

REVIEW

Open Access

# Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery

Ying Zhang<sup>1</sup>, Jing Jiao<sup>2</sup>, Pengyuan Yang<sup>1,2</sup> and Haojie Lu<sup>1,2\*</sup>

## Abstract

Glycosylation is estimated to be found in over 50% of human proteins. Aberrant protein glycosylation and alteration of glycans are closely related to many diseases. More than half of the cancer biomarkers are glycosylated-proteins, and specific glycoforms of glycosylated-proteins may serve as biomarkers for either the early detection of disease or the evaluation of therapeutic efficacy for treatment of diseases. Glycoproteomics, therefore, becomes an emerging field that can make unique contributions to the discovery of biomarkers of cancers. The recent advances in mass spectrometry (MS)-based glycoproteomics, which can analyze thousands of glycosylated-proteins in a single experiment, have shown great promise for this purpose. Herein, we described the MS-based strategies that are available for glycoproteomics, and discussed the sensitivity and high throughput in both qualitative and quantitative manners. The discovery of glycosylated-proteins as biomarkers in some representative diseases by employing glycoproteomics was also summarized.

**Keywords:** Glycosylation, Disease proteomics, Mass spectrometry

## Introduction

Proteomics, aiming at the large-scale protein analysis in biological samples, is useful for monitoring proteins that are involved in physiological changes in cells or organisms [1]. In the past two decades, proteomics has contributed tremendously to the discovery of biomarkers in a variety of diseases [2-5]. Although proteins perform a vast array of biological functions in living organisms, posttranslational modifications (PTMs) that are introduced after translation further extended the range of functions of the protein [3]. Glycosylation, one of the most common and functionally important PTMs in mammalian cells, plays a central role in many biological processes, including protein folding, host-pathogen interaction, immune response, and inflammation [4]. The most common types of protein glycosylations occur with the addition of specific glycan residues to asparagine (N-linked glycosylation). Alterations in either level or site of glycosylation can influence the cellular processes such as growth, differentiation, metastasis, and immune surveillance of tumors. Also, it is reported that modifications of the glycan structures are

associated with several diseases as well. Since glycosylated-proteins have been discovered to be the clinical biomarkers for many diseases, for example, Her2/neu in breast cancer, prostate-specific antigen (PSA) in prostate cancer, CA125 in ovarian cancer, and carcinoembryonic antigen (CEA) in colorectal, bladder, breast, pancreatic and lung cancers, glycoproteomics, as a significant branch of proteomics, has become an emerging field for biomarker discovery [5,6]. In a long time, proteomics and glycomics are two separate disciplines with the proteomics studying the whole proteome (the protein), while the glycomics focusing on the glycome (the glycan). However, neither of them can provide complete information during the progression of diseases [7]. For instance, in some cases only the concentration of glycosylated-protein, the glycosylation occupancy on a certain site, or the structure of the oligosaccharide changed, while in most of the others, all three factors may vary at the same time. Therefore, it is necessary to have an inter-discipline of glycoproteomics that can not only qualify and quantify the proteins as the traditional proteomics, but also take the following two more aspects into account: 1) protein glycosylation sites (the occupancy extent of glycosylation sites); and 2) glycosylation forms (the alterations from the attached glycan compositions and structures). Glycoproteomics, as a

\* Correspondence: luhaojie@fudan.edu.cn

<sup>1</sup>Key Laboratory of Glycoconjugates Research Ministry of Public Health and Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

<sup>2</sup>Department of Chemistry, Fudan University, Shanghai 200433, China

combination of proteomics and glycomics, studying the profile of glycosylated proteins from both protein and glycan parts, would make unique contributions to the discovery of biomarkers of cancers.

Typically, there are several goals to be achieved in the development of potential biomarkers by glycoproteomics: 1) to identify the glycosylated-proteins, their glycosylation sites, and the glycan compositions; 2) to quantify the glycosylated-proteins and their glycosylation extent; 3) to validate the potential glycosylated-protein biomarkers; and 4) to establish the clinical test method. Complete glycoproteome analysis requires multiple methods, due to the complex linkage of the glycan to the protein and the variety of glycan compositions. In recent years, considerable works have been done to characterize the sequences of the glycosylated-proteins, the primary structures of the glycan attached to proteins, and the related glycosylation site occupancy. Mass Spectrometry analysis in combination with modern separation methodologies is one of the most powerful and versatile techniques used to study

glycoproteome, because of its strong capabilities of providing both qualitative and quantitative information [8,9]. In general, N-glycoproteomics methodology consists of isolation and/or enrichment of glycosylated-proteins, proteolytic digestion, and detection of peptides for protein characterization and glycosylated-peptides for glycan composition analysis using MS (Figure 1). In N-linked glycosylated-proteins identification, the glycan are usually detached from the glycosylated-proteins/glycosylated-peptides using enzyme such as PNGase F [10]. Information of glycan composition, sequence, branching, glycosylation site, and relative quantification can thus all be obtained using MS-based methods. The conventional proteomics strategies have enabled enormous advances in the field during the past decade; however, they are still facing a bioanalytical challenge with the complex structure of glycans, which requires multiple MS-based approaches to gain different types of information.

In this review, we have summarized the current existing MS-based technologies that are utilized for characterization

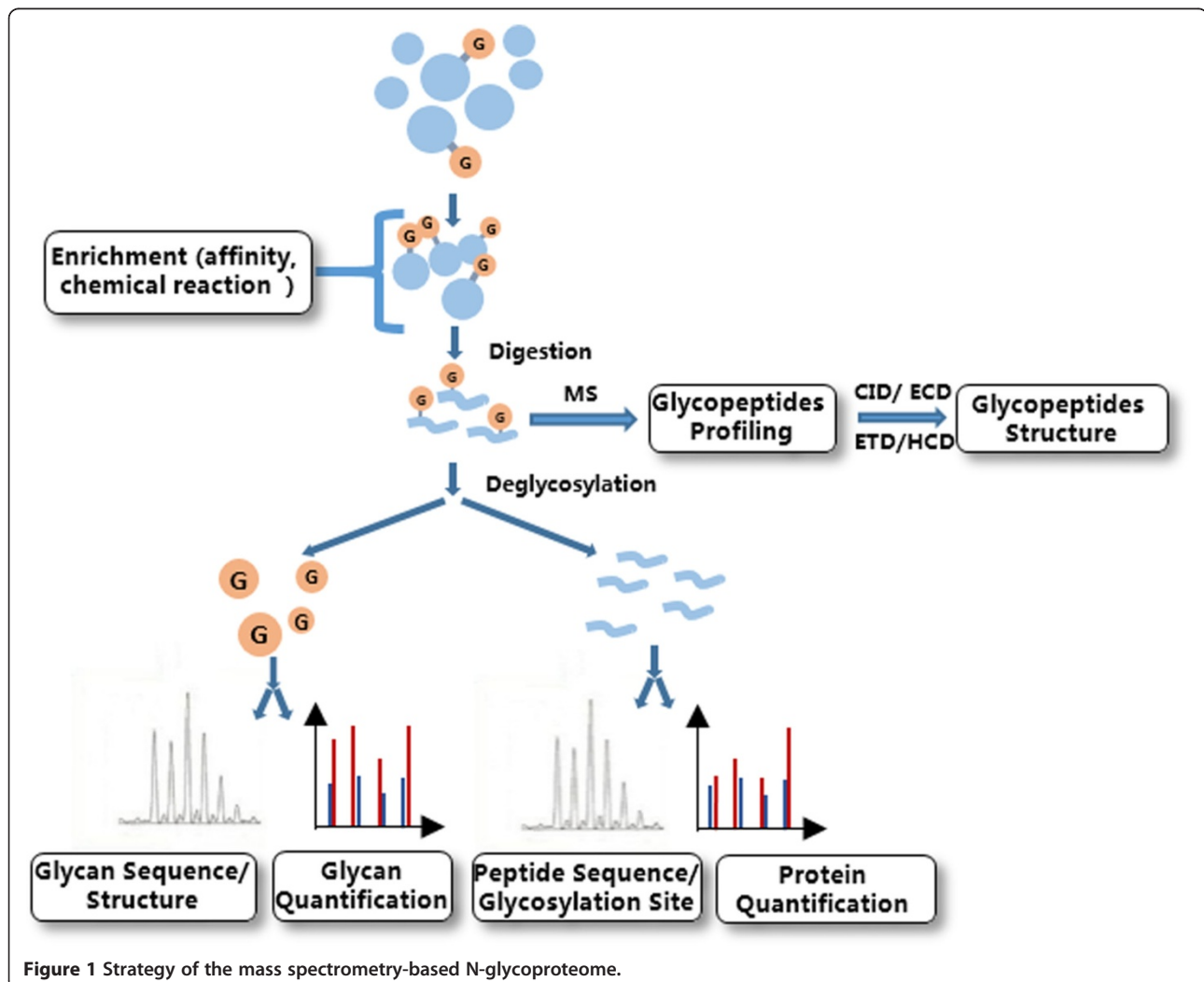


Figure 1 Strategy of the mass spectrometry-based N-glycoproteome.

and quantification of N-glycosylated-proteins, and also the understanding of glycosylated-proteins as biomarkers of some representative diseases.

### Characterization of glycosylated-proteins by mass spectrometry

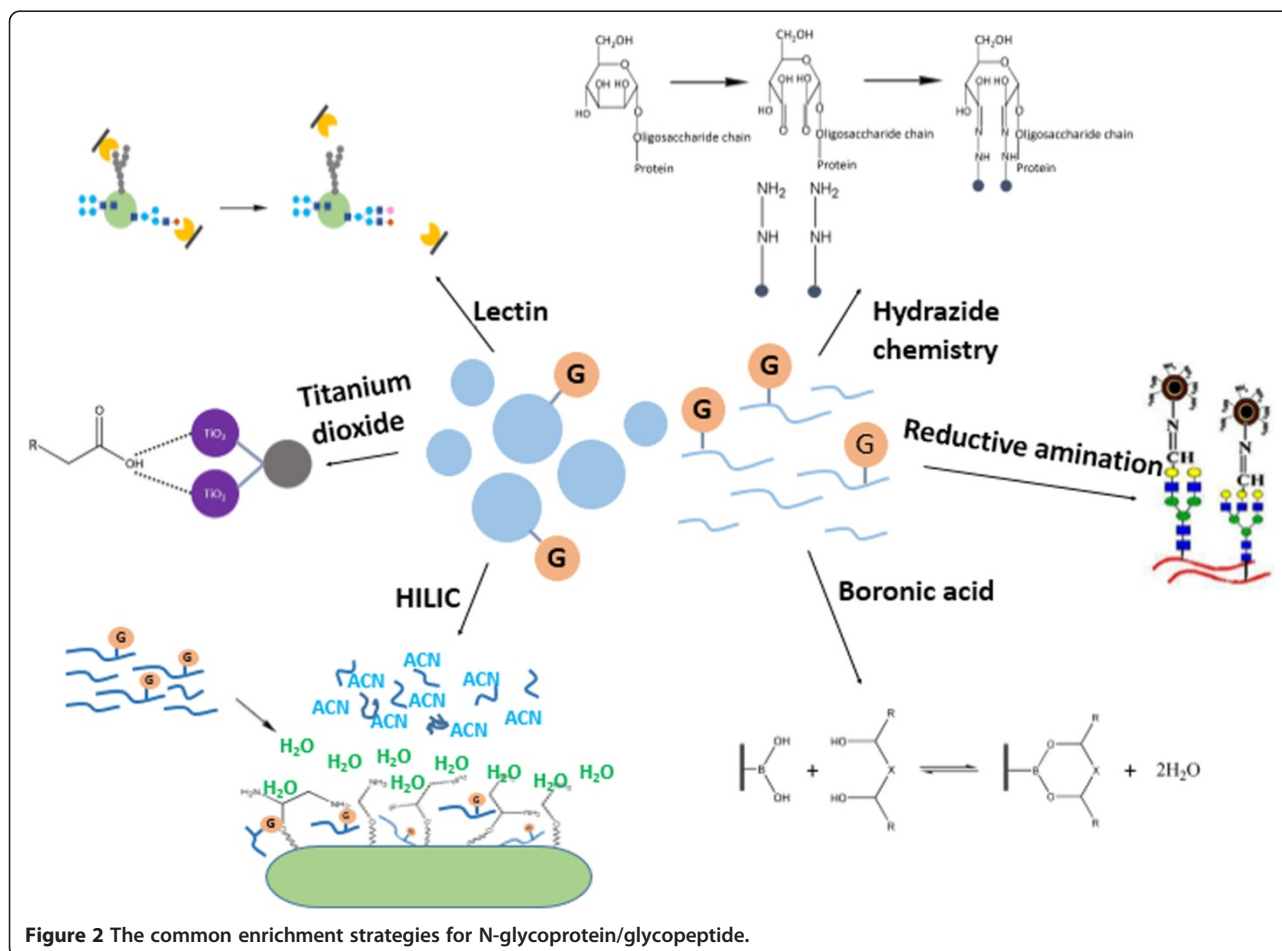
#### Glycosylated-proteins/ glycosylated-peptides enrichment

A complete characterization of glycosylated-protein includes identification, glycosylated site location, glycan structure elucidation and the functional study of the glycosylated-proteins. However, the glycoproteome is overshadowed by the complicated cellular environment, the relatively low abundance (2% to 5%) and low ionization efficiency of glycosylated-peptides. Therefore, the effective separation of glycosylated-proteins/glycosylated-peptides combined with appropriate mass spectrometry fragmentation mode would facilitate more desirable identification result [1,11]. Generally speaking, the common enrichment method can be divided into affinity approach and chemical approach. The former is based on the interaction of sugar chains and stationary phase such as lectin affinity chromatography, titanium dioxide affinity chromatography and hydrophilic interaction

liquid chromatography, while the latter realize immobilization of glycosylated-protein/glycosylated-peptide relying on the chemical reaction of glycans and specific functional group, such as hydrazide chemistry, reductive amination chemistry and boronic acid chemistry shown in Figure 2.

#### Lectin affinity chromatography

In the last decades, carbohydrate-binding proteins, known as lectins, have been widely exploited for the isolation of glycosylated-proteins or glycosylated-peptides from complex samples. Lectin has the capability to bind to distinct oligosaccharide epitopes. As early as 1970, Aspberg and Lloyd first applied lectin affinity for enrichment of glycosylated-proteins [12,13]. Since then, numerous research using lectins to enrich glycosylated-proteins/glycosylated-peptides have been developed. According to a fresh review, at least 160 lectins are known to researchers. Among them, more than 60 kinds are commercially available [14]. One kind of lectins includes concanavalin A (Con A) [15], wheat germ agglutinin (WGA) [16] and jacalin (JAC) [17], etc. They are all known to have broad specificity and can react with many different glycosylated-proteins. While, the more specific lectins, such as Sambucus



nigra lectin (SNA), which is preferentially to sialic acid attached to terminal galactose in  $\alpha$ -2,6 linkage, are often applied to adsorb a definite subset of the glycoproteome [14]. Therefore, the lectin method not only can target a broad type of glycans (e.g. Con A is often used for binding high-Man N-linked glycans) but also is appropriate for enrichment of glycans containing specific structures (e.g. L-phytohemagglutinin, or L-PHA, will recognize  $\beta$  (1,6)-branched N-linked glycans [18]) Lectin affinity chromatography based glycosylated-proteins/glycosylated-peptides enrichment strategy followed by downstream proteomic analysis has been successfully applied to variety kinds of biological samples, most frequently, in blood [19-23].

In the recent study of serum biomarker, Jung et al. took the advantage of narrow specificity lectins, which target more specific features of glycosylated-proteins to study disease associated aberrations in glycosylation [24]. They observed that approximately 30% of the plasma secretome can be captured using Con A alone, which has broad reactivity for the mannose cores typically found in N-linked glycans. Two fucose-specific lectins, Aleuria aurantia lectin (AAL) and Lens culinaris agglutinin (LCA), were found to bind 4% and 8% of the total plasma proteins, respectively. Highly specific Wisteria floribunda agglutinin (WFA)-assisted glycoproteomics approach was developed adopting dual enrichment. The combination of dissection of the WFA-stained cholangiocarcinoma tissue region and WFA-agarose column chromatography successfully identified 71 proteins as candidate markers with the assistance of subsequent analysis by mass spectrometry [23]. The limitation of enrichment approaches based on lectin is that no single lectin has the ability to bind the whole glycoproteome. Due to maximizing capture of majority of glycosylated-proteins in one step, enhanced binding strengths were often acquired by multi-lectin affinity chromatography. Kullolli et al. recently developed a High Performance Multi-Lectin Affinity Chromatography (HP-MLAC) composed of lectins ConA, JAC, and WGA. The effective integration of protein depletion, glycosylated-proteins enrichment and in-line sample concentration/desalting allowed the minimization of sample loss and the improvement of the depth of the plasma proteome analysis [25].

#### **Titanium dioxide affinity chromatography**

Titanium dioxide ( $\text{TiO}_2$ ) was often used to capture phosphopeptides from the complex sample. Larsen et al. presented a simple but highly selective strategy for enrichment of sialic acid-containing glycosylated-peptides [26]. The negatively charged sialic acid, which contains both carboxylic acid and hydroxyl group may interact with  $\text{TiO}_2$  molecules via a multipoint binding. Therefore, after dephosphorylation by alkaline phosphatase,

interaction will specifically occur between  $\text{TiO}_2$  and sialic acid. Under specific buffer conditions, nonglycosylated-peptides or neutral glycosylated-peptides were no longer bound onto the  $\text{TiO}_2$ . A total of 192 and 97 sialic acid-containing glycosylation sites from depleted human plasma and saliva were identified, respectively in this way [26]. Zhao et al. made a profound profiling of the sialylated N-glycoproteomics in human plasma by an optimized strategy of combination of a filter-aided sample preparation,  $\text{TiO}_2$  chromatography enrichment, multiple enzyme digestion, and high-pH reversed-phase peptide separation. 982 glycosylation sites in 413 proteins were identified, providing the largest database in human plasma until now [27]. Considering that  $\text{TiO}_2$  displays high specificity to enrich the phosphopeptides, it can be utilized for comprehensive studies of both types of two PTMs simultaneously in one enrichment procedure [28].

#### **Hydrophilic interaction liquid chromatography (HILIC)**

Although hydrophilic interaction liquid chromatography (HILIC) was first introduced as early as 1970s to analyze saccharides [29], recent years have seen its versatile application for the separation of biomolecules. Relative to reversed phase liquid chromatography (RPLC), HILIC is featured by the use of a hydrophilic materials as stationary phase and a hydrophobic organic solvent as mobile phase, which can be regarded as a kind of normal-phase (NP) chromatography. HILIC makes it viable for the multidimensional separation of complex samples in proteomics applications, especially analysis of post-translational modifications. Yet, NP is often performed with non-water-miscible solvent buffers but the mobile phase used by HILIC is similar to those employed in the RP-LC mode. Therefore, HILIC has specific advantages such as less complicated eluent preparation, good solubility of polar samples and greater compatibility with electrospray MS. Richness of hydroxyls makes glycosylated-peptides more hydrophilic, which means that HILIC is suitable for glycosylated-peptide separation. Analytes and polar stationary phase are interacted with each other in the form of hydrogen bonding. However, the deep mechanism of HILIC is different because of various stationary phases.

Classical HILIC stationary phases were made of bare silica or silica gels modified with polar functional groups. For example, polyhydroxyethyl aspartamide column was used to help the enrichment of glycosylated-peptides from depleted plasma prior to MS analysis. Interestingly, the author demonstrated that TFA as an ion-pairing reagent could improve enrichment efficiency [30]. The idea of employing carbohydrate materials as modification of HILIC was based on hydrophilic binding principle of carbohydrate matrices such as cellulose and sepharose to oligosaccharides [31]. IgG isotype-specific Fc glycosylation profiles were characterized relying on HILIC-SPE with



cellulose and sepharose stationary phases. The author pointed out sepharose-SPE displayed allowed a faster and more constant flow due to more homogeneous particle sizes [32]. Cotton wool, recently, was creatively introduced into HILIC SPE microtips to remove salts, most non-glycosylated peptides, and detergents. Purification of N-glycans after PNGase F deglycosylation treatment of IgG and transferrin was also obtained by cotton HILIC microtips [33].

Huang et al. put forward a new material bonded on HILIC silica with the help of diazotransfer reagent to transfer the partial amino groups in chitooligosaccharides to azido groups [34]. Compared to Sepharose, click-maltose on silica beads is more flexible, producing adequate hydroxyl groups to capture glycosylated-peptides and improving selectivity [35]. An automatical sample pretreatment workflow, composed of a click maltose-HILIC column for glycosylated-peptide enrichment, a strong cation exchange (SCX) precolumn for sample buffer exchange, and a PNGase F immobilized enzymatic reactor (IMER) for online deglycosylation was established, allowing improved sensitivity (detection limit of 5 fmol glycosylated-peptide) and selectivity (interfering 50 times (mass ratio) of BSA) [36].

The application of HILIC with zwitterionic materials is developing rapidly in the glycoproteomics. A charged liquid layer is generated near the stationary phase surface, creating a region of high ionic strength for polar interactions with glycan. The electrostatic interactions between the two oppositely charged groups is almost at a stoichiometric ratio, which is relatively weaker when compared to normal ionic exchangers such as SCX [37]. A two-dimensional chromatographic protocol with a combination of off-line ZIC-HILIC and RP chromatography including column packing, LC setup and gradient optimization was described. The strategy made it possible to improve sensitivity and handle small sample volumes, compatible with the stable isotopic label methods for quantitation used in glycoproteomics [38]. However, it is worthwhile to note that the selectivity of HILIC is relatively low as the coseparation of hydrophilic non-glycosylated-peptides.

#### **Hydrazide chemistry**

A solid-phase extraction method to enrich glycosylated-proteins based on hydrazide chemistry was first developed by Zhang et al. [39]. The methodology includes four steps: 1) Periodate oxidation of carbohydrate cis-diol groups to aldehydes. 2) Hydrazone formation between aldehydes and hydrazide groups on the support medium. 3) Tryptic digestion of the immobilized proteins in order to remove the nonglycosylated peptides. 4) Liberation of glycosylated-peptides from the support with PNGase F. Attributed to the good specificity and

high sensitivity of hydrazide chemistry-based enrichment approach, many kinds of complex biological samples have been investigated under this method. However, the glycan moiety was destroyed so the analysis of the structure of glycan was difficult. In order to make this isolation of glycosylated-protein/glycosylated-peptide easier, magnetic particles modified with hydrazide group may be synthesized [40-42]. Due to the enormous complexity of real biological samples, hydrazide chemistry combined with other pretreatment protocols will further increase the sensitivity and coverage of detection. For example, some pretreatments such as the depletion of high-abundant protein or separations by two-dimensional chromatographic will increase the sensitivity of the method [41]. Furthermore, increased glycosylated-protein sequence coverage was proved by a two-step protease digestion (Lys-C digestion, hydrazide resin separation, trypsin digestion and PNGase F deglycosylation) [43]. Multiple enzyme digestion (trypsin, pepsin and thermolysin) significantly improved the coverage of N-glycosites [44]. In addition, Parker et al. adopted parallel enrichments strategy to reveal a deeper analysis of glycoproteome, with a total of 1556 nonredundant N-linked glycosylation sites from 972 protein groups identifications [45]. Compared with lectin method, hydrazide chemistry-based enrichment turned out to be more specific due to its covalent bonding to the carbohydrates.

#### **Reductive amination chemistry**

Recently, our group improved the traditional hydrazide chemistry by introducing reductive amination to enrich N-Linked glycosylated-peptides for the first time. The enrichment strategy is based on the reaction between aldehyde groups which were produced by oxidation of the cis-diol groups of glycans and amino groups modified on the magnetic nanoparticles. Compared to the hydrazide chemistry-based solid phase extraction, this protocol eliminated the desalting step and shortened the extraction time to 4 h, improving the detection limit of glycosylated-peptides by 2 orders of magnitude. A total of 111 N-glycosylation sites were identified in 108 glycosylated-peptides in only 5  $\mu$ L of human serum [46].

#### **Boronic acid chemistry**

Boronate esters are selectively formed under alkaline condition and the reversed-reaction can occur under acidic conditions [47]. Therefore, boronic acid chemistry can be utilized for the capture and release of cis-diol-containing molecules such as glycans. The reversible, covalent bonds between boronic group and vicinal diols of glycans endow boronic acid "lectin-like" properties. However, the recognition between the glycosylated-peptide and boronic acid does not need a complicated discrimination motif, which means that boronic acid have a more broad binding ability to glycans than lectin

alone. Besides, an easy switch of the pH will control the capture/liberation of glycoconjugates, which make boronic acid-based enrichment for glycosylated-proteins/glycosylated-peptides attractive.

Our group took advantage of FDU-12 which showed good selectivity towards glycosylated-peptides to dramatically enhance the LOD of glycosylated-peptides by close to 2 orders of magnitude [48]. We further made use of mesoporous silica MCM-41 to be functionalized with boronic acid groups. The boronic acid-functionalized MCM-41 silica was demonstrated to have excellent selectivity (glycosylated-peptides/nonglycosylated-peptides at molar ratio of 1:100), good binding capacity ( $40 \text{ mg g}^{-1}$ ) and high recovery of glycosylated-peptides (88.10%) and was applied to reveal rat serum glycopeptidome for the first time, identifying 15 unique glycosylation sites originating from 15 different endogenous glycosylated-peptides [49].

### Mass spectrometry analysis

Mass spectrometry is by far the most powerful analytical technique which allow the qualitative and quantitative analysis of glycosylated-proteins. Among the “modern” ionization methods, two methodologically distinct approaches to investigate the structures of glycoconjugates in complex biological sample with increasing effectiveness and measurement sensitivity are Matrix-Assisted Laser Desorption/Ionization (MALDI) and Electrospray Ionization (ESI). The heterogeneous characteristic of glycosylated-peptide complicates the corresponding MS analysis and therefore promotes different ionization experimental conditions for different samples. For example, neutral and basic glycosylated-peptides may be analyzed under positively charged ions mode, while acidic glycoconjugates (sialic acid and sulfate containing) tend to be detected more effectively as negatively charged [50].

A complete characterization of the structure of glycosylated-peptides should include not only determination of the peptide’s amino acid sequence, but also the confirmation the glycans and the site of glycosylation. The structure elucidation mainly relies on tandem MS analysis, through which important piece of data from the overall the glycosylated-peptide may be obtained. Therefore, an appropriate fragmentation method or a combination of multiple techniques is really necessary for a comprehensive analysis, which include both the information of glycan structure and glycosylation site. Collision-induced dissociation (CID) is one of the most widely applied fragmentation mode in which molecules are usually bombarded with an inert buffer gas to increase their internal vibrational energy and fragmentation occurs when enough energy is applied on the molecule. Therefore, in most cases, CID is not suitable for structural characterization because the obtained tandem mass spectrum is dominated by the glycans cleavages fragments. In order to

generate maximum oxonium ion abundance and minimum peptide backbone fragmentation, Roepstorff et al. optimized the collision energies and realize more sensitive and selective glycopeptides identification [51].

A complementary fragmentation mode was electron-based methods, including electron-capture dissociation (ECD) and electron-transfer dissociation (ETD). Compared to CID, electron-based fragmentation pathways are prior to break the peptide bonds than glycans bonds, therefore more suitable for providing the reliable structural information about the glycosylation site because the modifications still remain onto the peptides. In ECD, low-energy electrons ( $<0.2 \text{ eV}$ ) were applied on multiply protonated peptide lead to the cleavage of N–C $\alpha$  bonds to generate a series of c and z. fragments ions [52]. Similar to ECD, in ETD, the electron delivery is achieved through the singly charged anions (e.g., fluoranthene anions) to the multiply protonated peptides. The reactions can occur in ion traps, which are less expensive than the Fourier transform ion cyclotron resonance (FT-ICR) mass, in which ECD is performed. Both ECD and ETD results preferentially in c and z fragments ions. Therefore, ECD and ETD are two excellent tools for post translational modifications such as glycosylation and phosphorylation because the labile modifications still remain onto the peptide and are largely unaffected by the fragmentation process [53]. When used in combination with CID, the glycosylated-peptides may be characterized to the fullest. Four model glycosylated-proteins were used to evaluate the power of ETD in conjunction with CID in the characterization of tryptic glycosylated-peptides. Results showed that ETD have the ability to provide reliable structural information about the glycosylation site and when it was used in tandem with CID, and both the glycan structure and the amino acid sequence of the glycosylated-peptide could be deduced [54]. A systematic approach including enrichment, deglycosylation, mass spectrometric analysis with CID and ETD was established for the study of serum of hepatocellular carcinoma (HCC). By the comparison of matched spectra of glycopeptides with and without fucosyl residue, ETD can help to localize core fucosylation sites. The combination of the CID and ETD increased the coverage and confidence of fucosylation identification [55].

By coupling an linear ion trap (LTQ) to an orbitrap analyzer, the new type of mass spectrometer boasts significantly higher resolution much more improved mass accuracies than ordinary ion-trap [56]. Olsen et al. reasoned that the C-trap, which is originally used to store ions could be used as a collision chamber for fragmentation at higher energies [57]. This higher-energy collisional dissociation (HCD) was first used to characterize protein glycosylation by Segu et al. [58]. Distinct Y1 ions (peptide + GlcNAc) generated from HCD mode allowed

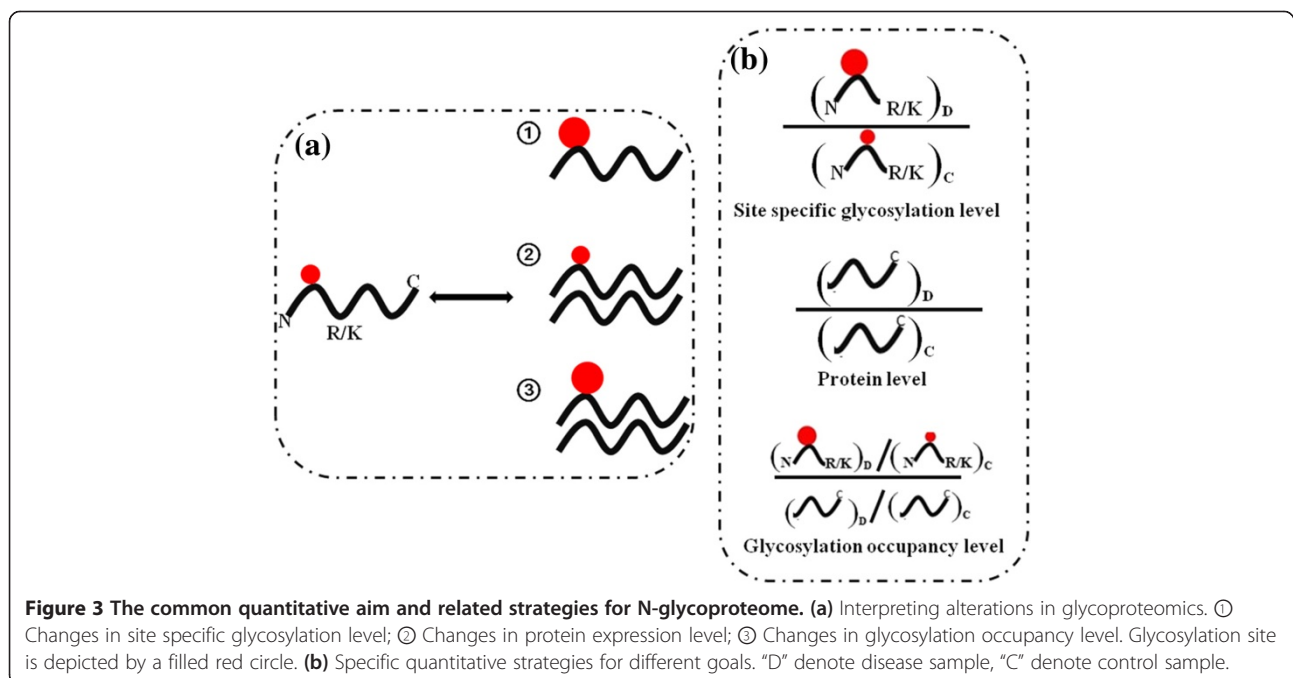
location of N-glycosylation sites of glycosylated-proteins. Furthermore, fragment ions produced by HCD were detected in the orbitrap, overcoming 1/3 cut-off limitation. Therefore, the glycan oxonium ions can also be detected. Scott et al. enriched *Campylobacter jejuni* glycosylated-peptides by using ZIC-HILIC followed by CID/HCD as well as CID/ETD. Based on their respective advantages, CID/HCD allowed the identification of glycan structure and peptide amino acid sequence, while CID/ETD enabled the elucidation of glycosylation. 130 glycosylated-peptides and 75 glycosylation sites were identified among which CID/HCD provided the majority of the results (73 sites) [59]. Ye et al. adopted HCD as the supplemental dissociation technique for CID and ETD, which can generate complementary fragments for structural elucidation of the glycosylated-peptides. Five glycosylated-peptides were analyzed with 23 glycoforms identification and 100% protein sequence coverage [60]. In a word, despite the dissociation behaviors of glycosylated-peptide ions are complex, the diversity of dissociation pathways can be regarded as an opportunity to help us fully understand glycosylated-peptide fragmentation processes and lay foundation of glycosylated-peptide profiling [61,62].

### Quantification

The high-throughput and accurate quantification of proteins is another essential component of proteomics strategies for studying cellular functions and processes. The progresses in quantitative proteomics have provided important insights into many biological processes and in

discovering proteins as biomarkers. A number of robust and accurate MS-based quantitative methods have been developed for quantification of protein expression levels in large scale. Besides monitoring changes in protein expression levels as the traditional quantitative proteomics, quantitative glycoproteomic analysis should also determine the both the site specific glycosylation level and the glycosylation occupancy level. As shown in Figure 3 (a), a glycosylated-protein can change in three typical manner (taking upregulated in glycosylation level and/or upregulated in protein expression level as an example), ① Changes in glycosylation; ② Changes in protein expression level; ③ Changes in protein expression level with simultaneous changes in glycosylation level. In the whole proteome digests, the nonglycosylated-peptides come from both of the glycosylated-proteins and nonglycosylated-proteins, while the glycosylated-peptides are only from the glycosylated-proteins. Therefore, in case ①, the abundance of nonglycosylated-peptides keeps constant, while the abundance of glycosylated-peptides may change significantly; in case ②, the abundance of glycosylated-peptides remain unchanged, while the abundance of nonglycosylated-peptides may vary dramatically; in case ③, both the abundances of glycosylated-peptides and nonglycosylated-peptides change. According to the different purposes, specific quantitative strategies can be applied as shown in Figure 3 (b).

Different methods of quantitative glycoproteomics can be categorized into two groups: label-based and label-free. Label-based quantification is accomplished by calculating the ratio of mass spectrometric intensities between the



**Figure 3** The common quantitative aim and related strategies for N-glycoproteome. (a) Interpreting alterations in glycoproteomics. ① Changes in site specific glycosylation level; ② Changes in protein expression level; ③ Changes in glycosylation occupancy level. Glycosylation site is depicted by a filled red circle. (b) Specific quantitative strategies for different goals. "D" denote disease sample, "C" denote control sample.

“light” and “heavy” samples. These techniques are based on different mechanisms to introduce labeling tags on the peptides or proteins, including stable isotopic or isobaric labeling using chemical reactions (e.g. dimethylation and iTRAQ), metabolic incorporation (e.g. SILAC) and enzymatic reactions (e.g.  $^{18}\text{O}$  labeling). While label-free quantification, which does not use a compound containing stable isotope to bind to the protein, is also a widely used technique for quantifying relative changes in complex protein samples.

### Relative quantification

#### *Alteration at the site specific glycosylation level*

The determination of the site specific glycosylation level usually combines glycosylated-proteins/glycosylated-peptides enrichment strategy, since separation of the glycosylated-proteins/ glycosylated-peptides from the backgrounds can reduce the sample complexity to a large extent. Quantification can be then performed by comparing the glycosylated-peptides with or without detaching glycan in different samples.

In most cases, quantifications are performed using peptides from glycosylated-peptides after detached glycan. Due to the multiple glycoforms that can be attached to the peptide backbone, glycosylated-peptides always have a high degree of heterogeneity, which contributes to the difficulties in glycosylated-peptides identification as well as quantification. Meantime, attachment of the glycans makes the mass spectrum difficult to decipher. Therefore, detaching the glycans from glycosylated-peptides is beneficial in the following aspects for quantification. First, when a glycosylated-peptide is identified in mass spectra, it would produce many spectral peaks corresponding to those different glycoforms attached to this glycosylated-peptides; the spreading of the glycosylated-peptide ion signal among all these glycoforms would lower the abundance of the MS signal of each peak. After detached the glycans, deglycosylated-peptides with the same amino acid composition would display as one peak in the mass spectra, making it easy to be elucidated.; Second, it is known that the deglycosylation step using peptide-N-glycosidase F (PNGase F) results in a conversion of asparagine to aspartic acid in the peptide sequence, introducing a mass difference of 0.9840 Da. The fixed mass difference combined with N-glycosylation sequence motif is conducive to further confirmation of the glycosylation site.

Once the glycans are detached, different glycoforms on the specific glycosylation sites would contribute to the total abundance of deglycosylated-peptides, the quantitative results would in fact reflect the total glycosylation profile on a specific glycosylation. Zhang et al. first described a method for the quantification of glycosylation profiles using deglycosylated-peptides after detaching

glycans, for which, glycosylated-proteins were bound to a solid support using hydrazide chemistry, and labeled with isotopically light (d0, contains no deuteriums) or heavy (d4, contains four deuteriums) forms of succinic anhydride, followed by releasing the labeled deglycosylated-peptides via PNGase F. d0 and d4 labeled peptides were then sent to mass spectrometric analysis, and the relative MS intensity ratio between the two samples reflected the glycosylation profile. The robustness of the method has been demonstrated by the standard glycosylated-protein analysis, and the observed ratios were close to the expected, with the differences ranging between 0% and 29% with a mean of 8%. The authors then successfully applied this method in quantifying the glycoproteome of different samples, including cell membranes and human serum [39]. Glycosylation quantification can be completed with the combination of various enrichment methods and labeling methods. In Sun's study, a solid phase based labeling approach by integration of glycosylated-peptides enrichment and stable isotope labeling on hydrazide beads was developed, with washing, labeling, and release of the glycosylated-peptides were all performed on the hydrazide beads sequentially. This approach was proved to be accurate and had a good linearity range with 2 orders of magnitude for quantification of deglycosylated-peptides. In addition, when comparing with dimethyl labeling conventionally performed in solution, the developed approach has better enrichment recovery (10 – 330% improvement) and high detection sensitivity. For example, when using only 10  $\mu\text{g}$  of the four standard glycosylated-protein mixtures and 400  $\mu\text{g}$  of bovine serum albumin interference as the starting sample, 42% of the annotated glycosites could still be quantified., higher than the conventional method of 26% [63]. Cordwell and co-workers quantified the glycoproteome of myocardial ischemia/reperfusion injury via iTRAQ labeling, and validated these changes with dimethyl labeling. The iTRAQ approach revealed 80/437 deglycosylated-peptides with altered abundance, while dimethyl labeling confirmed 46 of these and revealed an additional 62 significant changes. In their strategy, they employed a series of different enrichment methods including hydrazide capture, titanium dioxide and HILIC with multiple proteases to increase glycoproteome coverage [45]. Moreover, to enable the accurate and large scale quantification of the glycosylated-protein, Mann and co-workers developed a method combined “filter aided sample preparation” (FASP) and super-SILAC to quantify N-glycosylated secretome during breast cancer progression and in human blood samples. The N-glycoproteome from the 11 cell lines that are representative of different stages of breast cancer development were obtained by the FASP flow and then analyzed by LC-MS/MS. Then a super-SILAC mix was developed by collecting conditioned medium from three representative different heavy stable



isotope labeled cell lines, which served as an internal spike-in standard when being mixed with the conditioned medium of the tested cell lines for all samples accurate quantification. A total of 1398 unique N-glycosylation sites were identified and quantified, and biologically relevant differences were clearly reflected by N-glycosylated secretome profiles [64].

If the enrichment strategy is performed at the protein level, nonglycosylated-peptides from the glycosylated-proteins can also be used to quantify the glycosylation profile. As glycosylations only occur on a less part of the amino acid, it is obvious that the nonglycosylated-peptides are dominated in the whole proteome digest. If all of the non glycosylated-peptides are derived from glycosylated-proteins, the nonglycosylated-peptides can be used as the substitutes of glycosylated-peptides for quantification the total glycosylation profile, which would make the quantification more accurate and reliable due to more peptides being employed. Zou et al. developed a relative quantitative method for profiling the glycoproteome of human serum from hepatocellular carcinoma patient and healthy people, using nonglycosylated-peptides derived from glycosylated-proteins instead of glycosylated-peptides. They first captured glycosylated-proteins from serum samples on a solid-phase to ensure that the nonglycosylated-proteins would not be retained for further analysis. Therefore, all of the nonglycosylated-peptides came from the glycosylated-proteins. Then, nonglycosylated-peptides were released from the solid-phase of by trypsin digestion, separated by two-dimensional liquid chromatography, and analyzed by tandem MS. Protein quantification was achieved by a label-free quantitative method, which compared the spectrum counts of identified nonglycosylated-peptides between different samples. This method was demonstrated to have almost the same specificity and sensitivity in glycosylated-proteins quantification as capture at glycosylated-peptides level. The differential abundance of proteins present were quantified with high confidence at the concentration as low as nanogram per milliliter levels [65]. For the accurate and specific quantification of the cell surface glycosylation profile, they further proposed a modified cell surface-capturing strategy, where the glycosylated-peptides were submitted to LC-MS/MS analysis directly for identification of glycosylated-proteins, and the nonglycosylated-peptides were isotopically labeled for quantification of glycosylated-proteins. 33 glycosylated-proteins were quantified with significant expression change between the two cell lines of Chang Liver and HepG2 cells [66].

While most of the quantitation methods were based on deglycosylated-peptides, Ye and co-authors reported a quantification method achieved by comparing glycosylated-peptides without detaching the glycans in a recent study, combining a comprehensive LC-MS analysis method with

tandem mass tag (TMT) labeling. Because the glycans are retained, the quantitative results are closely related to the glycoforms, it is a glycoform specific glycosylation quantification. The glycosylated-peptides were directly qualitatively and quantitatively characterized on an Orbitrap mass analyzer with data-dependent tandem mass spectrometry (MS/MS) by alternating collision-induced dissociation (CID), electron-transfer dissociation (ETD), and higher-energy collisional dissociation (HCD) scans. The CID and low-energy HCD MS/MS analysis were performed to generate fragment ions by cleavage of glycosidic bonds, and these fragments were used for peptides sequencing and glycan structure elucidation. In addition, isomeric glycosylated-peptides with a slightly different sequence of glycan moiety could also be differentiated by CID. ETD, which tended to preserve labile PTMs, allowed not only the identification of the amino acid sequence of a glycosylated-peptide but also the assignment of its glycosylation sites. HCD combined with a TriVersa NanoMate that used to collect and reinfuse target glycosylated-peptides fractions, were performed to generate reporter ions from the tandem mass tag, and thus, provided the quantitative information for glycosylated-peptides. By alternatively using three different dissociation methods, 23 glycoforms from all 5 corresponding glycosylated-peptides were identified from tryptic digests of bovine fetuin. Nonglycosylated-peptides and glycosylated-peptides were simultaneously quantified with quantification errors less than 9% for all examined peptides and the results from nonglycosylated-peptides and glycosylated-peptides were comparable with % RSD less than 10 in most cases. However, because of the complexity of real sample, this method has only been applied to the standard sample analysis [60]. Radoslav et al. reported a LC-MS-multiple reaction monitoring (MRM) workflow that using glycosylated-peptides to quantify the glycan-specific glycosylation of Haptoglobin in liver cirrhosis and hepatocellular carcinoma. To achieve quantification of the Hp-T3 glycoforms, intense oxonium ions that are easily generated in CID of glycosylated-peptides, were selected for quantification. The glycoform-specific oxonium and peptide-HexNAc fragments were used to assure specificity of detection. The quantitative analysis showed that 14 multiply fucosylated glycoforms increased significantly in the liver disease group, comparing to healthy controls with an average 5-fold increase in intensity ( $p < 0.05$ ). At the same time, 2 tri-antennary glycoforms without fucoses did not increase in the liver disease group and 2 tetra-antennary glycoforms without fucoses showed a marginal 40% increase in intensity at most. However, the complex glycosylated-peptide mixtures made it difficult to realize the specificity of these transitions, therefore, this strategy should begin with the purification of target glycosylated-protein from patient plasma [67].

### **Alteration at the protein level**

Quantitation of the alteration at the protein level is quite similar to the traditional quantitative proteomics analysis and it can be performed through label-based or label-free approaches. In these cases, enrichment strategy can be omitted because the quantification focuses on the total amount of proteins including their native states and glycosylated counterparts. Glycosylation site is only qualitative analysis by mass spectrometry to affirm the glycosylated-protein. Because the whole procedure can be found in traditional quantitative proteomics, we will not discuss in detail in this review.

### **Alteration at the glycosylation occupancy level**

Glycosylated site occupancy is defined as the percentage of the protein that a site is occupied by glycan. So when analyzing the alteration in the glycosylation occupancy level, the alteration of glycosylation extent and the alteration of protein should be both taken into consideration. Because nonglycosylated-peptides can be from either glycosylated-protein and or their glycosylated-protein counterparts, the total nonglycosylated-peptides of a protein reflect the level of this protein including its native state and glycosylation state. While to obtain the glycosylation profile, glycosylated-proteins are always first digested into peptides, among which, the glycosylated-peptides are then selectively enriched to reflect the glycosylation profile. The glycosylation site occupancy, then, can be calculated from the ratio between the glycosylated-peptides and nonglycosylated-peptides (including ones that are from non glycosylated-proteins and its glycosylated counterparts). Our group has developed a tandem  $^{18}\text{O}$  stable isotope labeling (TOSIL) method to quantify the changes in glycosylated-proteins expression and the individual N-glycosylation site occupancy. Digests of two samples for comparison were labeled in  $\text{H}_2^{18}\text{O}$  or  $\text{H}_2^{16}\text{O}$  catalyzed by trypsin to introduce a mass tag on the C-termini of peptides, and the glycans were then released with PNGase F hydrolysis in  $\text{H}_2^{18}\text{O}$  or  $\text{H}_2^{16}\text{O}$  respectively to introduce a mass tag of  $^{18}\text{O}$  or  $^{16}\text{O}$  on the glycosylation site. As a result, the relative quantities of total protein level were obtained by measuring the  $^{18}\text{O}/^{16}\text{O}$  intensity ratios of nonglycosylated-peptides, while the relative quantities of glycosylated-protein level could be obtained by measuring the intensity of deglycosylated-peptides. Therefore, N-glycosylated site occupancy could be obtained by comparing the ratio of glycosylated-proteins level and total protein level of the two samples. By employing the TOSIL approach, it yielded a good linearity for the quantitative response of the N-glycosylated site occupancy, which was within a 10-fold dynamic range with a correlation coefficient  $r^2 > 0.99$ . This strategy was then successfully applied to the analysis of serum glycoproteome between ovarian cancer patients and healthy controls, and results showed

that changes in N-glycosylation occupancy at specific sites did not always have the same trends as those of protein expression levels [68]. In addition, Shakey et al. developed a similar strategy with the same labeling methods but different enrichment method for relative quantitation of glycosylated-proteins site occupancy [42]. Ueda et al. reported an approach termed isotopic glycosidase elution and labeling on lectin column chromatography (IGEL) for profiling serum biomarkers in lung cancer patient serum. Protein digests were first labeled with 8-plex iTRAQ reagent, followed by the capturing of glycosylated-peptides with lectin-column chromatography and the tagging of the N-glycosylation sites with  $^{18}\text{O}$  by PNGase F. Nonglycosylated-peptides labeled with iTRAQ were used to quantify the glycosylated-proteins in the protein expression level, whereas deglycosylated-peptides labeled with iTRAQ were used to quantify the glycosylation degree. With the aid of a 8-plex labeling reagent, this technology allowed the comparisons of up to eight different samples within a single experiment [69].

### **Absolute quantification**

Quite often, the “hypothesis-generating” relative quantification experiments mentioned above are worthless for defining the abundance of a protein. Therefore, subsequent experiments are strongly needed to target them explicitly through absolute quantification. Targeted quantification of proteins is a daily task in biological research, but traditional immunoaffinity-based methods cannot reach desired levels of sensitivity and throughput. In conjunction with the stable isotope dilution method, MS-based selected reaction monitoring (SRM) and MRM assays have opened new avenues for absolute quantification of the target protein, with unparalleled sensitivity and high-speed [70,71]. Domon et al. designed a heavy isotope-labeled synthetic peptide with C-terminal lysine and arginine labeled with  $^{13}\text{C}$  and  $^{15}\text{N}$ , respectively. Especially, the asparagine (N) at the N-X-S/T motifs was replaced with aspartic acid (D) to mimic the deglycosylated-peptides. Then the heavily labeled peptides were used in the SRM workflow for absolute quantitative of the target glycosylated-proteins. However, this protocol can only quantify the concentration at the protein level, while the degree of glycosylation could not be distinguished [72]. In recent work, Ruedi et al. reported a large scale absolute quantitative method for N-glycoproteome by SWATH-MS. SWATH-MS is a data-independent (DIA) acquisition mass spectrometric method in which data are acquired on a fast, high resolution Q-TOF instrument. By repeatedly cycling through sequential isolation windows over the whole chromatographic elution range, SWATH generates a complete digital archive of all possible ions in a single analysis for permanent, retrospective, quantitative MS/MS analysis of complex samples. It thus can thus potentially

combine the advantages of shotgun (high throughput) with those of SRM (high reproducibility and sensitivity). The author utilized isotopically labeled heavy forms (containing either a C-terminal [ $^{13}\text{C}_6^{15}\text{N}_4$ ] Arg or [ $^{13}\text{C}_6^{15}\text{N}_2$ ] Lys residue) of deglycosylated-peptides as the “absolute quantification” (AQUA) peptides to realize the quantification. Their results showed that compared to SRM, SWATH acquisition allowed the quantification of plasma glycosylated-proteins with equal variances, similar accuracy, and dynamic range, except for a slightly higher LOQ (2–3 times less sensitive) [73]. While the above-mentioned works focused on the absolute quantification of glycosylated-proteins, Ruhaak et al. have developed a method using MRM to examine quantitative changes in glycosylation at a site-specific level. They first performed studies on tuning the MS conditions for nonglycosylated-peptides and glycosylated-peptides, using standard IgG to construct suitable transitions. Then, the absolute concentrations of proteins were obtained using nonglycosylated-peptides and absolute glycosylated-proteins concentration and site-specific quantitation of individual glycoforms were obtained using glycosylated-peptides. Both nonglycosylated-peptides and glycosylated-peptides were quantified using MRM in the same run. Consequently, this new approach provided information regarding both the absolute amount of protein and the site-specific glycosylation profile, which is useful to determine if altered glycosylation profiles in serum/plasma were due to a change in protein glycosylation or a change in protein concentration. This strategy also showed a low limit of detection of 60 attomoles, and a wide dynamic range of 3 orders magnitude for IgG protein quantitation [74].

### Diseases

N-linked glycosylated-proteins, have been shown to be increasingly important in biomarker analyses and biopharmaceutical drug development [75]. Aberrant protein glycosylation may result in abnormal changes in biological function/activity, protein folding, and molecular recognition in cancer. The site of protein glycosylation and the structure of the oligosaccharide could also be altered during initiation or progression of disease [75,76]. Cancer has emerged as the most dreadful disease in the world. Although there has been a great deal of progress in treating it in the last decades, cancer still claims hundreds of thousands of lives each year. The discovery of these cancer-associated modifications of glycans on the glycosylated-proteins may also improve on the specificity of existing cancer biomarkers. For example, elevation of serum alpha-fetoprotein (AFP), a common marker for hepatocellular carcinoma (HCC), also occurs in non-HCC conditions such as pregnancy. In contrast, AFP-L3, consisting of core-fucosylated glycoforms of AFP, provides better specificity for HCC. The most deadly forms

of cancer in the world include lung cancer, colon cancer, breast cancer, pancreatic cancer, liver cancer et al. We will discuss recent advances of glycoproteomics in these common cancers in the following parts to demonstrate the usefulness of glycoproteomics in biomarker discovery, diagnosis and subtype classification et al.

Breast cancer is the most frequently diagnosed cancer, and it is the leading cause of cancer death among females. Recent research reported by Yen et al. showed that glycosylated-proteins could be used to distinguish various subtypes of breast cancer, as well as to distinguish normal and benign breast cells from breast cancer cells. They obtained glycosylated-protein profiles, characterized these glycosylated-proteins from 14 breast cell lines, and then sent the glycoproteome to unsupervised hierarchical cluster analysis. Results showed that the differential expression of glycosylated-proteins in these breast cancer cell lines readily allows the classification of the lines into normal, benign, malignant, basal, and luminal groups [77]. Lung cancer is the leading cancer in males. Through quantification using MS-based glycoproteomic approaches, Tsai and co-authors noticed that when compared to normal donors, the fucosylated haptoglobin (Hp) level increased significantly in serum of each subtype of lung cancer. They also determined the glycan structure to be an alpha 2,6-linked tri-sialylated triantennary glycan, containing alpha 1,3-linked fucose which is attached to the four-linked position of the three-arm mannose of N-linked core pentasaccharide [78]. As hepatocellular carcinoma (HCC) is the most common primary malignant tumor of the liver, Liu et al. developed an integrated platform to discover the glycosylated-protein biomarkers in early HCC. The approach they employed is to use lectin arrays combined with LC-MS/MS-based quantitative glycoproteomics. Lectin arrays analysis showed an elevation of fucosylation level of serum glycosylated-proteins from HCC compared with cirrhosis patients. Finally, C3, CE, HRG, CD14 and HGF were found to be biomarker candidates for distinguishing early HCC from cirrhosis, with a sensitivity of 72% and specificity of 79% [79]. Among the newly discovered potential biomarkers in HCC, AFP-L3 (a glycoform of alpha-fetoprotein, AFP) is recently approved by FDA as a supplemental test in patients with elevated total AFP. Because elevation of serum AFP, a common marker for hepatocellular carcinoma (HCC), also occurs in non-HCC conditions such as pregnancy, AFP-L3, consisting of core-fucosylated glycoforms of AFP, provides better specificity for HCC. AFP-L3 levels have been found to be related to progression from moderately differentiated to poorly differentiated tumors, therefore, it may be used as an early diagnosis of HCC when the tumor diameter is <2 cm [80]. Similarly, in order to discover potential glycosylated-proteins biomarkers in ovarian cancer, Wu et al. applied a lectin array and



Exactag labeling based quantitative glycoproteomics approach. After being labeled with isobaric chemical tags, glycosylated-proteins were enriched by fucose-specific lectin array. The enriched glycosylated-proteins were then identified and quantified by LC-MS/MS, and five glycosylated-proteins were found to be differentially expressed in the serum of ovarian cancer patients, compared to benign diseases. With the further confirmation by lectin-ELISA and ELISA assay, corticosteroid-binding globulin (CBG) and other four glycosylated-proteins were identified as potential markers for differentiating ovarian cancer from benign diseases or healthy controls. Especially, the combination of CBG, SAP, and CA125 showed improved performance for distinguishing stage III ovarian cancer from benign diseases compared to CA125 alone [81]. These above findings clearly suggest that glycosylated-protein modifications may be a useful means to identify novel markers for detection of cancer and monitoring of cancer progression.

## Conclusions

Despite intense activities in the field of biomarker research, new biomarkers are still urgently needed to accelerate efforts in developing new drugs and treatments of known diseases. Though the field of glycoproteomics is moving forward, and more glycosylated-protein biomarkers have been discovered, there is still much work to do, for the future glycoproteomics research. First, since biologically significant glycosylated-proteins often exist in low abundance, enrichment of these glycosylated-proteins with high specificity is still worth attention. Second, although traditional proteomics identification strategies can provide information on novel sites of glycosylation, software and bioinformatics methods still need to be developed to improve glycan composition and structural analysis. Third, in order to realize effective biomarker screening, it is required to accurately quantify the glycosylation site occupancy from both relative and absolute aspects, as well as to analyze the relationship between the changes in glycosylation site occupancy and protein expression. With the progress of glycoproteomics research, we believe that the growth in the development of methods for high-throughput glycoproteomics will shed new lights on the discovery of new biomarkers.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

YZ and JJ drafted the manuscript. All authors edited and approved the final manuscript.

## Acknowledgement

The work was supported by the NST (Grants 2012CB910602, 2012AA020203), NSF (Grants 21025519, 21335002, and 21375026), the Ph.D. Programs Foundation of Ministry of Education of China (20130071110034) and Shanghai Projects (Grants Eastern Scholar and B109).

Received: 30 September 2013 Accepted: 20 March 2014  
Published: 5 May 2014

## References

1. Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, Bergeron J, Borchers CH, Corthals GL, Costello CE: **The human proteome project: current state and future direction.** *Mol Cell Proteomics* 2011, **10**:1–5.
2. Tan HT, Lee YH, Chung M: **Cancer proteomics.** *Mass Spectrom Rev* 2012, **31**:583–605.
3. Deribe YL, Pawson T, Dikic I: **Post-translational modifications in signal integration.** *Nat Struct Mol Biol* 2010, **17**:666–672.
4. Kolarich D, Lepenies B, Seeberger PH: **Glycomics, glycoproteomics and the immune system.** *Curr Opin Chem Biol* 2012, **16**:214–220.
5. Kim EH, Misek DE: **Glycoproteomics-based identification of cancer biomarkers.** *Int J Proteomics* 2011. doi:10.1155/2011/601937.
6. Chandler K, Goldman R: **Glycoprotein disease markers and single protein-omics.** *Mol Cell Proteomics* 2013, **12**:836–845.
7. Doerr A: **Glycoproteomics.** *Nat Methods* 2012, **9**:36–36.
8. Pan S, Chen R, Aebersold R, Brentnall TA: **Mass spectrometry based glycoproteomics—from a proteomics perspective.** *Mol Cell Proteomics* 2011, **10**:1–14.
9. Desaire H: **Glycopeptide analysis, recent developments and applications.** *Mol Cell Proteomics* 2013, **12**:893–901.
10. Lazar IM, Lee W, Lazar AC: **Glycoproteomics on the rise: established methods, advanced techniques, sophisticated biological applications.** *Electrophoresis* 2013, **34**:113–125.
11. Pasing Y, Sickmann A, Lewandrowski U: **N-glycoproteomics: mass spectrometry-based glycosylation site annotation.** *Biol Chem* 2012, **393**:249–258.
12. Aspberg K, Porath J: **Group-specific adsorption of glycoproteins.** *Acta Chem Scand* 1970, **24**:1841.
13. Lloyd KO: **The preparation of two insoluble forms of the phytohemagglutinin, concanavalin A, and their interactions with polysaccharides and glycoproteins.** *Arch Biochem Biophys* 1970, **137**:460–468.
14. Fanayan S, Hincapie M, Hancock WS: **Using lectins to harvest the plasma/serum glycoproteome.** *Electrophoresis* 2012, **33**:1746–1754.
15. Brewer CF, Bhattacharyya L: **Specificity of concanavalin A binding to asparagine-linked glycopeptides. A nuclear magnetic relaxation dispersion study.** *J Biol Chem* 1986, **261**:7306–7310.
16. Nagata Y, Burger MM: **Wheat germ agglutinin molecular characteristics and specificity for sugar binding.** *J Biol Chem* 1974, **249**:3116–3122.
17. Roque-Barreira MC, Campos-Neto A: **Jacalin: an IgA-binding lectin.** *J Immunol* 1985, **134**:1740–1743.
18. Abbott KL, Aoki K, Lim J-M, 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.
19. Mann B, Madera M, Klouckova I, Mechref Y, Dobrolecki LE, Hickey RJ, Hammoud ZT, Novotny MV: **A quantitative investigation of fucosylated serum glycoproteins with application to esophageal adenocarcinoma.** *Electrophoresis* 2010, **31**:1833–1841.
20. Ahn YH, Kim KH, Shin PM, Ji ES, Kim H, Yoo JS: **Identification of low-abundance cancer biomarker candidate TIMP1 from serum with lectin fractionation and peptide affinity enrichment by ultrahigh-resolution mass spectrometry.** *Anal Chem* 2012, **84**:1425–1431.
21. Wen C-L, Chen K-Y, Chen C-T, Chuang J-G, Yang P-C, Chow L-P: **Development of an AlphaLISA assay to quantify serum core-fucosylated E-cadherin as a metastatic lung adenocarcinoma biomarker.** *J Proteomics* 2012, **75**:3963–3976.
22. Uen Y-H, Lin K-Y, Sun D-P, Liao C-C, Hsieh M-S, Huang Y-K, Chen Y-W, Huang P-H, Chen W-J, Tai C-C: **Comparative proteomics, network analysis and post-translational modification identification reveal differential profiles of plasma Con A-bound glycoprotein biomarkers in gastric cancer.** *J Proteomics* 2013, **83**:197–213.
23. Matsuda A, Kuno A, Matsuzaki H, Kawamoto T, Shikanai T, Nakanuma Y, Yamamoto M, Ohkohchi N, Ikehara Y, Shoda J: **Glycoproteomics-based cancer marker discovery adopting dual enrichment with Wisteria floribunda agglutinin for high specific glyco-diagnosis of cholangiocarcinoma.** *J Proteomics* 2013, **85**:1–11.
24. Jung K, Cho W, Regnier FE: **Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography.** *J Proteome Res* 2008, **8**:643–650.



25. Kullolli M, Hancock WS, Hincapie M: **Automated platform for fractionation of human plasma glycoproteome in clinical proteomics.** *Anal Chem* 2009, **82**:115–120.
26. Larsen MR, Jensen SS, Jakobsen LA, Heegaard NH: **Exploring the sialome using titanium dioxide chromatography and mass spectrometry.** *Mol Cell Proteomics* 2007, **6**:1778–1787.
27. Zhao X, Ma C, Han H, Jiang J, Tian F, Wang J, Ying W, Qian X: **Comparison and optimization of strategies for a more profound profiling of the sialylated N-glycoproteomics in human plasma using metal oxide enrichment.** *Anal Bioanal Chem* 2013, **405**(16):5519–5529.
28. Palmisano G, Parker BL, Engholm-Keller K, Lendal SE, Kulej K, Schulz M, Schwämmle V, Graham ME, Saxtorph H, Cordwell SJ: **A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development.** *Mol Cell Proteomics* 2012, **11**:1191–1202.
29. Linden JC, Lawhead CL: **Liquid chromatography of saccharides.** *J Chromatogr A* 1975, **105**:125–133.
30. Myslting S, Palmisano G, Højrup P, Thaysen-Andersen M: **Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics.** *Anal Chem* 2010, **82**:5598–5609.
31. Wada Y, Tajiri M, Yoshida S: **Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics.** *Anal Chem* 2004, **76**:6560–6565.
32. Selman MH, McDonnell LA, Palmblad M, Ruhaak LR, Deelder AM, Wuhrer M: **Immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry.** *Anal Chem* 2010, **82**:1073–1081.
33. Selman MH, Hemayatkar M, Deelder AM, Wuhrer M: **Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides.** *Anal Chem* 2011, **83**:2492–2499.
34. Huang H, Jin Y, Xue M, Yu L, Fu Q, Ke Y, Chu C, Liang X: **A novel click chitooligosaccharide for hydrophilic interaction liquid chromatography.** *Chem Commun* 2009, **45**:6973–6975.
35. Yu L, Li X, Guo Z, Zhang X, Liang X: **Hydrophilic interaction chromatography based enrichment of glycopeptides by using click maltose: a matrix with high selectivity and glycosylation heterogeneity coverage.** *Chem-A Eur J* 2009, **15**:12618–12626.
36. Qu Y, Xia S, Yuan H, Wu Q, Li M, Zou L, Zhang L, Liang Z, Zhang Y: **Integrated sample pretreatment system for N-linked glycosylation site profiling with combination of hydrophilic interaction chromatography and PNGase F immobilized enzymatic reactor via a strong cation exchange precolumn.** *Anal Chem* 2011, **83**:7457–7463.
37. Di Palma S, Boersema PJ, Heck AJ, Mohammed S: **Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-C-HILIC) provide high resolution separation and increase sensitivity in proteome analysis.** *Anal Chem* 2011, **83**:3440–3447.
38. Di Palma S, Mohammed S, Heck AJ: **ZIC-C-HILIC as a fractionation method for sensitive and powerful shotgun proteomics.** *Nat Protoc* 2012, **7**:2041–2055.
39. Zhang H, Li X-j, Martin DB, Aebersold R: **Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.** *Nat Biotechnol* 2003, **21**:660–666.
40. Wang L, Aryal UK, Dai Z, Mason AC, Monroe ME, Tian Z-X, Zhou J-Y, Su D, Weitz KK, Liu T: **Mapping N-Linked Glycosylation Sites in the Secretome and Whole Cells of *Aspergillus niger* Using Hydrazide Chemistry and Mass Spectrometry.** *J Proteome Res* 2011, **11**:143–156.
41. Berven FS, Ahmad R, Clauser KR, Carr SA: **Optimizing performance of glycopeptide capture for plasma proteomics.** *J Proteome Res* 2010, **9**:1706–1715.
42. Shakey Q, Bates B, Wu J: **An approach to quantifying N-linked glycoproteins by enzyme-catalyzed 18O3-labeling of solid-phase enriched glycopeptides.** *Anal Chem* 2010, **82**:7722–7728.
43. Chen Y, Cao J, Yan G, Lu H, Yang P: **Two-step protease digestion and glycopeptide capture approach for accurate glycosite identification and glycoprotein sequence coverage improvement.** *Talanta* 2011, **85**:70–75.
44. Chen R, Jiang X, Sun D, Han G, Wang F, Ye M, Wang L, Zou H: **Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry.** *J Proteome Res* 2009, **8**:651–661.
45. Parker BL, Palmisano G, Edwards AV, White MY, Engholm-Keller K, Lee A, Scott NE, Kolarich D, Hambly BD, Packer NH: **Quantitative N-linked glycoproteomics of myocardial ischemia and reperfusion injury reveals early remodeling in the extracellular environment.** *Mol Cell Proteomics* 2011, **10**:1449–1458.
46. Zhang Y, Kuang M, Zhang L, Yang P, Lu H: **An accessible protocol for solid-phase extraction of N-linked glycopeptides through reductive amination by amine-functionalized magnetic nanoparticles.** *Anal Chem* 2013, **85**:5535–5541.
47. Rawn JD, Lienhard GE: **Binding of boronic acids to chymotrypsin.** *Biochemistry* 1974, **13**:3124–3130.
48. Xu Y, Wu Z, Zhang L, Lu H, Yang P, Webley PA, Zhao D: **Highly specific enrichment of glycopeptides using boronic acid-functionalized mesoporous silica.** *Anal Chem* 2008, **81**:503–508.
49. Liu L, Zhang Y, Zhang L, Yan G, Yao J, Yang P, Lu H: **Highly specific revelation of rat serum glycopeptidome by boronic acid-functionalized mesoporous silica.** *Anal Chim Acta* 2012, **753**:64–72.
50. Harvey DJ: **Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2007–2008.** *Mass Spectrom Rev* 2012, **31**:183–311.
51. Jebanathirajah J, Steen H, Roepstorff P: **Using optimized collision energies and high resolution, high accuracy fragment ion selection to improve glycopeptide detection by precursor ion scanning.** *J Am Soc Mass Spectrom* 2003, **14**:777–784.
52. Zubarev RA, Kelleher NL, McLafferty FW: **Electron capture dissociation of multiply charged protein cations. A nonergodic process.** *J Am Chem Soc* 1998, **120**:3265–3266.
53. Catalina MI, Koeleman CA, Deelder AM, Wuhrer M: **Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain.** *Rapid Commun Mass Spectrom* 2007, **21**:1053–1061.
54. Alley WR, Mechref Y, Novotny MV: **Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data.** *Rapid Commun Mass Spectrom* 2009, **23**:161–170.
55. Chen R, Wang F, Tan Y, Sun Z, Song C, Ye M, Wang H, Zou H: **Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation.** *J Proteome* 2012, **75**:1666–1674.
56. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R: **The Orbitrap: a new mass spectrometer.** *J Mass Spectrom* 2005, **40**:430–443.
57. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M: **Higher-energy C-trap dissociation for peptide modification analysis.** *Nat Methods* 2007, **4**:709–712.
58. Segu ZM, Mechref Y: **Characterizing protein glycosylation sites through higher-energy C-trap dissociation.** *Rapid Commun Mass Spectrom* 2010, **24**:1217–1225.
59. Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AV, Crossett B, Falconer L, Kolarich D, Djordjevic SP, Højrup P: **Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*.** *Mol Cell Proteomics* 2011, **10**:1–18.
60. Ye H, Boyne MT, Buhs LF, Hill J: **Direct approach for qualitative and quantitative characterization of glycoproteins using tandem mass tags and an LTQ orbitrap XL electron transfer dissociation hybrid mass spectrometer.** *Anal Chem* 2013, **85**:1531–1539.
61. Dodds ED: **Gas-phase dissociation of glycosylated peptide ions.** *Mass Spectrom Rev* 2012, **31**:666–682.
62. Cooper CA, Gastegger E, Packer NH: **GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data.** *Proteomics* 2001, **1**:340–349.
63. Sun Z, Qin H, Wang F, Cheng K, Dong M, Ye M, Zou H: **Capture and dimethyl labeling of glycopeptides on hydrazide beads for quantitative glycoproteomics analysis.** *Anal Chem* 2012, **84**:8452–8456.
64. Boersema PJ, Geiger T, Wiśniewski JR, Mann M: **Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples.** *Mol Cell Proteomics* 2013, **12**:158–171.
65. Chen R, Tan Y, Wang M, Wang F, Yao Z, Dong L, Ye M, Wang H, Zou H: **Development of glycoprotein capture-based label-free method for the high-throughput screening of differential glycoproteins in hepatocellular carcinoma.** *Mol Cell Proteomics* 2011, **10**. doi:10.1074/mcp.M110.006445.

66. Sun Z, Chen R, Cheng K, Liu H, Qin H, Ye M, Zou H: **A new method for quantitative analysis of cell surface glycoproteome.** *Proteomics* 2012, **12**:3328–3337.
67. Sanda M, Pompach P, Brnakova Z, Wu J, Makambi K, Goldman R: **Quantitative liquid chromatography-mass spectrometry-multiple reaction monitoring (LC-MS-MRM) analysis of site-specific glycoforms of haptoglobin in liver disease.** *Mol Cell Proteomics* 2013, **12**:1294–1305.
68. Liu Z, Cao J, He Y, Qiao L, Xu C, Lu H, Yang P: **Tandem 18O stable isotope labeling for quantification of N-glycoproteome.** *J Proteome Res* 2009, **9**:227–236.
69. Ueda K, Takami S, Saichi N, Daigo Y, Ishikawa N, Kohno N, Katsumata M, Yamane A, Ota M, Sato T-A: **Development of serum glycoproteomic profiling technique; simultaneous identification of glycosylation sites and site-specific quantification of glycan structure changes.** *Mol Cell Proteomics* 2010, **9**:1819–1828.
70. Whiteaker JR, Lin C, Kennedy J, Hou L, Trute M, Sokal I, Yan P, Schoenherr RM, Zhao L, Voytovich UJ: **A targeted proteomics-based pipeline for verification of biomarkers in plasma.** *Nat Biotechnol* 2011, **29**:625–634.
71. Picotti P, Aebersold R: **Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions.** *Nat Methods* 2012, **9**:555–566.
72. Kim YJ, Zaidi-Ainouch Z, Gallien S, Domon B: **Mass spectrometry-based detection and quantification of plasma glycoproteins using selective reaction monitoring.** *Nat Protoc* 2012, **7**:859–871.
73. Liu Y, Hüttenhain R, Surinova S, Gillet LC, Mouritsen J, Brunner R, Navarro P, Aebersold R: **Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS.** *Proteomics* 2013, **13**:1247–1256.
74. Hong Q, Lebrilla CB, Miyamoto S, Ruhaak LR: **Absolute quantitation of immunoglobulin G and its glycoforms using multiple reaction monitoring.** *Anal Chem* 2013, **85**:8585–8593.
75. Tian Y, Zhang H: **Glycoproteomics and clinical applications.** *Proteomics Clin Appl* 2010, **4**:124–132.
76. Kay Li Q, Gabrielson E, Zhang H: **Application of glycoproteomics for the discovery of biomarkers in lung cancer.** *Proteomics Clin Appl* 2012, **6**:244–256.
77. Yen T-Y, Macher BA, McDonald CA, Alleyne-Chin C, Timpe LC: **Glycoprotein profiles of human breast cells demonstrate a clear clustering of normal/benign versus malignant cell lines and basal versus luminal cell lines.** *J Proteome Res* 2011, **11**:656–667.
78. Tsai HY, Boonyapranai K, Sriyam S, Yu CJ, Wu SW, Khoo KH, Phutrakul S, Chen ST: **Glycoproteomics analysis to identify a glycoform on haptoglobin associated with lung cancer.** *Proteomics* 2011, **11**:2162–2170.
79. Liu Y, He J, Li C, Benitez R, Fu S, Marrero J, Lubman DM: **Identification and confirmation of biomarkers using an integrated platform for quantitative analysis of glycoproteins and their glycosylations.** *J Proteome Res* 2009, **9**:798–805.
80. Malaguarnera G, Giordano M, Paladina I, Berretta M, Cappellani A, Malaguarnera M: **Serum markers of hepatocellular carcinoma.** *Dig Dis Sci* 2010, **55**:2744–2755.
81. Wu J, Xie X, Liu Y, He J, Benitez R, Buckanovich RJ, Lubman DM: **Identification and confirmation of differentially expressed fucosylated glycoproteins in the serum of ovarian cancer patients using a lectin array and LC-MS/MS.** *J Proteome Res* 2012, **11**:4541–4552.

doi:10.1186/1559-0275-11-18

**Cite this article as:** Zhang *et al.*: Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. *Clinical Proteomics* 2014 **11**:18.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

