Review

Towards an integrated proteomic and glycomic approach to finding cancer biomarkers

Allen D Taylor*, William S Hancock†, Marina Hincapie†,
Naoyuki Taniguchi‡§ and Samir M Hanash*

Addresses: *Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA. †Department of Chemistry and Chemical Biology, Barnett Institute, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA. ‡Systems Glycobiology Group, Disease Glycomics Team, Advanced Science Institute, RIKEN, Japan. §Department of Disease Glycomics, Research Institute for Microbial Diseases, Osaka, Japan.

Correspondence: Samir M Hanash. Email: shanash@fhcrc.org

Abstract

Advances in mass spectrometry have had a great impact on the field of proteomics. A major challenge of proteomic analysis has been the elucidation of glycan modifications of proteins in complex proteomes. Glycosylation is the most structurally elaborate and diverse type of protein post-translational modification and, because of this, proteomics and glycomics have largely developed independently. However, given that such a large proportion of proteins contain glycan modifications, and that these may be important for their function or may produce biologically relevant protein variation, a convergence of the fields of glycomics and proteomics would be highly desirable. Here we review the current status of glycoproteomic efforts, focusing on the identification of glycoproteins as cancer biomarkers.

Introduction

The sequencing of the human genome and the spectacular advances in mass spectrometry (MS) have had a substantial impact on the field of proteomics. MS has evolved from a tool for the identification and characterization of isolated proteins (mass peak profiling) to a platform for interrogating complex proteomes and identifying differentially expressed proteins, whether in cells, tissues, or body fluids, by matching mass spectra to sequence databases. Remaining challenges that are gradually being conquered include increased depth and throughput of proteomic analysis and increased emphasis on elucidation of post-translational modifications. Elucidation of glycan modifications of proteins in complex proteomes has been a major challenge for proteomics. Glycosylation is the most structurally elaborate and diverse kind of protein post-translational modification and has been shown to have significant impact on protein function and confirmation. It has been shown that more than half of all proteins in human serum are glycosylated [1], so glycoproteins are particularly interesting in serum diagnostics for cancer and other diseases. Glycomics and proteomics have largely developed independently, but a convergence of the two fields is highly desirable.

Here we review the current status of glycoproteomic efforts relevant to the identification of cancer biomarkers. We also discuss what lies ahead and various options for comprehensive analyses that encompass both the cancer proteome and its related glycome in search of biomarkers for early cancer detection, for disease classification, and for monitoring response to cancer therapy.

Glycoprotein alterations in cancer

Glycan modification of proteins occurs primarily at asparagine residues (*N*-linked glycans) and at serine or threonine residues (*O*-linked glycans). Typically glyco-proteins that have complex glycan structures are membrane-bound or secreted. Proteins with glycosylation that are pre-dominantly nuclear or cytoplasmic often have a mono-saccharide *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) at serine residues, which is also a site of protein phosphory-lation. Research going back several decades has yielded evidence that glycosylation is altered in cancer. Some cancer cells have proteins with such differences in glycosylation from non-cancerous cells that the proteins are categorized as tumor-associated antigens, and they may even elicit a humoral immune response, as reviewed a quarter of a century ago by Hakomori [2] and recently by others [3]. Many initial studies with naturally occurring and hybridoma-derived monoclonal antibodies that were targeted against tumor antigens yielded evidence of reactivity that was directed against carbohydrate epitopes, as in the case of so-called oncofetal antigens [4]. Some glycomic alterations found in cancer cells have been attributed to the activity and localization in the Golgi of glycosyltranferases.

Mucins are among the most investigated glycoproteins pro-duced by epithelial cancer cells. Mucins contain numerous *O*-glycans that are clustered along the Ser/Thr/Pro-rich 'variable number of tandem repeat' (VNTR) domains and have several cancer-associated structures, including the Thomsen-Fredenreich antigen (T-antigen), the Thomsen-nouveau antigen (Tn-antigen), and certain Lewis antigens [5]. Cell-surface-bound and secreted mucin glycoproteins contain *N*-acetylgalactosamine (GalNAc)-Ser/Thr *O*-linked sugars that constitute more than half of the mass of the mucin. The glycans of mucins expressed on the cell surface are involved in interactions with the microenvironment. Several well known cancer serological biomarkers are mucins or mucin-like glycoproteins, including tumor-associated antigens cancer antigen 125 (CA125), cancer antigen 19-9 (CA19-9), cancer antigen 15-3 (CA15-3), and cancer antigen 72-4 (CA72-4). The discovery of CA125 resulted from an assay developed using a murine monoclonal antibody that reacts with an antigen found to be common to most non-mucinous epithelial ovarian carcinomas [6]. The application of this assay showed that only 1% of 888 apparently healthy subjects and 6% of 143 patients with non-malignant disease had serum CA125 levels above 35 units (U)/ml. In contrast, 83 of 101 patients with surgically demonstrated ovarian carcinoma had elevated levels of antigen above 35 U/ml [6]. CA125 was subsequently identified as a new mucin, mucin 16 (MUC16) [7]. CA19-9 is a tumor marker for pancreatic cancer that resulted from the development of a monoclonal antibody [8] that was found to interact with sialyl-Lewis a (sLe[a]) antigen on pancreatic carcinoma mucins [9,10]. CA15-3 is the antigen expressed on mucin 1 (MUC1) in breast cancer that was defined by two monoclonal antibodies raised against breast cancer cells [11]. CA72-4 is a marker for epithelial-cell-derived tumors resulting from the develop-ment of murine monoclonal antibodies that recognized TAG-72, a mucin-like protein, in human serum [12]. Although these glycoproteins do not show optimal performance as screening tests for cancer, they have utility for monitoring disease progression because their serum levels correlate with tumor mass.

Some mucins can be identified and quantified in serum and plasma by proteomic analysis. For example, in the Human Proteome Organisation plasma proteome study [13], MUC16/CA125, a large protein with 22,152 amino acids and mass of 2.35 MDa before glycosylation, was well represented among the identified proteins. Its plasma concentration by immunoassay was only 67 pg/ml, yet as many as 15 MUC16 peptides were observed in some analyses using glycoprotein enrichment. Even without glycoprotein enrichment but with some fractionation before MS analysis of individual frac-tions, we have identified a large number of mucins in plasma studies (Table 1). A snapshot of the peptide representation for MUC16 in these data is shown in Figure 1. Despite the substantial peptide identifications derived from MUC16, the lack of identified peptides (red rectangles) overlapping glycosylation sites (black bars) demonstrates that domains containing glycan modifications fail to be identified in a traditional shotgun proteomic analysis.

Some effort in proteomics has focused on glycoproteins because of their biological significance and importance as a source of biomarkers. Much of the effort has used enrich-ment of glycoproteins followed by identification through MS analysis of non-glycosylated or deglycosylated peptides, without elucidating any disease-associated glycan modifica-tions. Glycomic approaches have mainly been focused on the analysis of individual targeted glycoproteins obtained through a preparative process, followed by detailed analysis of their glycan composition and structure, without making full use of the capabilities of proteomics. A more glycan-informative approach is outlined by the flow chart in Figure 2, which reflects the integration of traditional proteomics methods with glycomics techniques. In this methodology, glycoproteins can be enriched by using their affinity for lectins, which bind to specific glycans, followed by MS analysis. Glycoproteins can then be classified on the basis of their lectin-binding properties. Furthermore, im-proved glycoproteomic technologies and methodologies enable detailed glycan structure analysis and sequencing of glycopeptide backbones. Glycopeptides can be enriched and/or fractionated by lectin affinity followed by MS analysis of glycan structures in parallel with peptide sequencing of deglycosylated peptides by MS analysis. Alternative approaches to MS for glycoprotein profiling include lectin affinity array-based analyses with which glycoproteins in complex mixtures are sub-classified according to their binding to lectin arrays, and are then

Table 1

Identification of mucin proteins from plasma proteomic analysis

IPI	Gene	Description	Length (aa)	MW	N
IPI00552461	*CD164L2*	Cd164 sialomucin-like 2	171	17,874	2
IPI00452565	*EMCN*	Isoform 1 of endomucin precursor	261	27,452	2
IPI00452566	*EMCN*	Isoform 2 of endomucin precursor	178	18,848	1
IPI00400813	*EMR4*	Isoform 2 of putative EGF-like module-containing mucin-like hormone receptor-like 4 precursor	474	52,857	1
IPI00289931	*MADCAM1*	Isoform 1 of mucosal addressin cell adhesion molecule 1 precursor	406	42,691	24
IPI00397424	*MADCAM1*	Mucosal vascular addressin cell adhesion molecule 1 isoform a precursor	382	40,155	1
IPI00016334	*MCAM*	Isoform 1 of cell surface glycoprotein muc18 precursor	646	71,607	69
IPI00445227	*MCAM*	Isoform 2 of cell surface glycoprotein muc18 precursor	527	57,605	43
IPI00013955	*MUC1*	Isoform 1 of mucin-1 precursor	1,255	122,072	9
IPI00103552	*MUC16*	Mucin-16	22,152	2353428	18
IPI00646572	*MUC16*	Uncharacterized protein muc16	14,495	1518170	61
IPI00384775	*MUC17*	Small intestinal mucin muc3 (fragment)	648	65,520	1
IPI00470673	*MUC17*	Isoform 1 of mucin-17 precursor	4,493	451,723	1
IPI00737886	*MUC19*	Similar to submaxillary apomucin	1,509	158,995	8
IPI00829796	*MUC19*	Muc19	246	23,202	1
IPI00027201	*MUC2*	Mucin-2 precursor	5,179	540,300	20
IPI00178316	*MUC4*	Isoform 12 of mucin-4 precursor	1,125	125,170	1
IPI00181286	*MUC4*	52 kDa protein	475	51,806	2
IPI00103397	*MUC5AC*	Mucin-5ac precursor (fragment)	5,030	526,608	18
IPI00386327	*MUC5AC*	Mucin-5ac (fragment)	1,233	130,073	2
IPI00479313	*MUC5AC*	Gastric mucin (fragment)	1,373	147,788	1
IPI00744035	*MUC5AC*	591 kDa protein	5,709	591,191	3
IPI00869170	*MUC5AC*	Mucin 5ac, oligomeric mucus/gel-forming	6,207	648,803	1
IPI00012165	*MUC5B*	Predicted: mucin 5, subtype b, tracheobronchial	5,708	591,045	15
IPI00747373	*MUC5B*	Mucin-5b precursor	5,703	590,499	1
IPI00401776	*MUC6*	Mucin 6, gastric	2,439	257,051	23
IPI00030366	*MUC8*	Mucin	274	29,038	51
IPI00180687	*MUPCDH*	Isoform 3 of mucin and cadherin-like protein precursor	487	53,233	3
IPI00301869	*MUPCDH*	Isoform 2 of mucin and cadherin-like protein precursor	651	69,541	1
IPI00029147		Mucin (fragment)	171	17,863	1
IPI00376207		Similar to muc19	598	61,206	1
IPI00398026		Predicted: similar to mucin-4	270	27,866	1
IPI00455879		Predicted: similar to mucin 1 precursor (muc-1)	879	93,058	8
IPI00457080		Predicted: similar to submaxillary apomucin	7,202	694,615	66

Abbreviations: aa, amino acids; IPI, International Protein Index [55]; MW, molecular weight; N, number of experiments. Descriptions are taken from annotations in the IPI database.

individually quantitatively analyzed using antibodies against particular proteins.

Glycan profiling in cancer

Efforts to elucidate the spectrum of glycan alterations associated with particular tumor types have accelerated because of methodological improvements, as exemplified by two breast cancer studies that include the analysis of glycans from tumor tissue and cells [14] and the analysis of serum and other biological fluids [15]. An approach using tumor tissue or cells takes advantage of knowledge that has been derived from studies at multiple levels. Abbott *et al.* [16] used a glycosyl-transferase-directed approach that used the
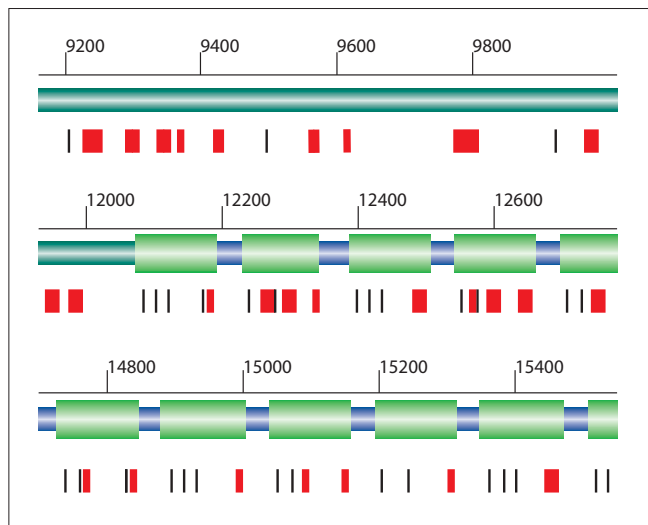
Figure 1
Distribution of mucin 16 peptides identified by mass spectrometry along part of the sequence of this large protein. Red rectangles, peptides identified; black bars, potential sites of glycosylation; green sequence, repeat domains that are heavily glycosylated. Note the lack of identified peptides spanning glycosylation sites.
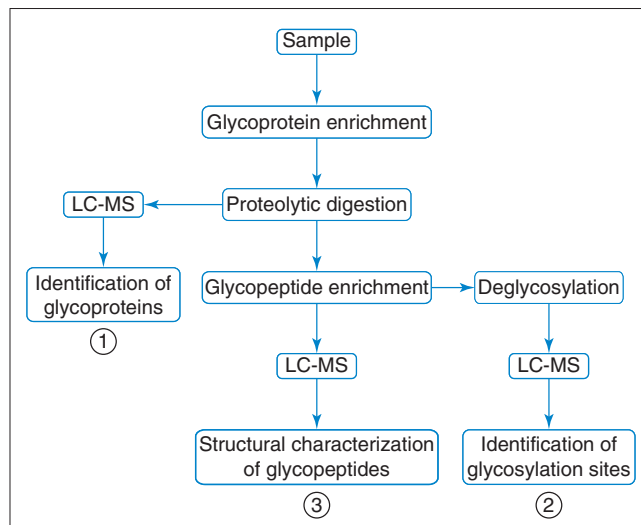

Figure 2
Glycoproteomic flow chart reflecting the integration of traditional proteomics methods with glycomic techniques. This addresses the need to identify a glycoprotein (1), identify a glycosylation site (2), and determine the structure of the glycan (3). The ability to successfully characterize glycoproteins requires an emphasis on purification, enrichment, and fractionation strategies at both the protein and peptide level.

observation that *N*-acetylglucosaminyltransferase Va (GnT-Va) transcript levels and activity are increased in breast cancer. Elevated GnT-Va levels lead to increased β(1,6)-branched *N*-linked glycan structures on glycoproteins that can be measured using *Phaseolus vulgaris* leucoagglutinin lectin (PHA-L). During the progression to invasive carcinoma, cells show a progressive increase in PHA-L binding. A PHA-L-affinity enrichment procedure, followed by nanospray ionization tandem MS (NSI-MS/MS), was used to identify a set of proteins enriched by the PHA-L-affinity fractionation in tumor relative to normal tissue. These proteins could have potential as breast cancer biomarkers [16].

Studies of serum and other biological fluids have followed a variety of approaches. In a study by Adb Hamid *et al.* [17], *N*-glycans of the total serum glycoproteins from breast cancer patients and healthy controls were characterized by high performance liquid chromatography (HPLC) with fluorescence detection coupled with exoglycosidase digestions and MS. A significant increase in a tri-sialylated tri-antennary glycan-containing α1,3-linked fucose, which forms part of the sialyl Lewis x (sLe^x) epitope, was observed in cancer samples, and this led to the isolation of a mono-galactosylated tri-antennary structure containing α1,3-linked fucose. (Glycan structures with several branches are referred to as bi-antennary, tri-antennary and so on.) This glycan was found to occur at higher levels in serum from breast cancer patients than in controls, with evidence of a positive correlation between the glycan marker and disease progression. The marker showed a stronger correlation with metastasis than did CA15-3 or carcinoembryonic antigen.

The proteins α1-acid glycoprotein, α1-antichymotrypsin and haptoglobin β-chain were identified as contributors to the increase in the glycan marker.

Matrix assisted laser desorption ionization (MALDI)-MS has been applied to glycomic profile analyses of biological fluids. MALDI-MS analysis of permethylated glycans in sera from breast cancer patients and disease-free controls identified several sialylated and fucosylated *N*-glycan structures as potential biomarkers [18]. Increases in sialylation and fucosylation of glycan structures were associated with tumor progression. The findings led to the suggestion that MS-based *N*-glycomic profiling of serum-derived constituents holds promise for staging cancer progression.

In another study by Goldman *et al.* [19], enzymatically released *N*-glycans from serum glycoproteins of hepato-cellular carcinoma (HCC), normal, and chronic liver disease were profiled by MALDI-time of flight (TOF)-MS for quantitative comparison of 83 *N*-glycans. The abundance of 57 *N*-glycans was significantly altered in HCC patients compared with controls. The combination of three selected glycans could classify HCC with 90% sensitivity and 89% specificity in an independent validation set.

Yet another glycan-centric approach applied to cancer used immunoaffinity chromatography (IAC) to isolate and identify potential cancer biomarker glycoproteins through their disease-associated glycans [15]. Glycoproteins were selected from plasma of disease-free and breast cancer patients with

an anti-Lewis x (Le^x) IAC column. The selected proteins were eluted with an acidic mobile phase and identified following tryptic digestion, reversed-phase chromatographic fractionation of the digest, and identification of peptides in collected reversed-phase liquid chromatography (LC) fractions by MALDI-MS/MS. Nine glycoproteins were found to be potential breast cancer marker candidates because of their increased levels in breast cancer patients.

Glycoprotein capture from biological fluids for proteomic and glycan analysis

In the past decade, with the evolution of proteomics and the availability of new analytical separation tools and advanced MS instrumentation, we have observed a renaissance in the use of lectins. Given the urgent need to discover cancer biomarkers in readily available clinical samples, much of the scientific interest and thus effort by investigators using lectins has been their application to complex biological samples, such as plasma and serum. Several laboratories have reported on improved analytical methods using lectin affinity chromatography (LAC) [20-24]. Given the complexity and wide range of protein concentrations in the plasma/serum proteome, most glycoproteomic workflows have incorporated one of the several available methods for depletion of abundant proteins as the first step in the fractionation of serum or plasma. Lectins are being used singly, in serial combinations or as mixtures of lectins [25-28].

Zhao *et al.* [29] used three separate lectin-agarose conjugates consisting of wheat germ agglutinin (WGA), *Sambucus nigra* lectin (SNA) and *Maackia amurensis* lectin (MAL) as part of sample preparation to selectively enrich sialic-acid-containing glycoproteins from human cancer serum. The eluate from each lectin column was further separated using other analytical separation methods (reversed-phase HPLC, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and glycopeptide mapping). Proteins that showed profile differences were first identified by peptide sequencing using LC-MS/MS. Glycopeptide mapping from the target proteins was performed in parallel using LC-electrospray ionization (ESI)-TOF-MS to determine changes in the carbohydrate structure and to identify altered glycosylation sites in the cancer serum samples. In this work [29], a change in glycosylation site occupancy, such as at Asn83, was observed for α-1 anti-trypsin in pancreatic cancer compared with the control sample. Thus, the strategy of combining targeted glycoprotein enrichment with MS analysis enabled an assessment of changes both in the abundance of the glycoprotein and in the glycosylation state of the protein.

Serial lectin affinity chromatography

Other lectin chromatography approaches, such as serial lectin affinity chromatography (SLAC), have relied on multiple lectins so that different subsets of glycoproteins can be targeted simultaneously to compare and detect changes in glycosylation patterns or to elucidate structural differences between glycoforms. SLAC involves a serial arrangement of lectin columns, whereby the flow-through of the first column is applied to the second column, and the flow-through from the second column applied to a third, and so forth. The use of SLAC for the enrichment of glycoproteins from complex biological samples requires a careful evaluation and optimization of the order of lectins, as well as the tailoring of lectin types to be used in series, in order to minimize overlap in binding specificity.

The application of SLAC is best demonstrated by the work of Sumi *et al.* [30], who proposed that *N*-acetylglucosaminyl-transferase IV is upregulated in patients with early stage prostate cancer compared with those with benign prostatic hyperplasia. This was determined by monitoring the elution of prostate-specific antigen from a SLAC system consisting of Concanavalin A (ConA), WGA, *Pisum sativum* PS, *Phaseolus vulgaris* erythroagglutinin, and *Phaseolus vulgaris* erythroagglutinin L4 lectin columns. Less binding of prostate-specific antigen to ConA and PS was observed in prostate cancer than in benign prostatic hyperplasia, indicating a shift from bi-antennary complex glycoforms to multi-antennary structures with branched GlcNAc oligosaccharides.

In another application, Qiu and Regnier [31] used a SNA column inline with a ConA column to enrich glycopeptides instead of glycoproteins containing sialylated complex *N*-linked glycans from human serum. Serum samples were digested with trypsin and fractionated over the SLAC system. Glycopeptide fractions were then isotopically labeled with heavy and light tags and deglycosylated with peptide-*N*-glycosidase F before analysis by LC-MS. A quantitative comparison of the bound and unbound fractions eluting from the ConA column showed that bi-antennary *N*-linked glycoforms were present in higher concentrations than tri- and tetra-antennary *N*-linked glycoforms in serum. Different degrees of *N*-linked glycan branching were also detected in the case of ferroxidase. Two glycopeptides were detected for ferroxidase but were observed in different ratios, indicating a different degree of binding of these glycoforms to ConA and variable glycan branching at different sites in the same protein.

Multi-lectin affinity chromatography

A strategy introduced by Yang and Hancock [32], designated multi-lectin affinity chromatography (M-LAC), uses physical admixtures of different immobilized lectins. This approach is based on the well established 'glycoside cluster effect', in which binding affinities can be dramatically increased by a clustering of both lectin binding sites and carbohydrate recognition units on the surface of cells or tissues [33]. Using kinetic measurements of a printed panel of lectins using elliptically polarized light (LFIRE) [25], the binding strength

of lectins to carbohydrates was found to be enhanced by using a multi-lectin strategy, suggesting that this approach can indeed lead to greater functional avidity than the use of individual lectins. For example, the binding affinity of thyroglobulin for the mixture of lectins (K_a = 2.1 × 10⁶ M⁻¹) was found to be about 6 and 15 times higher than the corresponding individual lectins, ConA (3.4 × 10⁵ M⁻¹) and WGA (1.4 × 10⁵ M⁻¹), respectively [25].

In addition to affinity enhancement effects, the incorporation of different ligands into a multi-ligand cluster can give rise to exquisite selectivity for specific glycan structures. For example, the M-LAC column that combines ConA, WGA, and Jacalin (a lectin derived from seeds of the jackfruit *Artocarpus integrifolia*) has been optimized to capture the majority of glycoproteins present in human serum or plasma in a single step. Furthermore, because of the enhancement in binding affinities it is possible to enrich for low- to medium-level glycoproteins despite the large dynamic range of concentrations in plasma and serum. In a M-LAC study of serum samples collected from breast cancer patients, it was possible to identify the cancer marker human epidermal growth factor receptor 2 (HER-2) at a level of 5-20 ng/ml, which was confirmed by enzyme-linked immunosorbent assay (ELISA) quantification [34].

Targeted analysis of individual glycoproteins

The targeted analysis of specific glycoproteins is performed to elucidate details of the glycan structures associated with those glycoproteins. Studies have been performed by purifying proteins followed by glycomic analysis of cleaved glycan structures or analysis of glycans attached to intact glycopeptides, enabling identification of glycan structure and peptide sequence. White *et al.* [35] used a glycomic approach to characterize glycans cleaved from prostate-specific antigen and prostatic acid phosphatase purified from seminal samples representative of normal controls, benign prostatic disease, and prostate cancer patients. The cleaved glycans were permethylated by chemically converting them to methylated derivatives, which improves MS analysis and stabilizes sialic acids and has, therefore, become a common method in glycomics analysis. The released *N*-linked glycan constituents of both proteins by normal phase HPLC and MALDI-TOF MS resulted in 40 putative glycoforms of prostate-specific antigen and 21 glycoforms of prostatic acid phosphatase.

A set of ongoing studies by Miyoshi and co-workers [36-39] has characterized increased levels of fucosylated haptoglobin in pancreatic cancer. They discovered elevated levels of fucosylated haptoglobin in the serum of patients with pancreatic cancer by western blot analyses using *Aleruia aurantica* lectin. Subsequent LC-ESI-MS analyses of haptoglobin purified from human serum of pancreatic cancer patients and normal controls demonstrated that total

concentrations of fucosylated glycan of haptoglobin increased in pancreatic cancer. Furthermore, tri-antennary *N*-glycans containing Leˣ-type fucose increased at the Asn211 site of haptoglobin in pancreatic cancer, and a di-fucosylated tetra-antennary *N*-glycan was observed only at this site in pancreatic cancer patients.

In another study, the glycan structure analysis by LC-MS/MS of intact glycopeptides of epidermal growth factor receptor (EGFR) from the A431 epithelial carcinoma cell line and four reference pools of plasma from healthy subjects and patients with cancer resulted in identification of a glycosylation site with a very different glycan structure in human plasma than in the A431 cell-line-derived EGFR (Figure 3) [40]. The four reference pools of plasma analyzed resulted in the identification of peptides shown in Figure 3a. Inspection of the identified peptides reveals that only peptides from the extracellular domain of EGFR are found in the plasma samples and only one peptide identification was for a known glycosylation site. A detailed glycan structure analysis of glycopeptides of EGFR from the A431 cell line resulted in identification of ten proposed *N*-linked glycan structures and glycosylation sites (Figure 3b) [41]. Glycan analysis of the plasma samples known to contain EGFR resulted in identification of a complex type glycan at Asn328, which differed from the high-mannose glycan at Asn328 in the A431 sample.

Lectin affinity assays

Lectin microarrays are also rapidly becoming a powerful tool for glycosylation profiling studies [42]. An advantage of lectin array platforms is that they allow highly parallel analysis of a minimal amount of body fluid or biopsy material and have good throughput and sensitivity. Recently, lectin-antibody sandwich arrays have been used to screen for glycosylation changes of target glycoproteins [43,44]. The use of an antibody as capture agent confers specificity, and several different fluorescently labeled lectins can be rapidly screened to detect different patterns of glycan structures. Chen *et al.* [43] found cancer-associated glycan alterations on the proteins MUC1 and carcinoembryonic antigen in the serum of pancreatic cancer patients compared with normal using a lectin-antibody sandwich microarray. In another study by Li *et al.* [44], antibodies for four potential glycoprotein markers were printed onto microarrays and hybridized against serum of 89 normal control, 35 chronic pancreatitis, 37 diabetic, and 22 pancreatic cancer samples. The captured glycoproteins were then probed with four different fluorescently labeled lectins in a sandwich array format, followed by characterization *in situ* by on-plate digestion and direct analysis using MALDI-QIT (quadrupole ion trap)-TOF-MS. They show that the response of α-1-β glycoprotein to SNA lectin was 69% higher in the cancer sample than in the non-cancer groups, resulting in the specific detection of pancreatic cancer with high sensitivity and specificity.
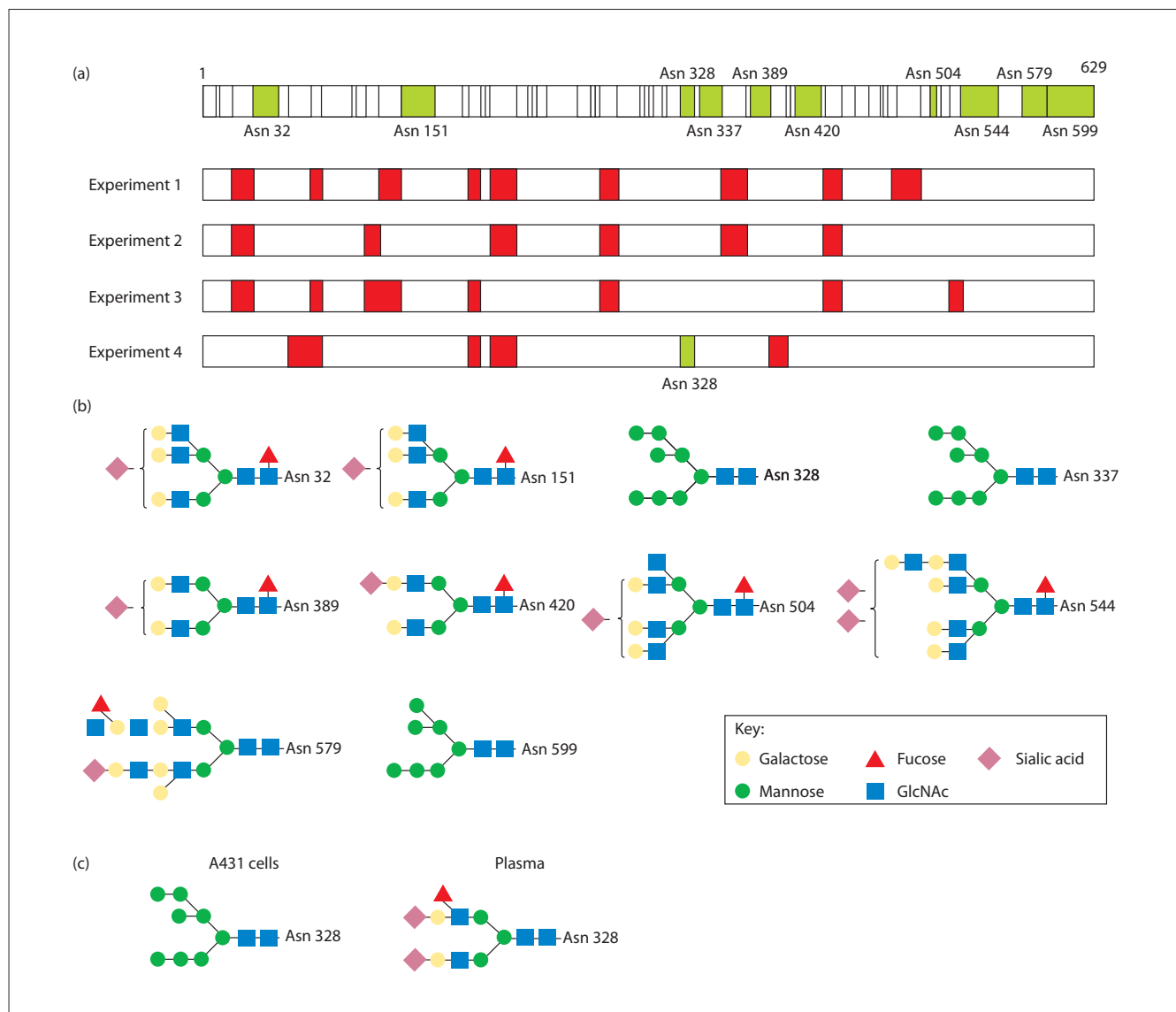
Figure 3

Identification of EGFR glycoforms. **(a)** The extracellular-domain sequence of EGFR (629 amino acids) is shown at the top. Black bars, peptides that are predicted to be generated by tryptic digestion; green boxes, peptides with known *N*-linked glycosylation sites; red boxes, peptides identified in a series of four LC-MS/MS experiments to identify proteins in reference pools of plasma from healthy subjects and patients with cancer. A manual inspection for known glycosylated peptides from EGFR was then carried out, leading to the identification of the peptide containing the Asn328 glycosylation (green box in experiment 4). **(b)** EGFR from the A431 cell line was subjected to MS analysis. Ten *N*-linked glycosylation sites were identified, and the main glycan proposed to be at each site is shown [41]. Curly brackets indicate that the monomer (sialic acid) is present in one of these positions (or two in the case of the glycan attached to Asn544). **(c)** The proposed glycan structure for one EGFR peptide differs from that of the corresponding peptide identified in A431 cells. This structure was determined following LC-MS/MS analysis of human plasma from a mixture of control individuals and patients with lung cancer (SL Wu and BL Karger, personal communication). Reproduced with permission from [40].

Other lectin array technologies that use detection methods, such as those using evanescent-field fluorescence assisted detection [45] or LFIRE [25], are emerging and offer the opportunity to detect carbohydrate-lectin interactions with very high sensitivity. A new microarray procedure using an evanescent-field fluorescence-detection principle has allowed the sensitive measurement of multiple lectin-carbohydrate interactions [46]. This method used 39 different lectins and allowed the quantitative detection of even weak lectin-carbohydrate interactions (dissociation constant, $K_d > 10^{-6}$ M). Scanning ellipsometry was used to profile glycosylation patterns from four different species with a panel of eight different lectins immobilized on a gold wafer [47].

Lectin affinity methods have potential as clinical assays for glycan biomarker quantification. Recently, an assay of the *Lens culinaris* agglutinin-reactive form of α-fetoprotein (AFP-L3%) has been approved by the US Food and Drug Administration as a diagnostic for HCC [48]. AFP is a glycoprotein with a molecular weight of about 70 kDa and has a single asparagine-linked carbohydrate chain. The level of AFP in blood is currently used in the diagnosis of HCC. However, AFP results are often not definitive until a tumor has reached a large size or an advanced state. AFP has positive predictive value ranges from 15% to 35%, with limited detection of early stage disease and many false-positive results. The study by Leerapun *et al.* [48] demonstrated the utility of AFP-L3% as a prognostic marker for HCC. They showed that for patients with indeterminate total AFP values of 10-200 ng/ml, a ratio of AFP-L3% to total AFP over 35% had 100% specificity for HCC.

Lectin affinity methods can be implemented as a targeted glycoproteomic approach for accelerated glycan biomarker discovery. The targeted glycoproteomic approach first uses LAC to enrich for a selected set of protein glycoforms and then uses lectin microarrays to screen for differences in glycosylation patterns in a set of clinical samples. Furthermore, such a lectin approach can lead to the identification of biomarker candidates solely on the basis of glycosylation differences and not necessarily of changes in protein abundance. The high-throughput capabilities of lectin arrays could accommodate hundreds of clinical samples and appropriate controls that can be rapidly screened with several lectins to validate candidate markers that had been discovered by MS-based glycomic studies.

The need for better glycoproteomic methods and platforms

To a large extent, proteomic approaches that rely on capture of glycoproteins followed by removal of their glycans as a means to allow identification of their protein constituents have not yielded unique glycoprotein forms with cancer specificity. Likewise, strategies to profile glycan constituents in cancer have resulted in the identification of abundant glycoproteins that show glycan changes related to processes that are not cancer specific, such as inflammation and acute phase response. Therefore, the approaches of stripping glycans of their proteins or stripping proteins of their glycans, followed by profiling of either glycans or protein 'backbones', are low-yield strategies for defining glycoprotein forms associated with cancer or other diseases. In addition, it is not possible, despite substantial effort and some progress in bioinformatics approaches, such as the program GlycoX [36], to predict actual glycosylation sites in a given protein, especially for *O*-glycosylation. There is a clear need for improved strategies for mining the proteome and its associated glycome to identify biomarkers and novel targets for imaging and therapeutics.

There have been substantial recent developments in sample preparation and MS methods to allow systematic analysis of glycopeptides derived from biological samples, and quantification of their glycan related isoforms, thus yielding identification of specific glycoforms associated with cancer as potential biomarkers. Such characterization approaches, when combined with lists of potentially glycosylated peptides contained in databases such as UniPep [49], enable the identification of a significant fraction of the glycoproteome of a given sample. Examples of recent developments contributing to glycoproteomics include improved analytical methods by development of specific antibodies and lectin capture agents, immobilized protease (pronase) reactors and Fourier transform MS measurements [50], isotopically labeled standard peptides [51], variable enzymatic deglycosylation strategies combined with ^{18}O incorporation [52], and new MS approaches that use electron transfer dissociation and collisionally associated dissociation fragmentation of the target glycopeptides [53]. A good example of the integration of such strategies is isotope-coded glycosylation-site-specific tagging, in which a combination of lectin-based capture methods, generation of peptide digests that are fractionated by hydrophilic interaction chromatography, peptide-*N*-glycanase-mediated incorporation of an ^{18}O stable isotope tag, and identification of ^{18}O-tagged peptides by LC-MS allowed identification of about 1,000 glycoproteins and their sites of glycosylation within a week from a crude extract of the nematode *Caenorhabditis elegans* or of mouse liver [54].

There are significant remaining challenges for the monitoring of glycan changes of low-concentration glycoproteins from clinical samples. The success of proteomics for characterization of low-abundance proteins has driven the development of the technologies and methodologies enabling substantial progress towards defining the human proteome; now there is a need to characterize glycoforms of those proteins at even lower abundance. A major limitation for detailed glycan analysis of clinical samples is the amount of sample available, even with pooling strategies. The progress of glycoproteomics will depend on the development of improved technology and methods for glycoprotein and glycopeptide enrichment, MS instrumentation capable of defining both glycan structure and peptide backbone sequence with high resolution and sensitivity, bioinformatics software for detailed glycan structure analysis, and high-throughput affinity assays for screening large sample sets. A strategy for discovery and validation of glycan alterations in cancer and other diseases will rely on MS for the discovery of glycan alterations and on high-throughput affinity approaches for validation of potential glycoforms as biomarkers through analysis of large sample sets. The eventual success of a validated clinical assay for glycan-specific biomarkers will depend on development of appropriate glycoprotein standards that will minimize variability between individual samples and laboratories.

Abbreviations

AFP, α-fetoprotein; AFP-L3%, *Lens culinaris* agglutinin-reactive form of α-fetoprotein; CA, cancer antigen; ConA, Concanavalin A; EGFR, epidermal growth factor receptor; ESI, electrospray ionization; GnT-Va, *N*-acetylglucosaminyl-transferase Va; HCC, hepatocellular carcinoma; HPLC, high performance liquid chromatography; IAC, immunoaffinity chromatography; IPI, International Protein Index; LC, liquid chromatography; Le^x, Lewis x; LFIRE, label-free internal reflection ellipsometry; MAL, *Maackia amurensis* lectin; MALDI, matrix assisted laser desorption ionization; M-LAC, multi-lectin affinity chromatography; MS, mass spectrometry; MS/MS, tandem mass spectrometry; MW, molecular weight; NSI, nanospray ionization tandem; MUC16, mucin 16; PHA-L, *Phaseolus vulgaris* leucoagglutinin; PS, *Pisum sativum*; QIT, quadrupole ion trap; SLAC, serial lectin affinity chromatography; SNA, *Sambucus nigra* lectin; TOF, time of flight; WGA, wheat germ agglutinin.

Authors' contributions

All authors contributed to the drafting of the manuscript and have seen and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

References

1. Apweiler R, Hermjakob H, Sharon N: **On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database.** *Biochim Biophys Acta* 1999, **1473:**4-8.
2. Hakomori S: **Tumor-associated carbohydrate antigens.** *Annu Rev Immunol* 1984, **2:**103-126.
3. Dube DH, Bertozzi CR: **Glycans in cancer and inflammation - potential for therapeutics and diagnostics.** *Nat Rev Drug Discov* 2005, **4:**477-488.
4. Feizi T: **Carbohydrate antigens in human cancer.** *Cancer Surv* 1985, **4:**245-269.
5. Hollingsworth MA, Swanson BJ: **Mucins in cancer: protection and control of the cell surface.** *Nat Rev Cancer* 2004, **4:**45-60.
6. Bast RC Jr, Klug TL, St John E, Jenison E, Niloff JM, Lazarus H, Berkowitz RS, Leavitt T, Griffiths CT, Parker L, Zurawski VR Jr, Knapp RC: **A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer.** *NEJM* 1983, **309:**883-887.
7. Yin BW, Lloyd KO: **Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16.** *J Biol Chem* 2001, **276:**27371-27375.
8. Koprowski H, Steplewski Z, Mitchell K, Herlyn M, Herlyn D, Fuhrer P: **Colorectal carcinoma antigens detected by hybridoma antibodies.** *Somatic Cell Genet* 1979, **5:**957-971.
9. Sawabu N, Watanabe H, Yamaguchi Y, Ohtsubo K, Motoo Y: **Serum tumor markers and molecular biological diagnosis in pancreatic cancer.** *Pancreas* 2004, **28:**263-267.
10. Magnani J, Nilsson B, Brockhaus M: **The antigen of a tumor-specific monoclonal antibody is a ganglioside containing sialylated lacto-N-fuco-pentaose II.** *Fed Proc* 1982, **41:**898.
11. Burchell JM, Mungul A, Taylor-Papadimitriou J: **O-linked glycosylation in the mammary gland: changes that occur during malignancy.** *J Mammary Gland Biol Neoplasia* 2001, **6:**355-364.
12. Colcher D, Hand PH, Nuti M, Schlom J: **A spectrum of monoclonal-antibodies reactive with human mammary-tumor cells.** *Proc Natl Acad Sci USA* 1981, **78:**3199-3203.
13. States DJ, Omenn GS, Blackwell TW, Fermin D, Eng J, Speicher DW, Hanash SM: **Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study.** *Nat Biotechnol* 2006, **24:**333-338.
14. Kirmiz C, Li B, An HJ, Clowers BH, Chew HK, Lam KS, Ferrige A, Alecio R, Borowsky AD, Sulaimon S, Lebrilla CB, Miyamoto S: **A serum glycomics approach to breast cancer biomarkers.** *Mol Cell Proteomics* 2007, **6:**43-55.
15. Cho W, Jung K, Regnier FE: **Use of glycan targeting antibodies to identify cancer-associated glycoproteins in plasma of breast cancer patients.** *Anal Chem* 2008, **80:**5286-5292.
16. Abbott KL, Aoki K, Lim JM, Porterfield M, Johnson R, O'Regan RM, Wells L, Tiemeyer M, Pierce M: **Targeted glycoproteomic identification of biomarkers for human breast carcinoma.** *J Proteome Res* 2008, **7:**1470-1480.
17. Abd Hamid UM, Royle L, Saldova R, Radcliffe CM, Harvey DJ, Storr SJ, Pardo M, Antrobus R, Chapman CJ, Zitzmann N, Robertson JF, Dwek RA, Rudd PM: **A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression.** *Glycobiology* 2008, **18:**1105-1118.
18. Kyselova Z, Mechref Y, Kang P, Goetz JA, Dobrolecki LE, Sledge GW, Schnaper L, Hickey RJ, Malkas LH, Novotny MV: **Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles.** *Clin Chem* 2008, **54:**1166-1175.
19. Goldman R, Ressom HW, Varghese RS, Goldman L, Bascug G, Loffredo CA, Abdel-Hamid M, Gouda I, Ezzat S, Kyselova Z, Mechref Y, Novotny MV: **Detection of hepatocellular carcinoma using glycomic analysis.** *Clin Cancer Res* 2009, **15:**1808-1813.
20. Jung KY, Cho WR, Regnier FE: **Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography.** *J Proteome Res* 2009, **8:**643-650.
21. Madera M, Mann B, Mechref Y, Novotny MV: **Efficacy of glycoprotein enrichment by microscale lectin affinity chromatography.** *J Sep Sci* 2008, **31:**2722-2732.
22. Kullolli M, Hancock WS, Hincapie M: **Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins.** *J Sep Sci* 2008, **31:**2733-2739.
23. Lee YC, Block G, Chen H, Folch-Puy E, Foronjy R, Jalili R, Jendresen CB, Kimura M, Kraft E, Lindemose S, Lu J, McLain T, Nutt L, Ramon-Garcia S, Smith J, Spivak A, Wang ML, Zanic M, Lin SH: **One-step isolation of plasma membrane proteins using magnetic beads with immobilized concanavalin A.** *Protein Expr Purif* 2008, **62:**223-229.
24. Qiu RQ, Regnier FE: **Use of multidimensional lectin affinity chromatography in differential glycoproteomics.** *Anal Chem* 2005, **77:**2802-2809.
25. Ralin DW, Dultz SC, Silver JE, Travis JC, Kullolli M, Hancock WS, Hincapie M: **Kinetic analysis of glycoprotein-lectin interactions by label-free internal reflection ellipsometry.** *Clin Proteomics* 2008, **1-2:**37-46.
26. Mechref Y, Madera M, Novotny MV: **Glycoprotein enrichment through lectin affinity techniques.** *Methods Mol Biol* 2008, **424:**373-396.
27. Calvano CD, Zambonin CG, Jensen ON: **Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry.** *J Proteomics* 2008, **71:**304-317.
28. Wang Y, Ao X, Vuong H, Konanur M, Miller FR, Goodison S, Lubman DM: **Membrane glycoproteins associated with breast tumor cell progression identified by a lectin affinity approach.** *J Proteome Res* 2008, **7:**4313-4325.
29. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM: **Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum.** *J Proteome Res* 2006, **5:**1792-1802.
30. Sumi S, Arai K, Kitahara S, Yoshida K: **Serial lectin affinity chromatography demonstrates altered asparagine-linked sugar-chain structures of prostate-specific antigen in human prostate carcinoma.** *J Chromatogr B Biomed Sci Appl* 1999, **727:**9-14.

31. Qiu R, Regnier FE: **Comparative glycoproteomics of N-linked complex-type glycoforms containing sialic acid in human serum.** *Anal Chem* 2005, **77:**7225-7231.
32. Yang ZP, Hancock WS: **Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column.** *J Chromatogr A* 2004, **1053:**79-88.
33. Lundquist JJ, Toone EJ: **The cluster glycoside effect.** *Chem Rev* 2002, **102:**555-578.
34. Yang Z, Harris LE, Palmer-Toy DE, Hancock WS: **Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients.** *Clin Chem* 2006, **52:**1897-1905.
35. White KY, Rodemich L, Nyalwidhe JO, Comunale MA, Clements MA, Lance RS, Schellhammer PF, Mehta AS, Semmes OJ, Drake RR: **Glycomic characterization of prostate-specific antigen and prostatic acid phosphatase in prostate cancer and benign disease seminal plasma fluids.** *J Proteome Res* 2009, **8:**620-630.
36. Nakano M, Nakagawa T, Ito T, Kitada T, Hijioka T, Kasahara A, Tajiri M, Wada Y, Taniguchi N, Miyoshi E: **Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: a novel approach for the development of tumor markers.** *Int J Cancer* 2008, **122:**2301-2309.
37. Miyoshi E, Nakano M: **Fucosylated haptoglobin is a novel marker for pancreatic cancer: Detailed analyses of oligosaccharide structures.** *Proteomics* 2008, **8:**3257-3262.
38. Narisada M, Kawamoto S, Kuwamoto K, Moriwaki K, Nakagawa T, Matsumoto H, Asahi M, Koyama N, Miyoshi E: **Identification of an inducible factor secreted by pancreatic cancer cell lines that stimulates the production of fucosylated haptoglobin in hepatoma cells.** *Biochem Biophys Res Commun* 2008, **377:**792-796.
39. Okuyama N, Ide Y, Nakano M, Nakagawa T, Yamanaka K, Moriwaki K, Murata K, Ohigashi H, Yokoyama S, Eguchi H, Ishikawa O, Ito T, Kato M, Kasahara A, Kawano S, Gu J, Taniguchi N, Miyoshi E: **Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation.** *Int J Cancer* 2006, **118:**2803-2808.
40. Hanash SM, Pitteri SJ, Faca VM: **Mining the plasma proteome for cancer biomarkers.** *Nature* 2008, **452:**571-579.
41. Wu SL, Kim J, Bandle RW, Liotta L, Petricoin E, Karger BL: **Dynamic profiling of the post-translational modifications and interaction partners of epidermal growth factor receptor signaling after stimulation by epidermal growth factor using extended range proteomic analysis (ERPA).** *Mol Cell Proteomics* 2006, **5:**1610-1627.
42. Hirabayashi J: **Concept, strategy and realization of lectin-based glycan profiling.** *J Biochem* 2008, **144:**139-147.
43. Chen SM, LaRoche T, Hamelinck D, Bergsma D, Brenner D, Simeone D, Brand RE, Haab BB: **Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays.** *Nat Methods* 2007, **4:**437-444.
44. Li C, Simeone DM, Brenner DE, Anderson MA, Shedden KA, Ruffin MT, Lubman DM: **Pancreatic cancer serum detection using a lectin/glyco-antibody array method.** *J Proteome Res* 2009, **8:**483-492.
45. Uchiyama N, Kuno A, Tateno H, Kubo Y, Mizuno M, Noguchi M, Hirabayashi J: **Optimization of evanescent-field fluorescence-assisted lectin microarray for high-sensitivity detection of monovalent oligosaccharides and glycoproteins.** *Proteomics* 2008, **8:**3042-3050.
46. Kuno A, Uchiyama N, Koseki-Kuno S, Ebe Y, Takashima S, Yamada M, Hirabayashi J: **Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling.** *Nat Methods* 2005, **2:**851-856.
47. Carlsson J, Aqvist J: **Absolute and relative entropies from computer simulation with applications to ligand binding.** *J Phys Chem B* 2005, **109:**6448-6456.
48. Leerapun A, Suravarapu SV, Bida JP, Clark RJ, Sanders EL, Mettler TA, Stadheim LM, Aderca I, Moser CD, Nagorney DM, LaRusso NF, de Groen PC, Menon KV, Lazaridis KN, Gores GJ, Charlton MR, Roberts RO, Therneau TM, Katzmann JA, Roberts LR: **The utility of *Lens culinaris* agglutinin-reactive alpha-fetoprotein in the diagnosis of hepatocellular carcinoma: evaluation in a United States referral population.** *Clin Gastroenterol Hepatol* 2007, **5:**394-402; quiz 267.
49. Zhang H, Loriaux P, Eng J, Campbell D, Keller A, Moss P, Bonneau R, Zhang N, Zhou Y, Wollscheid B, Cooke K, Yi EC, Lee H, Peskind ER, Zhang J, Smith RD, Aebersold R: **UniPep - a database for human N-linked glycosites: a resource for biomarker discovery.** *Genome Biol* 2006, **7:**R73.
50. Clowers BH, Dodds ED, Seipert RR, Lebrilla CB: **Site determination of protein glycosylation based on digestion with immobilized nonspecific proteases and Fourier transform ion cyclotron resonance mass spectrometry.** *J Proteome Res* 2007, **6:**4032-4040.
51. Hulsmeier AJ, Paesold-Burda P, Hennet T: **N-glycosylation site occupancy in serum glycoproteins using multiple reaction monitoring liquid chromatography-mass spectrometry.** *Mol Cell Proteomics* 2007, **6:**2132-2138.
52. Hagglund P, Matthiesen R, Elortza F, Hojrup P, Roepstorff P, Jensen ON, Bunkenborg J: **An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins.** *J Proteome Res* 2007, **6:**3021-3031.
53. Wu SL, Huhmer AF, Hao Z, Karger BL: **On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications.** *J Proteome Res* 2007, **6:**4230-4244.
54. Kaji H, Yamauchi Y, Takahashi N, Isobe T: **Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging.** *Nat Protoc* 2006, **1:**3019-3027.
55. Kersey PJ, Duarte J, Williams A, Karavidopoulou Y, Birney E, Apweiler R: **The International Protein Index: an integrated database for proteomics experiments.** *Proteomics* 2004, **4:**1985-1988.