

## Review

---

# Cancer biomarker discovery: Lectin-based strategies targeting glycoproteins

David Clark and Li Mao\*

*Department of Oncology and Diagnostic Sciences, University of Maryland, School of Dentistry, Baltimore, MD, USA*

**Abstract.** Biomarker discovery can identify molecular markers in various cancers that can be used for detection, screening, diagnosis, and monitoring of disease progression. Lectin-affinity is a technique that can be used for the enrichment of glycoproteins from a complex sample, facilitating the discovery of novel cancer biomarkers associated with a disease state.

**Keywords:** Lectins, lectin-affinity, biomarkers, glycosylation, glycoproteins

## 1. Cancer biomarker discovery

A biomarker, as defined by the National Institute of Health, is “a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention”, thus summarizing the role they can play in identifying the occurrence or progression of a pathogenic disease [4]. With this in mind, discovery of biomarkers related to cancer would be beneficial in the scope of detection, diagnosis, screening, and monitoring disease progression. Development of cancer biomarkers is a challenging process, best described as occurring in five consecutive phases: (1) Preclinical exploratory studies, (2) clinical assay for development for clinical disease, (3) retrospective longitudinal repository studies, (4) prospective screening studies, and (5) cancer control studies [31].

Biomarkers related to cancer may also be used to direct drug therapy, serving to explain clinical results in response to chemotherapy treatment. Such biomarkers, termed predictive biomarkers, may be used not only to

direct which agent of combination of agents should be used in treating a disease state, but also identify the population of patients who may show the best response to treatment (as well as identify those who may show adverse effects) [14]. In clinical trials requiring long-term intervention strategies, which require long periods of observation and large patient groups, biomarkers can be substituted for clinical responses or clinical end points [4]. Furthermore, incorporating a pattern-based biomarker, constructed of several single biomarkers that when evaluated together may be used in detection, risk assessment, screening, and diagnosis, may be more informative in predicting responses to individual therapeutic regimens [21]. These biomarkers will be an important integrated component of personalized medicine strategy in cancer treatments.

Sources of biomarkers can include tumor tissue, as well as bodily fluids including blood (serum and plasma), urine, sputum, saliva and cerebrospinal fluid (CSF). Development of biomarkers that can be obtained via minimal or non-invasive methods are ideal, and blood fulfills this description [35]. Not only is blood easily accessible for sampling, but because of the network of arteries, veins, and capillaries that come in contact with organs, it can collect molecular indicators such as proteins secreted, shed, or otherwise released by tissues [36]. The challenge in biomarker discovery using blood is that the plasma proteome is

---

\*Corresponding author: Li Mao, MD, Department of Oncology and Diagnostic Sciences, University of Maryland School of Dentistry, 650 W. Baltimore St., Baltimore, MD 21201, USA. E-mail: LMao@umaryland.edu.

the most complex human proteome, consisting of proteins whose function is dependent on their presence in plasma and transient proteins which are secreted and shed into the bloodstream [13]. It is this latter group where candidate biomarkers would most likely be classified, as proteins of interest would be secreted, shed, or leaked from the affected tissue. Further complicating biomarker discovery is the dynamic range of protein concentrations found in plasma, varying up to twelve orders of magnitude, wherein highly abundant proteins (i.e. albumin and IgG) compose the majority of the plasma proteome, making low-abundant proteins that may serve as biomarkers difficult to identify [12]. One strategy to increase the identification of low abundant proteins is to deplete the serum sample of highly abundant proteins using immune-affinity, thus reducing the dynamic range. However, several drawbacks of this strategy include the loss of proteins of interest due to nonselective binding or depletion of albumin, which functions as a carrier protein to assist in the circulation and transport of other proteins [13]. With these challenges in mind, prior to a biomarker found in plasma being considered for clinical use in a cancer setting, several criteria must be met including: (1) release into the circulation in detectable amounts by a small, non-symptomatic tumor, (2) exhibit specificity for the tissue or organ of origin, and (3) exhibit specificity for the cancer state, meaning it has no association with a noncancer disease [7]. Of note, the target biomarker may not exist in the tumor proteome itself, but instead manifests in the tumor microenvironment as a response to the presence or progression of the tumor [27].

There have been many studies focusing on biomarkers discovery, based on various type of biomolecules such as DNA, mRNA, ncRNA, lipids, carbohydrates, and proteins. Alterations in DNA and mRNA expression ultimately result in aberrant expression or modification of protein products, indicative of a disease state, which make proteins attractive targets in biomarker discovery [22]. This review will focus on a sub-population of proteins, glycoproteins, as potential biomarkers and present challenges and promises of lectin-affinity based in biomarker discovery for oncology applications.

## 2. Targeting glycoproteins as biomarkers

Glycosylation is the most common post-translational modification (PTM) of proteins, involving the covalent attachment of a glycan, or sugar structure, to the peptide backbone [34]. Approximately 50% of all human pro-

teins are glycosylated, functioning in many biological processes, including development, immune response, cell division, inflammation, and cellular regulation [3, 30]. In addition to glycosylation's involvement in many cellular functions, reliant on a specific interaction that occurs with the extended glycan structure, it is also associated with proper protein folding. There are several types of glycan attachments in glycoproteins, with the two most common occurring as N-linked glycans or O-linked glycans. N-linked glycosylation involves the production of an oligosaccharide precursor and an en bloc transfer onto the side chain amide nitrogen of an asparagine residue of a newly synthesized protein in the endoplasmic reticulum (ER). O-linked glycosylation occurs at a later stage of protein processing in the Golgi apparatus, and involves the attachment of a glycan structure to a serine or threonine residue at the hydroxy oxygen of the amino acid [17].

Targeting glycosylation in biomarker discovery can be construed as a more direct approach based on the following thoughts. Aberrant glycosylation has been associated with a cancer disease state for many years [8]. Alterations in N-linked glycans that have a biological role in cell adhesion have been associated with cancer cell invasion and metastasis [34]. The majority of proteins localized at the cell membrane or secreted from the cell surface are glycosylated, thus may be detectable in the blood stream [36]. The structure of glycans, or the sites of glyco-post-translational modifications could be altered during cancer initiation and progression [15]. Glycoproteomic studies that analyze profile comparisons, of normal and disease states, provide potential new cancer biomarkers that can be used in clinical setting in the future [24]. In fact, the majority of the U.S. Food and Drug Administration (FDA) approved cancer biomarkers currently used in a clinical setting are glycoproteins or glycosylated (Table 1). Furthermore, targeting the differences in glycosylation, specifically alterations in the glycan structure, may provide opportunities to identify biomarkers that exhibit improved detection sensitivity and specificity [8]. Cancer biomarkers that detect specific glycoforms of a protein, resulting in increased sensitivity and specificity, may allow for earlier detection of a disease state [10]. Earlier detection of cancers may provide an improved therapeutic window and an increase in overall survival.

## 3. Lectin-based glycoprotein biomarker development

Methodologies to enrich glycoproteins have been employed previously and include the use of hydrazine

Table 1

Current US Food and Drug Administration (FDA) approved cancer biomarkers. Abbreviations: CA: cancer antigen, GIST: gastrointestinal stromal tumor, NMP22: nuclear matrix protein 22

Biomarker	Type	Glycosylated	Cancer type	Source	Clinical use
$\alpha$ -Fetoprotein (AFP)	Glycoprotein	Yes	Nonseminomatous testicular	Serum	Staging
Human chorionic gonadotropin- $\beta$	Glycoprotein	Yes	Testicular	Serum	Staging
CA 19-9	Glycan	Yes	Pancreatic	Serum	Monitoring
CA 125	Glycoprotein	Yes	Ovarian	Serum	Monitoring
Carcinoembryonic antigen (CEA)	Glycoprotein	Yes	Colon	Serum	Monitoring
Epidermal growth factor receptor (EGFR)	Protein	Yes	Colon	Colon	Therapy selection
KIT	Protein	Yes	GIST	Gastrointestinal tumor	Diagnosis, therapy selection
Thyroglobulin (Tg)	Protein		Thyroid	Serum	Monitoring
Prostate-specific antigen (PSA)	Glycoprotein	Yes	Prostate	Serum	Screening, monitoring
CA 15-3	Glycoprotein	Yes	Breast	Serum	Monitoring
CA 27-29	Glycoprotein	Yes	Breast	Serum	Monitoring
Oestrogen receptor and progesterone receptor	Protein	Yes	Breast	Breast tumor	Hormonal therapy selection
HER2/NEU	Protein	Yes	Breast	Breast tumor/Serum	Monitoring, prognosis, Therapy selection
NMP22	Protein		Bladder	Urine	Screening, monitoring
Fibrin/ Fibrin degradation protein	Protein	Yes	Bladder	Urine	Monitoring
Bladder tumor-associated antigen	Protein		Bladder	Urine	Monitoring

chemistry, carbohydrate-reactive antibodies, and studies that focus on analysis of glycan structures after release from the associated protein [8]. Most of these strategies incorporate mass spectrometry (MS) analysis after initial separation or enrichment to identify targeted glycoproteins or glycan structures. One methodology that has been employed extensively is the use of lectin affinity, and may hold promise in targeted glycoprotein analysis for cancer biomarker discovery.

Lectins are glycan-binding proteins that bind with great specificity to selected sugar moieties, or functional groups. The first lectin recognized, ricin, was isolated almost a century ago from castor seeds and additional lectins have been discovered since that are used to isolate sugar-bound molecules in biological systems. They are found both in animal and plant systems as well as lower organisms such as viruses, which use lectins to attach to host organism cells. A current list of lectins and the glycan structures they recognize are summarized in Table 2. With the multitude of lectins available to isolate sugar-containing molecules, their use has been employed in several studies both as single reagents for purification and enrichment, as well as in combination to identify a larger profile of glycosylated molecules. More recently, protein microarrays that in-

corporate lectins to serve as screening tools have also been developed.

The use of multi-lectin affinity has the benefit of obtaining a global glycoproteome, as the different affinities of each lectin included will yield a wider net of captured glycoproteins [45]. Development of multi-lectin chromatography (M-LAC) to study the plasma proteome has been detailed by one group, who developed single use agarose-bound lectin columns containing the mixture of lectins concanavalin A (Con A), jacalin (JAC), and wheat germ agglutinin (WGA) [33]. In addition to M-LAC, the authors also included depletion of albumin and IgG, two abundant plasma proteins, in an attempt to increase identification of low abundance proteins. To increase the robustness of this approach, making it more suitable to clinical studies, another group immobilized the same lectins, Con A, WGA, and JAC, onto a column [16]. Termed high-performance multi-lectin affinity chromatography (HP-M-LAC), this study sought to validate the approach of M-LAC and immunodepletion, while also generating a column that would allow for good reproducibility when analyzing multiple samples. One study that has employed the use of the M-LAC platform yielded several potential candidate proteins including thrombospondin-1 and 5,

Table 2  
Lectins and their corresponding recognized glycan moieties

Lectin (common name)	Abbv.	Glycan specificity
<i><math>\alpha/\beta</math>-D- Galactosyl general specificity</i>		
Agarius bisporous (common mushroom)	ABA	Gal $\beta$ 1-3GalNAc-serine/threonine
Allomyrina dichotoma (Japanese rhinoceros beetle)	AlloA	NeuAc $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc
Artocarpus integrifolia/Jacalin (jackfruit seeds)	ALL, JAC	Gal $\beta$ 1-3GalNAc, Gal $\alpha$ 1-6Gal
Arachis hypogaea (peanut)	PNA	Gal $\beta$ 1-3GalNAc
Ricinus communis (castor bean)	RCA <sub>120</sub>	Gal $\beta$ 1-4GalNAc
<i>D-N-Acetylgalactosaminyl general specificity</i>		
Dolichos biflorus (horse gram)	DBA	GalNAc $\alpha$ 1-3GalNAc GalNAc $\alpha$ 1-3Gal
Glycine max (soybean)	SBA	GalNAc $\alpha$ 1-3GalNAc GalNAc $\alpha$ v $\beta$ 1-3/4Gal
Helix pomatia (common snail)	HPA	$\alpha$ -GalNAc, GalNAc $\beta$ 1-4Gal
Phaseolus lunatus (lima bean)	LBA	GalNAc $\alpha$ 1-3(Fuc $\alpha$ 1-2)Gal $\beta$ GalNAc $\alpha$ 1-2Gal $\beta$
Vicia villosa (hairy vetch)	VVA, VVL	GalNAc-Ser/Thr
<i>D-N-Acetylglucosamine general specificity</i>		
Datura stramonium (jimson weed)	DSA, DSL	(GlcNAc $\beta$ 1-4) <sub>2-4</sub> , Gal $\beta$ 1-4GlcNAc
Griffonia simplicifolia II	GS-II	GlcNAc $\alpha$ 1-4Gal $\beta$ 1-4GlcNAc
Lycopersicon esculentum (tomato)	LEL	(GlcNAc $\beta$ 1-4) <sub>1-4</sub>
Triticum vulgare/Wheat Germ Agglutinin (wheat germ)	WGA	(GlcNAc $\beta$ 1-4) <sub>2-5</sub> , Neu5Ac Man $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc
<i><math>\alpha</math>-L-fucosyl general specificity</i>		
Aleuria aurantia (orange peel fungus)	AAL	Fuc $\alpha$ 1-6/3GlcNAc
Anguilla anguilla (European eel)	AAA	$\alpha$ -L-Fuc
Lotus tetragonolobus (winged pea)	LTL	Fuc $\alpha$ 1-2Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc
Ulex europaeus I (common gorse)	UEA I	Fuc $\alpha$ 1-2Gal $\beta$
<i>D-Mannosyl general specificity</i>		
Canavalia ensiformis/Concanavalin A (jack bean)	Con A	Branched N-linked hexa-saccharide
Galanthus nivalis (snowdrop)	GNA	Man $\alpha$ 1-3Man
Lens culinaris (lentil)	LCA	Fuc $\alpha$ 1-6GlcNAc-Asn containing N-linked oligosaccharides
Pisum sativum (pea)	PSA	Fuc $\alpha$ 1-6GlcNAc-Asn containing N-linked oligosaccharides
<i>Neu5Ac general specificity</i>		
Maackia amurensis agglutinin II	MAA II	NeuAc $\alpha$ 2-3Gal $\beta$ 1-4Glc/GlcNAc
Maackia amurensis leukoagglutinin	MAL	Neu5Ac/Gc $\alpha$ 2,3Gal $\beta$ 1, 4 Glc(Nac)
Maackia amurensis hemoagglutinin	MAH	Neu5Ac/Gc $\alpha$ 2,3Gal $\beta$ 1, 3(Neu5Ac $\alpha$ 2,6)GalNAc
Sambucus nigra (black elderberry)	SNA	NeuAc $\alpha$ 2-6Gal/GalNAc
<i>Lectins with heterogeneous specificities</i>		
Cicer arietinum (chick pea)	CPA	Complex structure, binding inhibited by human IgM, fetuin
Euonymus europaeus (European spindle)	EEA	Gal $\alpha$ 1-3(Fuc $\alpha$ 1-2)Gal $\beta$ 1-3/4GlcNAc
Phaseolus vulgaris erythroagglutinating (common bean)	PHA-E	N-linked bi-antennary
Phaseolus vulgaris leukoagglutinating (common bean)	PHA-L	N-linked tri-tetra-antennary

alpha-1B-glycoprotein, serum amyloid P-component, and tenascin-X from the serum of breast cancer patients to be investigated in future studies [44]. In addition to the use of a multi-lectin affinity column, which incorporated the mixture of lectins Con A, JAC, WGA immobilized onto the column resin, the authors also employed immunodepletion, iso-electric focusing (IEF) fractionation, and LC-MS identification. Further analysis of the plasma proteome following a similar methodology, of immunodepletion of albumin and IgG followed by M-LAC established the reproducibility of this platform [33].

To continue with the theme of a multi-lectin approach, another study employed the use of biotinylated lectins E-PHA, AAL, and DSL incubated with ovarian tumor tissue lysate, leading to glycoprotein capture [1]. The biotinylated lectins, bound with glycoproteins, were then removed from the tissue lysate with the aid of paramagnetic streptavidin beads, and the bound glycoproteins eluted. In a previous study, the authors identified changes in glycotransferase expression in ovarian tissue, the results of which influenced their lectin selection. In doing so, the authors were able to identify 47 potential lectin-reactive markers in ovarian tumors. An additional aspect of the previously mentioned study includes the validation of several of the identified candidate proteins in serum using a single labeled lectin as an antibody and Western Blotting. Their results indicated some of their candidate marker proteins were bound to a specific lectin, implying that it was the glycan structure, and not the protein alone, that could be used to distinguish between malignant and normal patient serum. After assessing the value of several strategies to enrich and identify glycoproteins, one group decided to perform a “double” multi-lectin capture approach, wherein patient serum was first passed through a multi-lectin column containing Con A, WGA, and SNA, the resulting eluate was then tryptic digested, and the digested peptides were, again, enriched using the multi-lectin column [5]. This two stage enrichment of glycoproteins, and then glycopeptides, yielded approximately 45 proteins that were differentially expressed in the sera of breast cancer patients and non-disease patients, and when combined with ESI-LC-MS, allowed for glycosylation-site mapping of the identified proteins.

Multi-lectin capture of glycoproteins allows for a global analysis of changes in glycosylation patterns of proteins between disease and non-disease states. As the previously described experiments illustrate, enrichment of glycoproteins with lectins allows for identifi-

cation for putative biomarkers, especially when those methodologies employ the analysis of patient serum. All of these studies identify potential candidates that would need further validation studies, but establish a groundwork of which future studies can be based on.

A single lectin enrichment approach of glycoproteins appears in greater frequency in the literature, wherein capturing glycosylated proteins of a specific moiety may allow for the identification of candidate markers that can be validate in future studies. In doing so, the resulting glycoprotein fraction is reduced in complexity compared to those derived from multi-lectin affinity. This reduction in complexity may also help for the identification of proteins expressed in low abundance that may not be detected in any other methodology. One study focused on analyzing samples from normal individual and patients with lung adenocarcinoma with the use of the lectin WGA in attempt to identify glycoprotein biomarkers [11]. First, serum samples from normal individuals and patients with lung adenocarcinoma were screened using FITC-labeled lectin staining method, and identified the lectin WGA as exhibiting the highest specificity for bound glycoproteins in the serum samples. Glycoproteins in serum samples were then enriched using the WGA lectin, and the enriched fractions were then loaded onto 2-DE gels and analyzed via 2D DIGE to observe expression differences between the two sample groups. Interestingly, the authors noted that WGA incubation not only enriched for the glycoproteins that bound specifically, but also resulted in the removal of high abundant serum proteins, that would normally mask potential markers of disease. Using the 2-DE methodology, the study yielded the identification of 39 differentially expressed protein spots. Another study that analyzed differences in glycosylation in lung cancer used the lectin, Con A, and employed 2-DE to detect changes in protein expression [40]. In this study, serum samples from healthy patients and lung cancer patients were enriched with Con A, and resulting glycoprotein samples separated on a 2D gel, spots excised, and analyzed via ESI-LC-MS. Several candidate proteins were identified, and again, the use of lectin affinity resulted in enrichment of glycoproteins, while also removing highly abundant serum proteins such as IgG and albumin. Analysis of colon cancer serum was attempted to reveal the potential link of  $\beta$ -haptoglobin and colon cancer. Increased  $\beta$ -haptoglobin expression has often been associated with various pathobiological processes, thus one group sought to determine whether altered glycosylation of  $\beta$ -haptoglobin could serve as a disease marker in colon cancer [28]. The lectin *Aleuria*

*aurantia* (AAL) was used to compare the level of fucosylation between colon cancer patients, irritable-bowel syndrome (IBS) patients, and normal subjects, and researchers observed that increased fucosylation of  $\beta$ -haptoglobin correlates not only with a cancer state, but was also indicative of the cancer stage.

To evaluate changes in glycoprotein expression in a model of breast cancer progression, one group studied the membrane fraction derived from two breast cancer cell line phenotypes: a normal, premalignant cell line and a malignant, metastatic cell line [42]. The lectins WGA and Con A bound to agarose were individually packed into spin columns, and the membrane fraction of each cell line incubated with agarose-bound lectins, and unbound proteins washed away. The resulting bound fraction was tryptically digested and analyzed via LC-MS/MS. To quantitate the glycoprotein differences between the two cell lines, label-free spectral counting was employed, and 27 membrane glycoproteins were detected and determined to be differentially expressed. Although this study analyzed cell lines, the approach could be applied to more relevant clinical samples such as tumor tissue biopsies. An additional study that used individual, agarose-bound lectins in spin columns, examined the difference in expressed glycoproteins between normal and pancreatic cancer serum [46]. In this study, not only were expression changes in glycoprotein analyzed and identified via LC-MS, but also the changes in expression of the glycan structures via MALDI. Interestingly, the latter aspect revealed that not only will a change in abundance of a protein relate to a disease state, but also a change in the glycosylation modification.

In one study, researchers overexpressed the glycotransferase glycotransferase cosaminyltransferase V (GnT-V) in WiDr cells to mimic the conditions witnessed in colon cancer cells, and analyzed glycosylation differences between the control and experimental cell lines using the lectin phytohemagglutinin-L<sub>4</sub> (L-PHA) [2]. The authors also employed the mass spectrometry based technique of multiple-reaction monitoring (MRM) to quantify the glycosylation changes expressed between the control and GnT-V overexpressed cells. The two glycoproteins, TIMP1 and PTP $\kappa$ , both of which are target proteins of GnT-V, were selected for MRM quantitation to assist in a proof-of-concept experiment that MRM quantitation could be employed, in conjunction with lectin affinity enrichment, to detect differential expression of glycosylated proteins.

An interesting proof-of concept experiment developed incorporated lectin affinity enrichment combined

with an immunodepletion step involving the use of antibodies raised against all proteins in the sub-proteome (derived from enriched glycosylated protein fraction) of healthy humans [25]. This methodology, termed Tandem Affinity Depletion, employed the use of the lectin, WGA, and antibodies derived from llamas, as the authors detailed that old world camels generate antibodies containing only a heavy chain, which are more robust in terms of pH and temperature. Using a known cancer marker, CEA, spiked into the plasma sample from several colon cancer patients, the authors were able to enrich the glycoprotein by a factor of 600–800, validating their approach. The technique, in contrast to immunodepletion experiments which often serve to identify as many low abundance proteins as possible, instead focused on reducing the complexity of the plasma proteome to better differentiate between normal and disease states.

The development of methodologies that allow for quantifying of glycoproteins in serum has been the focus of several recent studies. One group studied the changes of sialylation, after enrichment of sialylated glycoproteins with SNA lectin, in serum of patients with prostate cancer [37]. Termed lectin-directed tandem labeling, distinct sample group enriched glycoproteins were isotopically labeled with either light or heavy forms of acetic anhydride, and then quantified via nano-ESI-LC-MS/MS. Researchers were able to determine that it was an increase in sialylation of several glycoproteins, but not the protein concentration (as determined by analysis of non-lectin enriched fractions), that correlated with a disease state. Another study attempted to identify N-linked glycoproteins associated with hepatocellular carcinoma (HCC) in plasma without extensive enrichment steps, and quantify the differences via iTRAQ labeling [18]. Knowing that identification of glycoproteins via MS are often masked by the signal of nonglycosylated proteins, researchers generated an exclusion list of peptide masses from nonglycosylated proteins, thus training the MS instrument to acquire spectra from glycosylated peptides only. To increase the pool of potential identified glycosylated peptides, the inclusion of lectin enrichment was employed. After identifying several glycopeptides in the serum of HCC patients, iTRAQ labeling was coupled to the exclude peak list technique, to elucidate abundance differences between normal and HCC plasma samples. Two glycoproteins, vitronectin and antithrombin III, were identified via this method.

Covalent attachment of lectins to magnetic beads to enrich for glycopeptides or glycoproteins was de-

scribed, and the resulting bound fractions analyzed using MALDI [38]. To facilitate a high-throughput automated procedure, one group employed the use of a single lectin coupled to magnetic beads, resulting in a one-step glycoprotein purification platform [20]. Use of a robotic liquid handling system and 96-well plate format, three lectins, Con A, JAC, and WGA, were coupled to commercial magnetic beads individually, and the resulting pulled-down protein fractions were analyzed via 2-DE and the spot identified using MS. The benefit of this methodology is increased reproducibility as well application of a high-throughput biomarker and glycoproteomic analysis technique. Combined with bioinformatic analysis, the resulting datasets exhibit the potential to reveal and identify alterations in glycosylation profiles of disease states.

Previous studies have employed the use of lectin conjugated to beads or immobilized onto a column. One study that differs in the approach of lectin affinity is the use of free lectins to bind to glycoproteins, and then recover the enriched glycoproteins by either ammonium sulfate precipitation, or the use of a membrane filter unit and centrifugation [43]. Using a multitude of lectins individually in parallel, including Con A, WGA, *Ricinus communis* agglutinin (RCA<sub>120</sub>), *Sambucus sieboldiana* lectin (SSA), and *Maackia amurensis* agglutinin (MAM), the two distinct methodologies were employed and evaluated. The first, which employed ammonium sulfate precipitation oligosaccharides derived from albumin, revealed that the recovered oligosaccharides moiety recovered in each lectin fraction was specific to the lectin employed. The second, incorporating free lectins incubated with tryptically digested glycopeptides from glycoprotein standard on top of a centrifugation unit membrane, showed an increase of enrichment for glycopeptides in each of the fractions, with individual lectins showing more specificity for individual, distinct proteins found in the protein standard. The results of these two methodologies illustrate that the use of a free lectin solution approach was comparable to those results obtained in other affinity chromatography experiments.

Many of the previously mentioned studies require substantial amounts of sample, and in instances where sample amount is limited, one group developed a nanoscale Con A-chelating monolithic capillary [9]. The use of this monolithic capillary would allow for complex, biological samples occurring in minute amounts to be analyzed and enriched for glycoproteins. To assess the value of their platform, urine samples from healthy individuals and bladder cancer patients

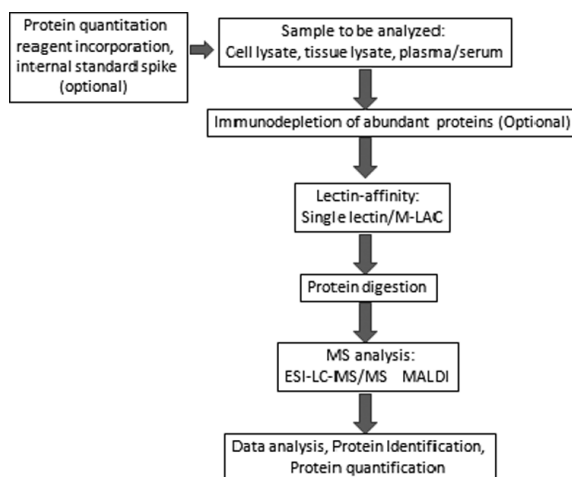


Fig. 1. Flowchart of generalized lectin-affinity strategies. Incorporation of quantitation reagents (i.e. iTRAQ or SILAC) would occur prior to cell lysis. Spiking of internal standard prior to analysis would allow for quantitation of tissue, plasma, or serum samples. Immunodepletion is an optional inclusion (as is other fractionation steps such as SEC, SAX, or iso-electric focusing), based on the complexity and composition of lectin-bound fractions.

were tested, using only 10  $\mu$ g of initial proteins. When compared to a standard lectin column, the nanoscale monolithic column resulted in the identification of more glycoproteins.

As these studies indicate, use of single lectin affinity has the benefit of reducing the complexity of a proteome, in most cases the plasma, or serum, proteome, thus identifying potential candidate biomarkers. A general workflow diagram for lectin-affinity based methodologies is illustrated in Fig. 1. Single lectins appear to not only enrich for glycoproteins, but also indirectly result in an immunodepletion of highly abundant serum proteins. The multitude of studies incorporating single lectin affinity generates a lengthy list of potential biomarkers that would need further validation in larger patient groups, and the use of methodologies that emphasize a high-throughput platform may be the best approach in doing so.

Several drawbacks of the lectin affinity approach are that overall, they are quite time-consuming with the multiple stages including enrichment, protein identification, and data analysis, and require a great deal of expertise regarding mass spectrometry analysis. In addition, the use of lectin affinity is based on a pull-down approach, thereby introducing the occurrence of additional proteins complexed with the lectin-target protein. Although this latter aspect has not been reported to interfere significantly with the selection of candidate proteins derived from lectin affinity experiments, it is

a concern that could arise when validating potential candidate biomarkers.

The development of protein microarrays incorporating lectins has also been employed to study changes in glycosylation. As noted, the use of lectin microarrays are beneficial as they are more suitable for high-throughput and systematic analysis of protein glycosylation, in contrast to more time consuming methodologies that incorporate affinity chromatography and MS analysis [46]. In a brief communication, the development of lectin microarrays was detailed, and experimentally shown to provide a distinct, glycosylation pattern for selected protein standards, comparable to previous methods [32].

To develop potential diagnostic tools in identifying aggressive prostate cancer, lectin microarrays were employed [19]. First, the authors used PSA, a known glycosylated protein found in prostate cancer to evaluate their methodology, and then expanded by applying their platform to differentiate glycosylation patterns between non-aggressive prostate cancer and aggressive prostate cancer. Using lectin microarrays, the researchers were able to detect several isoforms of PSA and membrane metallo-endopeptidase (MME), both are which are highly expressed in prostate cancer. Lectin microarrays were also used to determine surface glycan structures, and thus identify potential biomarkers of breast cancer stem-cells [39]. In attempt to generate additional cell surface markers that would be used to differentiate between cancer stem-like cells and cancer cell lines, the cell lines were analyzed with the use of lectin microarrays. The researchers revealed that the breast cancer cell line, MCF7, which can be grown under standard and sphere culture conditions, displayed distinct glycosylation patterns in response to the culture conditions. To identify differences in glycan expression between normal and tumorigenic breast cancer cell lines, researchers used lectin arrays [6]. The inclusion of lectin arrays in the study attempted to reveal cell surface expression changes in carbohydrates between invasive and noninvasive cell lines, and in doing so, the study also revealed a glycosylation pattern that related to tissue-specific homing in metastasis. As the authors noted, the differences revealed could be functionally relevant in the development of diagnostic and therapeutic targets for detecting and treating metastasis.

One study incorporated both multi-lectin affinity and lectin microarrays to analyze the glycosylation pattern in serum from patients with pancreatic cancer in an attempt to identify potential serum biomarkers [29].

First, glycoproteins from serum were enriched using a general lectin column, and the resulting fraction was separated via reversed-phase HPLC to reduce complexity. The researchers selected lectins that exhibited a great deal of specificity to avoid non-specific protein binding, and the resulting array results indicated there were differences in the glycosylation patterns between the two sample groups. As detailed by the authors, this approach not only allowed for the identification of glycosylation pattern differences, but also allows for the identification protein glycosylation differences related to the disease state. To identify carbohydrate-based tumor biomarkers, comparison of serum from patients with lung cancer and healthy individuals in a platform using lectin-coupled ProteinChip arrays was established [41]. To reduce interference from glycosylated serum proteins, immunodepletion was used prior to enrichment of the remaining glycoproteins with the lectins JAC and SNA coupled to ProteinChip arrays. The researchers discovered the specific glycosylation site for the protein apoC-III, and two, distinct glycan structures on this protein, one of which was elevated in the serum of lung cancer patients.

#### **4. Concluding remarks**

Biomarker research has been ongoing for many years now, and although perhaps hundreds of candidate biomarkers have been suggested, only a few have been selected for clinical validation and eventually approved for use in clinical settings. There are a number of factors such as the complexity of cancer biology, the selection bias in the biomarker discovery processes, and the difficulty to conduct prospective clinical validation studies, which all contribute to this low rate of biomarker development but are beyond the scope of this review. As discouraging as that may be, the benefit of biomarkers in the field of medicine and health management is immeasurable. Specifically in cancer, biomarkers may not only lead to early detection or verify a disease state, but may also have a role in diagnosing tumor stage, and classifying cancers based on their genetic/biologic properties. It is this latter aspect, wherein cancer biomarkers could play the biggest role in the development of personalized medicine. No longer would treatment be a trial and error process, but rather a directed-therapy based on the molecular makeup of the diseased individual.

With the advancement in technologies, we are now able to explore the entire molecular patterns in cells



or tissues. Biomarker discoveries using these global approaches were criticized as “fishing expeditions” in the past, but now show great potential in biomarker discovery. Using proteomics, we are able to compare the proteome profile between a disease and non-disease state, revealing a treasure trove of valuable information, however, limitations of the technologies used in proteomics currently makes deep analysis of biological fluids and tissues a challenge. Deep-mining of the proteome could potentially yield clues as to changes in normal biological functions due to a disease state, and with the approaches currently in use, generation of sub-proteomes may be the best method to accomplish this goal. One such sub-proteome, the glycoproteome, appears to be the most beneficial to study, for the reasons previously mentioned in this review, and lectin-based techniques may prove to be a preferred approach. As multiple studies have revealed, the use of lectin-based techniques to analyze the glycoproteome has yielded several proteins of interest for various cancers from different biological sources. These lectin-based strategies can be viewed as analyzing the sub-proteome of a sub-proteome, resulting in a deep-mining of the human proteome.

Identifying differences in protein profiles between normal and disease states may reveal potential biomarkers, but to determine increased expression of a protein(s) due to the disease state requires precise quantification. In this review, we discussed studies using 2D-DIGE, iTRAQ, and label-free methods to assess and quantitate relative protein expression differences, however, these methods are limited in the detection of slight protein expression differences, especially in samples with a large dynamic range. In contrast, stable isotope labeling by amino acids in cell culture (SILAC) is simple, highly sensitive, accurate in samples with a large dynamic range, and its workflow inherently reduces experimental error, allowing for the quantification of even minute protein expression differences [23]. Surprisingly, the incorporation SILAC with lectin-affinity is rare, and may be due to many of these studies analyzing serum or tissue samples, wherein SILAC requires living organisms or autotrophic samples. One technique to address shortcoming of the SILAC methodology is super-SILAC. The super-SILAC method involves the use of several SILAC-labeled cell lines that match the tissue of interest, serving as an internal standard for quantitation of proteins derived from the tissue sample [26]. The super-SILAC approach is sensitive and accurate, thus inclusion of this quantitation method for future biomarker discovery experiments may better re-

veal subtle differences in the proteomes of normal and disease states.

Although it appears many of the current lectin-based technologies are limited in their ability to be high-throughput, specific, sensitive, and reproducible, the combination of these distinct methodologies may be enough to overcome the limitations of each individual technique. The future may hold the development of a biomarker discovery/validation platform that can be applied in the detection candidate markers in many disease states. The development of such a platform would result in the transition of the identified biomarkers into clinically relevant applications.

## References

- [1] K.L. Abbot, J. Lim, L. Wells, B.B. Benigno, J.F. McDonald, M. Pierce, Identification of candidate biomarkers with cancer-specific glycosylation in the tissue and serum of endometrioid ovarian cancer patients by glycoproteomic analysis. *Proteomics* 10 (2010), 470-481.
- [2] Y.H. Ah, Y. Kim, E.S. Ji, J.Y. Lee, J. Jung, J.H. Ko, J.S. Yoo, Comparative quantitation of aberrant glycoforms by lectin-based glycoprotein enrichment coupled with multiple-reaction monitoring mass spectrometry. *Anal Chem* 82 (2010), 4441-4447.
- [3] H.J. An, S.R. Kronewitter, M.L. A de Leoz, C.B. Lebrilla, Glycomics and disease markers. *Current Opinion in Chemical Biology* 13 (2009), 601-607.
- [4] Biomarkers Definitions Working Group, Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69 (2001), 89-95.
- [5] C.D. Calvano, C.G. Zamboni, O.N. Jensen, Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *Proteomics* 71 (2008), 304-317.
- [6] S. Chen, T. Zheng, M.R. Shortreed, C. Alexander, L.M. Smith, Analysis of cell surface carbohydrate expression patterns in normal and tumorigenic human breast cell lines using lectin arrays. *Anal Chem* 79 (2007), 5698-5702.
- [7] E.P. Diamandis, Cancer biomarkers: can we turn recent failures into success? *J Natl Cancer Inst* 102 (2010), 1462-1467.
- [8] P.M. Drake, W. Cho, B. Li, A. Prakobphol, E. Johansen, N.L. Anderson, F.E. Regnier, B.W. Gibson, S.J. Fisher, Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem* 56 (2010), 223-236.
- [9] S. Feng, N. Yang, S. Pennathur, S. Goodison, D.M. Lubman, Enrichment of glycoproteins using nanoscale chelating concanavalin a monolithic capillary chromatography. *Anal Chem* 81 (2009), 3776-3783.
- [10] G.W. Hart and R.J. Copeland, Glycomics hits the big time. *Cell* 143 (2010), 672-676.
- [11] P. Hongsachart, R. Huang-Liu, S. Sinchaikul, F. Pan, S. Phutrakut, Y. Chaung, C. Yu, S. Chen, Glycoproteomic analysis of WGA-bound glycoprotein biomarkers in sera from patient with lung adenocarcinoma. *Electrophoresis* 30 (2009), 1206-1220.
- [12] G.L. Hortin, D. Sviridov, The dynamic range problem in the analysis of the plasma proteome. *Proteomics* 73 (2010), 629-636.

- [13] J.M. Jacobs, J.N. Adkins, W. Qian, T. Liu, Y. Shen, D.G. Camp II, R.D. Smith, Utilizing human blood plasma for proteomic biomarker discovery. *J Proteome Res* (2005), 1073-1085.
- [14] S. Khleif, J.H. Doroshow, W.N. Hait, AACR-FDA-NCI cancer biomarkers collaborative consensus report: advancing the use of biomarkers in cancer drug development. *Clin Can Res* 16 (2010), 3299-3318.
- [15] E.H. Kim and D.E. Misek, Glycoproteomics-based identification of cancer biomarkers. *Int J Proteomics* (2011), 84-94.
- [16] M. Kullolli, W.S. Hancock, M. Hincapie, Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins. *J Sep Sci* 31 (2008), 2733-2739.
- [17] A. Larkin and B. Imperiali, The expanding horizons of asparagine-linked glycosylation. *Biochemistry* 50 (2011), 4411-4426.
- [18] H. Lee, K. Na, E. Choi, K.S. Kim, H. Kim, Y. Paik, Simple method for quantitative analysis of n-linked glycoproteins in hepatocellular carcinoma specimens. *J Proteome Res* 9 (2010), 308-318.
- [19] Y. Li, S. Tao, G.S. Bova, A.Y. Liu, D.W. Chan, H. Zhu, H. Zhang, Detection and verification of glycosylation patterns of glycoproteins from clinical specimens using lectin microarrays and lectin-based immunosorbent assays. *Anal. Chem.* 83 (2011), 8509-8516.
- [20] D. Loo, A. Jones, M.M. Hill, Lectin magnetic bead array for biomarker discovery. *J Proteome Res* 9 (2010), 5496-5500.
- [21] J.A. Ludwig and J.N. Weinstein, Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer* 5 (2005), 845-856.
- [22] S. Makawita and E.P. Diamandis, The bottleneck in the cancer biomarker pipeline and protein quantitation through mass-spectrometry-based approaches: current strategies for candidate verification. *Clin Chem* 56 (2010), 212-222.
- [23] M. Mann and N.L. Kelleher, Precision proteomics: The case for high resolution and high mass accuracy. *PNAS* 105 (2008), 18132-18138.
- [24] D.L. Meany and D.W. Chan, Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clin Proteomics* 8 (2011), 7-21.
- [25] N. Mortezai, S. Harder, C. Schnabel, E. Moors, M. Gaulty, H. Schluter, C. Wagener, F. Buck, Tandem affinity depletion: a combination of affinity fractionation and immunoaffinity depletion allows the detection of low-abundance components in the complex proteomes of body fluids. *J Proteome Res* 9 (2010), 6126-6134.
- [26] T.A. Neuber and P. Tempst, Super-SILAC for tumors and tissues. *Nature Methods* 7 (2010), 361-362.
- [27] S. Ogino, J. Galon, C.S. Fuchs, G. Dranoff, Cancer immunology-analysis of host and tumor factors for personalized medicine. *Nat Rev Clin Oncol* (2011), doi: 10.1038/nr-clinonc.2011.122.
- [28] S. Park, S. Lee, N. Kawasaki, S. Itoh, K. Kang, S.H. Ryu, N. Hashii, J. Kim, J. Kim, J.H. Kim,  $\alpha$ 1-3/4 fucosylation at Asn 241 of  $\beta$ -haptoglobin is a novel marker for colon cancer: a combinatorial approach for development of glycan biomarkers. *Int J Cancer* (2011), doi: 10.1002/ijc.26288.
- [29] T.H. Patwa, J. Zhao, M.A. Anderson, D.M. Simeone, D.M. Lubman Screening of glycosylation patterns in serum using natural glycoprotein microarrays and multi-lectin fluorescence detection. *Anal Chem* 78 (2006), 6411-6421.
- [30] T. Patwa, C. Li, D.M. Simeone, D.M. Lubman, Glycoprotein analysis using protein microarrays and mass spectrometry. *Mass Spectrometry Reviews* 29 (2010), 830-844.
- [31] M.S. Pepe, R. Etzioni, Z. Feng, J.D. Potter, M.L. Thompson, M. Thornquist, M. Winget, Y. Yasui, Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 93 (2001), 1054-1061.
- [32] K.T. Pilobello, L. Krishnamoorthy, D. Slawek, L.K. Mahal, Development of a lectin microarray for the rapid analysis of protein glycopatterns. *ChemBio Chem* 6 (2005), 985-989.
- [33] T. Plavina, E. Wakshull, W.S. Hancock, M. Hincapie, Combination of abundant protein depletion and multi-lectin affinity chromatography (M-LAC) for plasma protein biomarker discovery. *J Proteome Res* (2007), 662-671.
- [34] C.A. Reis, H. Osorio, L. Silva, C. Gomes, L. David, Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol* (2010), 322-329.
- [35] J.M. Rhea and R.J. Molinaro, Cancer Biomarkers: Surviving the journey from bench to bedside. *MLO Med Lab Obs* 43 (2011), 10-18.
- [36] R. Schiess, B. Wollscheid, R. Aebersold, Targeted proteomic strategy for clinical biomarker discovery. *Mol Oncol* 3 (2009), 33-44.
- [37] V. Shetty, Z. Nickens, P. Shah, G. Sinnathamby, O.J. Semmes, R. Philip, Investigation of sialylation aberration in n-linked glycoproteins by lectin and tandem labeling (LTL) quantitative proteomics. *Anal Chem* 82 (2010), 9201-9210.
- [38] K. Sparbier, S. Koch, T. Wenzel, M. Kostrzewa, Selective Isolation of glycoproteins and glycopeptides for maldi-tof ms detection supported by magnetic particles. *Journal of Biomolecular Techniques* 16 (2005), 407-413.
- [39] S. Tao, Y. Li, J. Zhou, J. Qian, R.L. Schnaar, Y. Zhang, I.J. Goldstein, H. Zhu, J.P. Schneck, Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology* 18 (2008), 761-769.
- [40] T.T. Thanh, N.T.M. Phuong, N.B. Nhi, P.V. Chi, Changes of glycoproteins in lung cancer patients. *J Proteomics Bioinform* 1 (2008), 11-16.
- [41] K. Ueda, Y. Fukase, T. Katagiri, N. Ishikawa, S. Irie, T. Sato, H. Ito, H. Nakayama, Y. Miyagi, E. Tsuchiya, N. Kohno, M. Shiwa, Y. Nakamura, Y. Daigo, Targeted serum glycoproteomics for the discovery of lung cancer-associated glycosylation disorders using lectin-coupled ProteinChip arrays. *Proteomics* 9 (2009), 2182-2192.
- [42] Y. Wang, X. Ao, H. Vuong, M. Konanur, F.R. Miller, S. Goodison, D.M. Lubman, Membrane glycoproteins associated with breast tumor cell progression identified by a lectin affinity approach. *Journal of Proteome Research* 7 (2008), 4313-4325.
- [43] M. Yodoshi, T. Oyama, K. Masaki, K. Kakehi, T. Hayakawa, S. Suzuki, Affinity entrapment of oligosaccharides and glycopeptides using free lectin solution. *Analytical Sciences* 27 (2011), 395-400.
- [44] Z. Zeng, M. Hincapie, S.J. Pitteri, S. Hanash, J. Schalkwijk, J.M. Hogan, H. Wang, W.S. Hancock, A proteomics platform combining depletion, multi-lectin affinity chromatography (M-LAC), and isoelectric focusing to study the breast cancer proteome. *Anal Chem* 83 (2011), 4845-4854.
- [45] L. Zhang, H. Lu, P. Yang, Specific enrichment methods for glycoproteome research. *Anal Bioanal Chem* 396 (2010), 199-203.
- [46] J. Zhao, D.M. Simeone, D. Heldt, M.A. Anderson, D.M. Lubman, Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *Journal of Proteome Research* 5 (2006), 1792-1802.