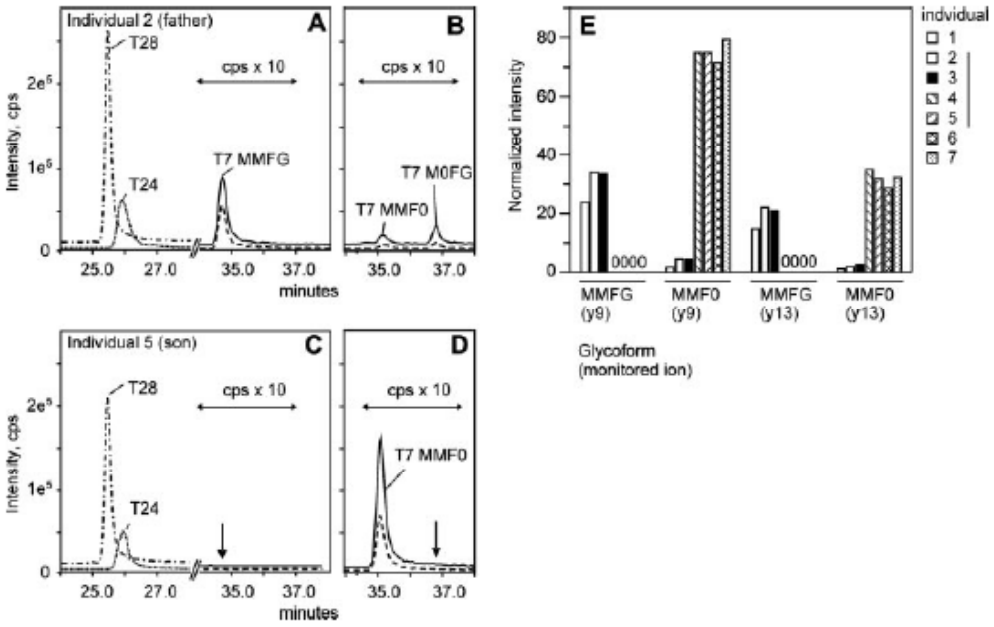


Figure 10 monitoring transitions of peptide T7 from properdin

(Taken from *Journal of Biological Chemistry* 2008, 283, 7)

Using similar technique, Thierry Hennet's group studied N-glycosylation site occupancy with CDG type-I patients. They monitored site occupancy of glycopeptides from two serum glycoproteins, transferrin and α1-antitrypsin. They found out in some glycoproteins site occupancy is stringently correlated with disease severity and proposed that glycosylation site occupancy study by MRM can be a good biomarker for disease diagnosis and prognosis.[63]

Conclusions and Future Directions:

Finding cancer biomarkers in human serum is such a daunting challenge. In this review, I talk about possible techniques that are useful in the work. Due to the nature of human serum, we need better separation and enrichment techniques be efficient. Problems faced are :

1. Scarcity of lectin and antibody to study analyte of interest
2. Using different peptide search algorithm results in different result

3. Ion mobility MS is still at early stage. Mass spectrometry sensitivity and resolution not enough, MRM is sensitive but can only be used as validation method

4. Glycan analysis technique, specific carbohydrate database is not developed, rely on skilled personal interpretation

5. Not knowing what possible markers can be.

6. Scarcity of sample which lacks statistics meaning

As we look into the future, when the markers are used in clinic trail, we also need to take into consideration tumor type, stage and grade, etc. we have to consider expense and expertise of personnel. Is setting up mass spectrometer possible in every clinic? Once we have identified a biomarker, the ultimate goal would be develop an antibody detection kit . It would be good idear to have a kit for every type of cancer since cancer biomarkers are non-specific. It maybe a combination of several proteins, lipids or carbohydrates that directly recognize significant biomarkers of the cancer type. Or it would be more fantastic to develop personal cancer detection kits or disease progression monitoring kits.

If we find a marker, it would be better idea to generate antibody to recognize such marker. For instance, H.Umeyama [64] described an antiviral protein, actinohivin, a lectin that contains 3 mannose binding sites, and it is active against human immunodeficiency virus(HIV) and simian immunodeficiency virus(SIV). The HIV is covered with gp120/gp41 glycoprotein complex, whose surface contains numorous N-glycan, most of which are high mannose. It is postulated that Actinohivin fix gp120 conformation in such a way as to block its interaction with host cell CD4. In the same sense, it is a good idea to find a lectin antibody that recognize not only glycan or peptide, but both glycan and peptide, so as to achieve higher detection accuracy.

Reference:

1. Apweiler R, H. H., Sharon N, ***Biochim Biophys Acta*** 1999, 1473, 5.

2. Mark M. Fuster, J. D. E., *Nature cancer review* 2005.

3. Chircolo, F. D. O. a. M. *Glycoconjugate Journal* 2001, (18), 10.

4. Umi M. Abd Hamid Louis Royle, R. S., Catherine M. Radcliffe, David J. Harvey, Sarah J. Storr, Maria Pardo, Robin Antrobus, Caroline J. Chapman, Nicole Zitzmann, John F. Robertson, Raymond A. Dwek, and Pauline M. Rudd, *Glycobiology* 2008.

5. Yunxue Zhao, J. L., Yanli Xing, Jingjian Wang, Congxiao Lu, Xianliang Xin and Meiyu Geng, *Molecular and Cellular biochemistry* 2008, 309, 10.

6. Guo HB Randolph M, P. M Journal of Biological Chemistry 2007, 282, (30), 13.

7. Dosaka-Akita H, M. E., Suzuki O, ItohT, Katoh H,Taniguchi N, *Clinical Cancer Research* 2004, 10, 9.

8. Eiji Miyoshi, K. M., and Tsutomu Nakagawa, *Journal of Biochemistry* 2008.

9. Maria Granovsky, J. F., Judy Pawling, William J. Muller, Rama Khokha&James W. Dennis, ***Nature Medicine*** 2000, 6, 7.

10. Bertozzi, J. A. P. a. C. B., *Cell* 2006, 126, 4.

11. Willy Morelle, J.-C. M.. *Nature Protocols* 2007, (2), 8.

12. Ashford.D, D., RA, Welply, J.K, Amatayukul,S, Homans, S.W, Lis, H,Taylor,G.N., Sharon, N.,ad Rademacher, T.W, *Eur.J .Biochem.* 1987, 166, 20.

13. Gerardo Alvarez-Manilla, N. L. W., Trina Abney, James Atwood, III, Parastoo Azadi, Will S. York, Michael Pierce and Ron Orlando, *Glycobiology* 2007.

14. Kazuhiro Aoki, M. P., Jae-Min Lim, Rebecca Cantu, Lance Wells, and Michael Tiemeyer, *Journal of Biological Chemistry* 2007, 282, (12), 16.

15. Atwood JA 3rd, C. L., Alvarez-Manilla G, Warren NL, York WS, Orlando R.. *Journal of proteome research* 2007, 7, 7.

16. David Ashline, S. S., Andy Hanneman, and Vernon Reinhold, *Anal.Chem*. 2005, 77, 13.

17. Anthony J. Lapadula, P. J. H., Andy J. Hanneman, David J. Ashline,Hailong Zhang, and Vernon N. Reinhold, *Anal. Chem.* 2005, 77, 9.

18. Hailong Zhang, S. S. a. V. N. R. Anal.Chem. 2005, 77, 8.

19. Kazuhiro Aoki, M. P., Samuel S. Lee, Brain Dong, Khoi Nguyen, Katherine H. McGlamry and Micheal Tiemeyer, *Journal of Biological Chemistry* 2008, 283, (44), 16.

20. Sarah J Storr, L. R., Caroline J Chapman, Umi M Abd Hamid, John F Robertson, Andrea Murray, Raymond A Dwek, and Pauline M Rudd. *Glycobiology* 2008, 18, (6), 7.

21. Lebrilla, Y. P. a. C. B., *Mass Spectrometry Reviews* 2005, 24, 32.

22. Bensheng Li, H. J. A., Crystal Kirmiz, Carlito B. Lebrialla, Kit S. Lam, and Suzanne Miyamoto, *Journal of Proteome Research* 2008, (7), 13.

23. Miyako Nakano, T. N., Toshifumi Ito, Takatoshi Kitada, Taizo Hijioka, Akinori Kasahara, Michiko Tajiri, Yoshinao Wada, Naoyuki Taniguchi and Eiji Miyoshi. *Int. J. Cancer* 2008, 122, 9.

24. N Ahmed, G. B., etal, *British Journal of Cancer* 2004, 91, 12.

25. Bin Ye, D. W. C., Steven J. Skates, Steven P. Gygi, Vanessa Pratomo, Lanfei Fu, Nora K. Horick, Larry J. Licklider, John O. Schorge, Ross S. Berkowitz and Samuel C. Mok, Haptoglobin-. *Clinical Cancer Research* 2003, 9, 8.

26. Ru Chen, T. A. B., Sheng Pan, Kelly Cooke, Kara White Moyes, Zhaoli Lane David A. Crispin, David R. Goodlett, Ruedi Aebersold and Mary P. Bronner *Molecular & Cellular Proteomics* 2007, 6, (8), 9.

27. *Proteomics* 2005, (5), 3.

28. Tam SW, P. J., Hinerfeld D, *Expert Rev Proteomics* 2004.

29. Hyong-Joo Lee, E.-Y. L., Min-Seok Kwon and Young-Ki Paik, *Current Opinion in Chemical Biology* 2006, (10), 8.

30. Vitor Faca, S. J. P., Lisa Newcomb, Veronika Glukhova, Doug Phanstiel, Alexei Krasnoselsky, Qing Zhang, Jason Struthers, Hong Wang, Jimmy Eng, Matt Fitzgibbon, Martin McIntosh and Samir Hanash *Journal of Proteome Research* 2007, (6), 8.

31. Hirabayashi, J., Concept, Strategy and Realization of Lectin-based glycan profiling. *J. Biochem* 2008, 144, 9.

32. Yehia Mechref, M. M., Milos V. Novotony, , *Methods in Molecular Biology* 424, 23.

33. Yang, Z. H., W. S., *Journal of Chromatography A* 2004, (1053), 10.

34. Zhao, J., *Journal of Proteome Research* 2007, (6), 11.

35. Zhao J, S. D., Heidt D, Anderson MA, Lubman DM, *Journal of Proteome Research* 2006, (5), 11.

36. Zhao J, P. T., Qiu W, Shedden K, Hinderer R, Misek DE, Anderson MA, Simeone DM, Lubman DM., *Journal of Proteome Research* 2007, (5), 11.

37. Bertozzi, S. T. L. a. C. R., *Nature Protocols* 2007, (2), 15.

38. Kohler, M. R. B. a. J. J., *Current Opinion in Chemical Biology* 2007, (11), 7.

39. Hui Zhang, X.-j. L., Daniel B Martin&Ruedi Aebersold *Nat. Biotechnology* 2003, (21), 7.

40. Yuan Tian, Y. Z., Sarah Elliott1, Ruedi Aebersold3 & Hui Zhang *Nat. Protocol* 2007, 2, 4.

41. Yong Zhou, R. A., and Hui Zhang, *Anal. Chem* 2007, 79, 12.

42. Bingyun Sun, J. A. R., Angelita G. Utleg, James T. White, Xiaowei Yan, Biaoyang Lin and Leroy Hood, *Molecular & Cellular Proteomics* 2007, (6), 10.

43. Hiroyuki Kaji, H. S. e., *Nature Biotechnology* 2003, 21, (6), 5.

44. Kate Koles, J.-M. L., Kazuhiro Aoki, Mindy Porterfield4, Michael Tiemeyer, Lance Wells, and Vlad Panin, *Glycobiology* 2007, 17, (12), 16.

45. Keith Vosserller, K. C. H., Robert J. Chalkley, Jonanthan C. Trinidad, Lance Wells, Gerald W. Hart and Alma L. Burlingame, *Proteomics* 2005, (5), 11.

46. John E. P. Syka, J. J. C., Melanie J. Schroeder§, Jeffrey Shabanowitz§, and Donald F. Hunt, *PNAS* ,2004, 101, (26), 6.

47. M. Isabel Catalina, C. A. M. K., Andre M. Deelder and Manfred Wuhrer , *Rapid Comm. Mass Spectrometry* 2007, 21, 9.

48. Hongling Han, Y. X., Min Yang, and Scott A McLuckey, *Anal Chem* 2008, 80, 6.

49. Deguchi, K., *Rapid Communication in Mass Spectrometry* 2007, (21), 8.

50. Manfred Wuhrer, J. C. S., Fleur E. van de Geijn, Carolien A. M. Koeleman, C. Theo Verrips, Radboud J.E. Dolhain, Cornelis H. Hokke and Andr M. Deelder, *Proteomics* 2007, 7, 12.

51. Abu B. Kanu, P. D., Maggie Tam, Laura Matz and Herbert H. Hill Jr. , *Journal of Mass Spectrometry* 2008, 43, 22.

52. Sergey Y. Vakhrushev, J. L., Lain Campuzano, Chris Hughes and Jasna Peter-Katalinic, *Anal Chem* 2008, 80, 8.

53. Manolo D. Plasencia, D. I., Samuel I. Merebloom, Yehia Mechref, and David E. Clemmer *J Am. Soc. Mass Spectrum* 2008.

54. D. Isailovic, R. T. K., M.D. Plasencia, S.T. Stokes, Z. Kyselova, R.Goldman, Y.Mechref, M.V Novotny, and D.E.Clemmer, , *Journal of Proteome Research* 2008, (7), 9.

55. Petra Olivova, W. C., Asish B. Chakraborty and John C. Gebler, *Rapid communication in mass spectrometry* 2008, (22), 12.

56. Sonja Vorderwiilbecke, S. C., Scott R. Weinberger & Andreas Wiesner, *Nature Methods* 2005, 2, (5), 3.

57. Emanuel F Petricoin, L. A. L., *Current Opinion in biotechnology* 2004, (15), 7.

58. *Cancer Research* 2006, 66, (11).

59. Nicol GR, H. M., Kim J, Birse CE, Brand E, Nguyen A, Mesri M, FitzHugh W, Kaminker P, Moore PA, Ruben SM, He T, *Molecular & Cellular Proteomics* 2008, 7, (10), 8.

60. Hasmik Keshishian, T. A., Michael Burgess, Eric Kuhn, and Steven A. Carr, *Molecular & Cellular Proteomics* 2007, 17.

61. Jeffery R. Whiteaker, H. Z., Lei Zhao, Pei Wang, Karen S. Kelly-Spratt, Richard G Ivey, Brain D. Piening, Li-Chi Feng, Erik Kasarda, Kay E. Gurley, Jimmy K. Eng, Lewis A. Chodosh, Christopher J. Kemp, Martin W. McIntosh and Amada G. Paulovich, *Journal of Proteome Research* 2007, (6), 13.

62. Daniel Hess , J. J. K., Saskia A. Lesnik Oberstein, Raoul C. M. Hennekam, *Journal of Biological Chemistry* 2008, 283, 7.

63. Andreas J. Hulsmeier, P. P.-B. a. T. H., *Molecular & Cellular Proteomics* 2007, 6, 7.

64. Chiba H, I. J., Nakashima H, Omura S, Tanaka H., Actinohivin, *Biochem Biophys Res Commun* 2004, 316, (11), 8.