Glycosylation is one of the most important and common forms of protein post-translational modification that is involved in many physiological functions and biological pathways. Altered glycosylation has been associated with a variety of diseases, including cancer, inflammatory and degenerative diseases. Glycoproteins are becoming important targets for the development of biomarkers for disease diagnosis, prognosis, and therapeutic response to drugs. The emerging technology of glycoproteomics, which focuses on glycoproteome analysis, is increasingly becoming an important tool for biomarker discovery. An in-depth, comprehensive identification of aberrant glycoproteins, and further, quantitative detection of specific glycosylation abnormalities in a complex environment require a concerted approach drawing from a variety of techniques. This report provides an overview of the recent advances in mass spectrometry based glycoproteomic methods and technology, in the context of biomarker discovery and clinical application. Molecular & Cellular Proteomics 10: 10.1074/mcp.R110.003251, 1–14, 2011.

With recent advances in proteomics, analytical and computational technologies, glycoproteomics—the global analysis of glycoproteins—is rapidly emerging as a subfield of proteomics with high biological and clinical relevance. Glycoproteomics integrates glycoprotein enrichment and proteomics technologies to support the systematic identification and quantification of glycoproteins in a complex sample. The recent development of these techniques has stimulated great interest in applying the technology in clinical translational studies, in particular, protein biomarker research.

While glycomics is the study of glycome (repertoire of glycans), glycoproteomics focuses on studying the profile of glycosylated proteins, i.e. the glycoproteome, in a biological system. Considerable work has been done to characterize the sequences and primary structure of the glycan moieties attached to proteins (1–3), and their structural alterations related to cancer (4–6). Recent reports have provided a comprehensive overview of the concept of glycomics and its prospective in biomarker research (7–10). In contrast, this review is focused on recent developments in glycoproteomic techniques and their unique application and technical challenge to biomarker discovery.

Glycoproteomics in Biomarker Discovery and Clinical Study—Most secretory and membrane-bound proteins produced by mammalian cells contain covalently linked glycans with diverse structures (2). The glycosylation form of a glycoprotein is highly specific at each glycosylation site and generally stable for a given cell type and physiological state. However, the glycosylation form of a protein can be altered significantly because of changes in cellular pathways and processes resulting from diseases, such as cancer, inflammation, and neurodegeneration. Such disease-associated alterations in glycoproteins can happen in one or both of two ways: 1) protein glycosylation sites are either hypo, hyper, or newly glycosylated and/or; 2) the glycosylation form of the attached carbohydrate moiety is altered. In fact, altered glycosylation patterns have long been recognized as hallmarks in cancer progression, in which tumor-specific glycoproteins are actively involved in neoplastic progression and metastasis (5, 6, 11, 12). Sensitive detection of such disease-associated glycosylation changes and abnormalities can provide a unique avenue to develop glycoprotein biomarkers for diagnosis and prognosis. In addition, intervention in the glycosylation and carbohydrate-dependent cellular pathways represent a potential new modality for cancer therapies (6, 11, 13). Table I lists some of the FDA approved cancer biomarkers (14, 15) that are glycosylated proteins or protein complexes.

Protein biomarker development is a complex and challenging task. The criteria and approach applied for developing each individual biomarker can vary, depending on the purpose of the biomarker and the performance requirement for its clinical application (16, 17). In general, it has been suggested that the preclinical exploratory phase of protein biomarker development can be technically defined into four stages (18), including initial discovery of differential proteins; testing and selection of qualified candidates; verification of a subset of candidates; assay development and pre-clinical validation of potential biomarkers. Thanks to recent technological advances, mass spectrometry based glycoproteomics is now playing a major role in the initial phase of discovering...
Glycoproteomics Methods

A comprehensive analysis of glycoproteins in a complex biological sample requires a concerted approach. Although the specific methods for sample preparation can be different for different types of samples (e.g. plasma, serum, tissue, and cell lysate), a glycoproteomics pipeline typically consists of glycoprotein or glycopeptide enrichment, multidimensional protein or peptide separation, tandem mass spectrometric analysis, and bioinformatic data interpretation. For glycoprotein-based enrichment methods, proteolytic digestion can be performed before or after glycan cleavage, depending on the specific workflow and enrichment methods used. For glycopeptide enrichment, proteolytic digestion is typically performed before the isolation step so that glycopeptides, instead of glycoproteins, can be captured. For quantitative glycoproteomics profiling, additional steps, such as differential stable isotope labeling of the sample and controls, are required. Fig. 1 illustrates the general strategy for an integrated glycoproteomics analysis.

Glycoproteins or glycopeptides can be effectively enriched using a variety of techniques (see below). Following the enrichment step, the workflow then splits into two directions: glycan analysis and glycoprotein analysis. The strategies for glycan analysis have been discussed in several reviews and will not be covered in this report. For glycoprotein analysis, bottom-up workflows (“shotgun proteomics”—peptide based proteomics analysis) (23) are still most common, providing not only detailed information of a glycoprotein profile, but also the specific mapping of glycosylation sites. It is notable that the reliable analysis of mass spectrometric data in glycoproteomic studies largely relies on bioinformatic tools and glycosylation databases that are available. An increasing number of algorithms and databases for glycan analysis have been developed and well documented in several recent reviews (24–26). For glycoprotein and glycopeptide sequence analysis, a large number of well-characterized and annotated glycoproteins

Table I: Listing of some of the US Food and Drug Administration (FDA) approved cancer biomarkers

<table>
<thead>
<tr>
<th>Protein target</th>
<th>Glycosylation Detection</th>
<th>Source</th>
<th>Disease</th>
<th>Clinical biomarker</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Fetoprotein</td>
<td>Yes</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Non-Hodgkin lymphoma</td>
</tr>
<tr>
<td>Human chorionic gonadotropin-β</td>
<td>Yes</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Testicular cancer</td>
</tr>
<tr>
<td>CA15-3</td>
<td>Yes</td>
<td>Carbohydrate</td>
<td>Serum</td>
<td>Pancreatic cancer</td>
</tr>
<tr>
<td>CA125</td>
<td>Yes</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Ovarian cancer</td>
</tr>
<tr>
<td>CEA (carcinoembryonic antigen)</td>
<td>Yes</td>
<td>Protein</td>
<td>Serum</td>
<td>Colon cancer</td>
</tr>
<tr>
<td>Epidermal growth factor receptor Kit</td>
<td>Yes</td>
<td>Protein</td>
<td>Tissue</td>
<td>Colon cancer</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>Yes</td>
<td>Protein</td>
<td>Serum</td>
<td>Thyroid cancer</td>
</tr>
<tr>
<td>PSA-prostate-specific antigen (Kallkrein 3)</td>
<td>Yes</td>
<td>Protein</td>
<td>Serum</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>CA15-3</td>
<td>Yes</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>CA27-29</td>
<td>Yes</td>
<td>Glycoprotein</td>
<td>Serum</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>HER2/NEU</td>
<td>Yes</td>
<td>Protein (IHC), Protein</td>
<td>Tissue</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>Fibrin/FDP-fibrin degradation protein</td>
<td>Yes</td>
<td>Protein</td>
<td>Urine</td>
<td>Bladder cancer</td>
</tr>
<tr>
<td>BTA-bladder tumour-associated antigen (Complement factor H related protein)</td>
<td>Yes</td>
<td>Protein</td>
<td>Urine</td>
<td>Bladder cancer</td>
</tr>
<tr>
<td>CEA and mucin (high molecular weight)</td>
<td>Yes</td>
<td>Protein (Immunofluorescence)</td>
<td>Urine</td>
<td>Bladder cancer</td>
</tr>
</tbody>
</table>
Proteins can be found in the UniProt Knowledgebase. In addition, many glycopeptide mass spectra are now available in the continually expanding PeptideAtlas library (27), which stores millions of high-resolution peptide fragment ion mass spectra acquired from a variety of biological and clinical samples for peptide and protein identification. Ultimately, all the data obtained from different aspects of the workflow need to be merged and interpreted in an integrated fashion so that the full extent of glycosylation changes associated with a particular biological state can be better revealed. To the best of our knowledge, the complete glycoform analysis of any glycoprotein in a specific cell type under any specific condition has not yet been accomplished for any glycoprotein with multiple glycosylation sites. Current technology can define the glycan compliment and profile the glycoproteins, but is not capable of putting them together to define the molecular species present. To date, such integrated studies still remain highly challenging, even with advanced tandem mass spectrometry technologies and growing bioinformatic resources (26, 28–31).

FIG. 1. The strategies of mass spectrometry based glycoproteomic analysis.

Enrichment of the Glycoproteome—Characterization of the glycoproteome in a complex biological sample such as plasma, serum, or tissue, is analytically challenging because of the enormous complexity of protein and glycan constituents and the vast dynamic range of protein concentration in the sample. The selective enrichment of the glycoproteome is one of the most efficient ways to simplify the enormous complexity of a biological sample to achieve an in-depth glycoprotein analysis. Two approaches for glycoprotein enrichment have been widely applied: lectin affinity based enrichment methods (31–36) and hydrazide chemistry-based solid phase extraction methods (37–42). Recent studies have demonstrated that the two methods are complementary and a very effective means for the enrichment of glycoproteins or glycopeptides from human plasma and other bodily fluids (38, 39, 43). In addition, glycoprotein and glycopeptide enrichment using boronic acid (44, 45), size-exclusion chromatography (46), hydrophilic interaction (47) and a graphite powder micro-column (48) have been reported.

Lectin affinity enrichment is based on the specific binding interaction between a lectin and a distinct glycan structure attached on a glycoprotein (49, 50). There are a variety of lectin species that can selectively bind to different oligosaccharide epitopes. For instance, concanavalin A (ConA) binds to mannosyl and glucosyl residues of glycoproteins (51); wheat germ agglutinin (WGA) binds to N-acetyl-glucosamine and sialic acid (52); and jacalin (JAC) specifically recognizes galactosyl (β-1,3)N-acetylgalactosamine and O-linked glycoproteins (53). Lectin affinity enrichment has been designed to enrich glycoproteins with specific glycan attachment from plasma, serum, tissue, and other biological samples through affinity chromatography and other methods. Multiple lectin species can also be combined to isolate multiple types of glycoproteins in complex biological samples (54–59). Concanavalin A and wheat germ agglutinin, as well as jacalin are often used together to achieve a more extensive glycoproteome characterization (31, 34, 57, 59, 60). Several reports have demonstrated a multilectin column approach to achieve a global enrichment of glycoproteins with various glycan at-
tachyisms from serum and plasma (31, 34, 59, 61, 62). A recent study has developed a “filter aided sample preparation (FASP)” based method, which allows highly efficient enrichment of glycopeptides using multi-lectins (63). To date, most of the work using lectin affinity for targeted glycoprotein enrichment has focused on N-glycosylation because the binding specificity of lectin for O-glycosylation is less satisfactory. To overcome such caveat, efforts have been made using serial lectin columns of concanavalin A and jacalin in tandem to isolate O-glycopeptides from human serum (35).

A hydrazide chemistry-based method has been applied to isolate glycoproteins and glycopeptides through the formation of covalent bonding between the glycans and the hydrazide groups (37). The carbohydrates on glycoproteins are first oxidized to form aldehyde groups, which sequentially react with hydrazide groups that are immobilized on a solid surface. The chemical reaction conjugates the glycoproteins to the solid phase by forming the covalent hydrazone bond. Although, conceptually, the majority of the glycoproteins in a biological sample can be captured using this method, the further analysis of the captured glycoproteins is practically limited by the method that can cleave glycoproteins or glycopeptides from the solid phase. Because there is a lack of efficient enzymes or chemicals that can specifically deglycosylate and/or release O-linked glycoproteins or glycopeptides from the solid phase, most of the studies have applied this method solely for N-linked glycoprotein analysis. PNGase F is the enzyme that can specifically release an N-glycosylated proteins or peptides (except those carrying α1–3 linked core fucose (38)) from its corresponding oligosaccharide groups. The hydrazide chemistry method is not only highly efficient in enriching N-linked glycoproteins or glycopeptides from a complex environment, but also allows great flexibility in its applications, such as capturing extracellular N-glycoproteins on live cells to monitor their abundant changes because of cell activation, differentiation, or other cellular activities (64). This method can be readily automated for analyzing a large quantity of samples.

Recent studies have compared glycoprotein isolation methods. One study assessed lectin-based protocols and hydrophilic interaction chromatography for their performance in enriching glycoproteins and glycopeptides from serum (65). Other studies compared lectin affinity and hydrazide chemistry methods for their efficiency in isolating glycoproteins and glycopeptides from a complex biological sample (39, 66, 67). The methods are complementary in enriching glycoproteins because of their different mechanisms of glycoprotein capturing. When both methods were applied, it significantly improves the coverage of the glycoproteome, resulting in an increased number of glycoproteins identified. The lectin affinity method can be tailored to target glycoproteins with specific glycan structure(s) for isolation using different lectins, thus, affording flexibility for its application in glycoproteomic studies. The application of hydrazide chemistry method has been widely used for N-linked glycosylation study. The hydrazide chemistry essentially reacts with all the proteins with carbonyl groups, which may include glycoproteins with oxidized glycans (37, 40) and other oxidized proteins that carry carbonyl groups (68–70). The high specificity of this method may mainly result from the specificity of PNGase F, the enzyme cleaving N-glycosidic bonds to release N-glycoproteins and peptides from the solid phase. This method affords high efficiency and specificity in enriching N-linked glycoproteins or glycopeptides from a complex sample, and can be easily incorporated into a proteomics workflow for integrated analysis. In addition to the lectin and hydrazide chemistry-based methods, it has been suggested that boronic acid-based solid phase extraction may also be useful for an overall glycoproteome enrichment (44, 45), on the basis of the evidence that boronic acid can form diester bonds with most glycans, including both N-linked and O-linked glycosylation (71).

Mass Spectrometric Analysis of Glycoproteome—Mass spectrometry, because of its high sensitivity and selectivity, has been one of the most versatile and powerful tools in glycoprotein analysis, to identify the glycoproteins, evaluate glycosylation sites, and elucidate the oligosaccharide structures (56, 72, 73). The utility of a top-down approach (intact protein based proteomics analysis) (74) for glycoprotein characterization in a complex sample is still technically challenging with the current technology. The most versatile and widely used current glycoproteomics methods are based on characterizing glycopeptides generated by the digestion of glycoproteins, analyzing either deglycosylated glycopeptides or intact glycopeptides with glycan attachment, as illustrated in Fig. 1.

The direct analysis of intact glycopeptides with carbohydrate attachments is complicated by the mixed information obtained from the fragment ion spectra, which may include fragment ions from the peptide backbone, the carbohydrate group and the combinations of both. Although it is technically challenging to comprehensively analyze intact glycopeptides in a global scale for a complex biological sample, complementary information regarding peptide backbone and glycan structure can likely be obtained in a single measurement. Early work using collision-induced dissociation (CID) has identified a few key features that are characteristics of the fragmentation of glycopeptides, providing the basis for intact glycopeptide identification (75–79). The analysis of intact glycopeptides has been carried out using a variety of different instruments, including electrospray ionization (EST)-based ion trap (IT) (80–84), quadrupole ion trap (QiT) (85–87), Fourier transform ion cyclotron resonance (FTICR) (31, 57, 88, 89), ion trap/time-of-flight (IT/TOF) (90, 91), and quadrupole/time-of-flight (Q/TOF) (92–97); matrix-assisted laser desorption/ionization (MALDI) based Q/TOF (98–100),

---

1 The abbreviations used are: CID, collision-induced dissociation; Q, quadrupole.
In general, however, a high throughput, large scale automated workflow for analyzing intact glycopeptides in mixtures still requires intensive manual assignment and evaluation. A comprehensive interpretation of the fragment spectrum of an intact glycopeptide remains a challenge with current technology.

The analysis of deglycosylated peptides requires the removal of glycan attachments from glycopeptides. Fortunately, for N-linked glycopeptides, the N-glycosidic bond can be specifically cleaved using the enzyme PNGase F, providing deglycosylated peptides, which can then be analyzed directly using shotgun proteomics. The PNGase F-catalyzed deglycosylation results in the conversion of asparagine to aspartic acid in the glycopeptide sequence, which introduces a mass difference of 0.9840 Da. Such distinct mass differences can be used to precisely map the N-glycosylation sites using high resolution mass spectrometers. Stable isotope labeling introduced by enzymatic cleavage of glycans in $\text{H}_2\text{O}$ has also been used to enhance the precise identification of N-glycosylation sites (33, 142, 143). The removal of O-linked glycans is less straightforward, most assays rely on chemical deglycosylation methods, such as trifluoromethanesulfonic acid (144), hydrazinolysis (145), $\beta$-elimination (146), and periodate oxidation (35, 147). The application of these methods suffers from a variety of limitations, such as low specificity for O-linked glycosylation, degradation of the peptide backbone, and modifications of the amino acid residues—all of which can complicate or compromise O-linked glycoproteomics analysis in a complex sample. Most of the large scale glycoproteomics studies using the deglycosylation approach have been focused on N-glycoproteins, which are prevalent in blood and a rich source for biomarker discovery. O-glycosylation lacks a common core, consensus sequence, and universal enzyme that can specifically remove the glycans from the peptide backbone, thus, is more challenging to analyze for large scale profiling.

Following deglycosylation, the glycopeptides can be treated and analyzed as stripped peptides using a shotgun proteomics pipeline. MS/MS fragment spectra with b-ions and y-ions generated from CID are searched against protein databases using search algorithms, such as SEQUEST (148), MASCOT (149), and X!tandem (150), and subsequently validated via statistical analysis (151–154), to provide peptide and protein identifications with known false discovery rate. The N-glycosylation sites can be precisely mapped using the consensus sequence of Asn-X-Ser/Thr, in which asparagine is converted to aspartic acid following enzyme cleavage introducing a mass difference of 0.9840 Dalton. A variety of mass spectrometers have been used to analyze glycopeptides, in particular N-linked glycoproteins, in complex biological and clinical samples using the deglycosylation approach. These studies include electrospray ionization-based ion trap (37–39, 41, 67, 155–157), Orbitrap (158), Q/TOF (33, 35, 142, 155), quadrupole ion trap/time-of-flight (QIT/TOF) (86, 101, 102), and tandem time-of-flight (TOF/TOF) (81, 82, 101, 103–105) mass spectrometers. In general, the CID generated MS/MS spectrum of a glycopeptide is dominated by B- and Y-type glycosidic cleavage ions (carbohydrate fragments) (106), and b- and y-type peptide fragments from the peptide backbone. However, the MS/MS fragmentation data obtained from different instruments can have pronounced difference in providing structure information on glycan and peptide backbone, depending on the experimental setting and instrumentation used for mass analysis, including ionization methods, collision techniques and mass analyzers. Low energy CID with electrospray ionization-based ion trap, Fourier transform-ion cyclotron resonance, and Q/TOF instrument predominantly generates fragments of glycosidic bonds. The increase of collision energy using Fourier transform-ion cyclotron resonance, and Q/TOF instruments result in the more efficient fragmentation of b- and y- ions from the peptide backbone. MALDI ionization generates predominantly singly charged precursor ions, which are more stable and usually fragmented using higher energies via CID or post-source decay (PSD), generating fragments from both the peptide backbone and the glycan (98–100, 103, 107–110). Although Q/TOF instruments have been widely used for intact glycopeptide characterization, one unique feature of the ion trap instrument is that it allows repeated ion isolation/CID fragmentation cycles, which can provide a wealth of complementary information to interpret the structure of a glycan moiety and peptide backbone (56, 86, 86, 111). Recently, fragmentation techniques using different mechanisms from CID have been introduced and applied for glycopeptide analysis, including infrared multiphoton dissociation (IRMPD) (112–115), electron-capture dissociation (ECD) (112–120) and electron-transfer disassociation (ETD) (85, 121–123). The application of infrared multiphoton dissociation and electron-capture dissociation is largely performed with Fourier transform-ion cyclotron resonance instruments. Complementary to CID fragmentation, electron-capture dissociation and electron-transfer disassociation tend to cleave the peptide backbone with no loss of the glycan moiety, providing specific information on localizing the glycosidic modification. More details regarding mass spectrometric analysis of intact glycopeptides can be found in recent reviews (56, 124). Although great efforts have been made to apply a variety of mass spectrometry techniques to study both N-linked (32, 56, 86, 87, 112–114, 125–130) and O-linked (90, 116, 119, 120, 130–140) glycopeptides, the interpretation of the fragment spectrum of an intact glycopeptide still requires intensive manual assignment and evaluation. A recent study has demonstrated the feasibility to develop an automated workflow for analyzing intact glycopeptides in mixtures (141). In general, however, a high throughput, large scale profiling of intact glycopeptides in a complex sample still remains a challenge with current technology.
Glycoproteomics Methods

to utilize available proteomics technology for large-scale glycoproteome profiling, especially N-linked glycoproteins, in a high-throughput fashion.

Glycoproteomics Analysis in Blood and Other Bodily Fluids—An important target for blood-based diagnostic assays involves the detection and quantification of glycosylated proteins. Glycosylated proteins, especially N-linked glycoproteins, are ubiquitous among the proteins destined for extracellular environments (163), such as plasma or serum. A systematic and in-depth global profiling of the blood glycoproteome can provide fundamental knowledge for blood biomarker development, and is now possible with the development of glycoproteomics technologies. In the past few years, several large scale proteomics studies on profiling the glycoproteome of human plasma and serum have been reported (34, 35, 37, 38, 43, 61, 65, 164–166), adding significant numbers of glycoproteins into the blood glycoproteome database. In one study (38), immunoaffinity subtraction and hydrazide chemistry were applied to enrich N-glycoproteins from human plasma. The captured plasma glycoproteins were subjected to two-dimensional liquid chromatography separation followed by tandem mass spectrometric analysis. A total of 2053 different N-glycopeptides were identified, covering 303 non-redundant glycoproteins, including many glycoproteins with low abundance in blood (38). In a different study, hydrazide chemistry-based solid phase extraction method was applied to enhance the detection of tissue-derived proteins in human plasma (167). Other studies have applied lectin affinity-based approaches to characterize the serum and plasma glycoproteome (34, 43, 166). These studies provide detailed identification regarding the individual N-glycosylation sites using high-resolution mass spectrometry. The efforts made in global profiling of glycoproteins in plasma and serum have not only greatly enhanced our understanding of the blood glycoproteome, but also have facilitated the development of new technologies that can be used for glycoprotein biomarker discovery. A variety of experimental designs and strategies for blood glycoprotein profiling have been applied for clinical disease studies, including prostate cancer (168), hepatocellular carcinoma (164, 168–170), lung adenocarcinoma (61, 171), breast cancer (58, 165, 172), atopic dermatitis (169), ovarian cancer (173, 174), congenital disorders of glycosylation (175), and pancreatic cancer (156, 176). Most of these studies focused on the early stages of glycoprotein biomarker discovery and many of them exploited multilectin affinity techniques to isolate glycoproteins from serum or plasma.

Glycoproteomics techniques have also been applied to study the glycoproteome of other bodily fluids. The complementary application of hydrazide chemistry-based solid phase extraction and lectin affinity method have led to the identification of 216 glycoproteins in human cerebrospinal fluid (CSF), including many low abundant ones (39). A hydrazide chemistry based study on human saliva has charact
selective identification of cell surface N-glycoproteins (64). A complementary application of hydrazide chemistry and lectin affinity methods was demonstrated to profile cell membrane glycoproteins, significantly enhancing the glycoprotein identification (67).

**Quantitative Glycoprotein Profiling**—One of the major goals of clinical proteomics is to effectively identify dysregulated proteins that are specifically associated with a biological state, such as a disease. In the past decade, different quantitative proteomics techniques have been introduced and applied to study a wide variety of disease settings. These techniques are based on different mechanisms to facilitate mass spectrometric-based quantitative analysis, including stable isotopic or isobaric labeling using chemical reactions (e.g., ICAT and iTRAQ) (193–195), metabolic incorporation (e.g., SILAC) (196) and enzymatic reactions (e.g., 18O labeling) (197, 198); as well as less quantitatively accurate label-free approaches (199, 200). The overview and comparison of these quantitative techniques can be found in several reports in the literature and are not discussed in this review. Most of these isotopic labeling techniques can be adapted and utilized for glycoproteomics analysis to quantitatively compare the glycoproteome of a diseased sample to a control, thus revealing the glycosylation occupancy of individual glycosylation sites that may be involved in a disease. In addition to the well-established labeling methods cited above, several more experimental labeling strategies have been described in the field of glycoproteomics. One study demonstrated the feasibility of using stable isotope labeled succinimide anhydride for quantitative analysis of glycoproteins isolated from serum via hydrazide chemistry (37). In a different report, the heavy and light version of N-acetoxy-succinimide combining with lectin affinity selection was used to quantitatively profile serum glycopeptides in canine lymphoma and transitional cell carcinoma (201). Stable isotope labeled 2-nitrobenzenesulfenyl was also used for chemical labeling in a quantitative glycoprotein profiling study on the sera from patients with lung adenocarcinoma (202). O-Linked N-acetylgalactosamine (O-GlcNAc) is an intracellular, reversible form of glycosylation that shares many features with phosphorylation (203). Studies have suggested that O-GlcNAc may play an important role in many biological processes (204). A quantitative study on O-GlcNAc glycosylation has been reported, in which a method termed quantitative isotopic and chemoenzymatic tagging (QUIC-Tag) was described using a biotin-avidin affinity strategy for O-GlcNAc glycopeptide enrichment and stable isotope-labeled formaldehyde for mass spectrometric quantification (205). Recently, the isobaric tag for relative and absolute quantitation (iTRAQ) technique, combined with different glycoprotein enrichment approaches, has been utilized in several quantitative glycoproteomics studies. In the study of hepatocellular carcinoma, N-linked glycoproteins were enriched from hepatocellular carcinoma patients and controls using multillectin column and then quantitatively compared using iTRAQ to reveal the differential proteins associated with hepatocellular carcinoma (206). In a different study, the approach of using narrow selectivity lectin affinity chromatography followed by iTRAQ labeling was demonstrated to selectively identify differential glycoproteins in plasma samples from breast cancer patients (165). Another study utilized hydrazide chemistry-based solid phase extraction and iTRAQ to investigate the tear fluid of patients with climatic droplet keratopathy in comparison of normal controls, identifying multiple N-glycosylation sites with differential occupancy associated with climatic droplet keratopathy (178).

In addition to using chemical reactions to incorporate stable isotope tag for quantitative mass spectrometric analysis, 18O can be introduced into N-glycopeptides during enzymatic reactions, such as tryptic digestion (incorporation of two 18O into the peptide carboxyl-terminal) and PNGase F mediated hydrolysis (incorporation of one 18O into the asparagine of N-glycosylation sites (33)). Attempts have been made to apply this approach to identify differentially expressed N-glycosylation associated with ovarian cancer in serum (207). In a different approach, the SILAC technique allows incorporation of stable isotope-labeled amino acids into proteins during cell culturing process (196), and was applied to investigate the difference in cell surface N-glycoproteins among different cell types (64). A label-free approach has also been used for glycoproteomics profiling, including a method developed to profile intact glycopeptides in a complex sample (208) and a study that compares the plasma glycoproteome between psoriasis patients and healthy controls (209).

**Targeted Glycoproteomics Analysis**—Mass spectrometry based targeted proteomics has recently emerged as a multiplexed quantitative technique that affords highly specific and candidate-based detection of targeted peptides and proteins in a complex biological sample (18, 210–214). The technique is based on the concept of stable isotope dilution utilizing stable isotope-labeled synthetic reference peptides, which precisely mimic their endogenous counterparts, to achieve targeted quantification (214). Such techniques can be applied to target specific glycoproteins or glycopeptides, to precisely quantify the status of candidate glycosylation sites and assess the glycosylation occupancy at the molecular level. However, it is technically impractical to use synthetic peptides to precisely mimic a large number of natural glycopeptides with intact a glycan moiety as internal standards because of the structure complexity and variation of the sugar chain. To overcome these technical obstacles, an alternative approach was proposed for targeted analysis of N-glycosylation occupancy, in which stable isotope-labeled peptides were synthesized to mimic the deglycosylated form of candidate glycopeptides as internal references (161). It is known that the deglycosylation step using PNGase F results in a conversion of asparagine to aspartic acid in the peptide sequence, introducing a mass difference of 0.9840 Da. This phenomenon was utilized to design a synthetic peptide to mimic the en-
Glycoproteomics Methods

![Diagram of targeted analysis of N-glycopeptides]

Fig. 2. Targeted analysis of N-glycopeptides.

A deglycosylated N-linked glycopeptide in its deglycosylation form with exact amino acid sequence of its endogenous counterpart and with $^{13}$C and $^{15}$N labeling on one of its amino acids (161). Therefore, each matched pair of reference and endogenous candidate glycopeptides should share the same chromatographic and mass spectrometric characteristics, and can only be distinguished by their mass difference and isotopic pattern because of isotopic labeling. This design conceptually ensures that the synthetic internal standard of a candidate glycopeptide will be detected simultaneously with its endogenous form under the same analytical conditions, thus, minimizing the systematic variation and providing reliable quantification (214). The strategy for targeted glycoproteomics analysis is schematically illustrated in Fig. 2.

The targeted glycoproteomics technique was first demonstrated to analyze N-glycopeptides that were extracted from human serum using an integrated pipeline combining a hydrazide chemistry-based solid phase extraction method and a data-driven liquid chromatography MALDI TOF/TOF mass spectrometric analysis to quantify 21 N-glycopeptides in human serum (161). A similar mass spectrometric platform was then applied in a different study to assess a subset of glycoprotein biomarker candidates in the sera from prostate cancer patients (215). The targeted glycoproteomics analysis has also been demonstrated using a triple Q/linear ion trap instrument with the selected reaction monitoring (also referred to as multiple reaction monitoring) technique for highly sensitive targeted detection of N-glycopeptides in plasma (159). The technique was applied to detect tissue inhibitor of metalloproteinase 1 (TIMP1), an aberrant glycoprotein associated with colorectal cancer, in the sera of colorectal cancer patients (216) using a tandem enrichment strategy, combing lectin glycoprotein enrichment followed by the method of stable isotope standards and capture by antipeptide antibodies (SISCAPA), to enhance the detection of tissue inhibitor of metalloproteinase 1 (216). These studies demonstrate an integrated pipeline for candidate-based glycoproteomics analysis with precise mapping of targeted N-linked motifs and absolute quantification of the glycoprotein targets in a complex biological sample. Such targeted glycoproteomics can reach a detection sensitivity at the nanogram per milliliter level for serum and plasma detection (159, 214–216).

Concluding Remarks—The major challenge for a comprehensive glycoproteomics analysis arises not only from the enormous complexity and nonlinear dynamic range in protein constituent in a clinical sample, but also the profound biological intricacy within the molecule of a glycoprotein, involving the flexibility in glycan structures and the complex linkage with the corresponding protein. In the past decade, significant efforts have been made to structurally or quantitatively characterize the glycoproteome of a variety of biological samples, and to investigate the significant glycoproteins in a wide assortment of diseases. Shotgun proteomics-based techniques are still the most effective and versatile approach in glycoproteomics analysis, allowing high throughput and detailed analysis on individual glycosylation sites. Although glycoproteomics is quickly emerging as an important technique for clinical proteomics study and biomarker discovery, a comprehensive, quantitative glycoproteomics analysis in a complex biological sample still remains challenging. It is anticipated that with the continued evolution in mass spectrometry, separation technology, and bioinformatics many of the technical limitations associated with current glycoproteomics may be transient. There is no doubt that glycoproteomics is playing an increasingly important role in biomarker discovery and clinical study.

* This work was supported in part with federal funds from the National Institutes of Health (K25CA137222, R01CA107209, K07CA116296, R01DK081368) and a grant from Canary Foundation (TAB). RA acknowledges funding from the Swiss National Science Foundation (3100A0-107679).

† Address correspondence to: Sheng Pan, Ph.D., University of Washington, Department of Pathology, Seattle, WA 98195, span@u.washington.edu.

REFERENCES

Molecular Proteomics 10.1

Glycoproteomics Methods


Glycoproteomics Methods


91. Ito, H., Takegawa, Y., Deguchi, K., Nagai, S., Nakagawa, H., and Shino-
    hara, Y., Nishimura, S. (2006) Direct structural assignment of neutral and
    sialylated N-glycans of glycopeptides using collision-induced dissociation
    carbohydrate profiling of human transferrin by nano-flow liquid chromato-
    Spectrom. 18, 2983–2988
    40, 1472–1483
    ishi, T., and Hayakawa, T. (2006) Site-specific N-glycosylation analysis of
    human plasma ceruloplasmin using liquid chromatography with elec-
    tron capture dissociation tandem mass spectrometry. Anal. Biochem. 348,
    259–268
    in cycloxygenase-2 using mass spectrometry. Biochemistry 40, 3109–3116
96. Henriksson, H., Denman, S. E., Campuzano, I. D., Ademark, P., Master,
    E. R., Teeri, T. T., and Bruner, H., 3rd (2003) N-linked glycosylation of
    native and recombinant calf intestinal alkaline phosphatase. J. Biol. Chem.
    278, 61–73
    (2006) Determination and characterization of site-specific N-glycosyla-
    tion using MALDI-Qq-TOF tandem mass spectrometry: case study with
    a plant protease. Anal. Chem. 78, 1093–1103
    Site-specific N-glycosylation analysis: matrix-assisted laser desorption/
    ionization quadrupole-quadrupole time-of-flight tandem mass spectromet-
    ry of native and recombinant proteinase K from the human keratinocyte
    combined with off-line HPLC for characterization of protein primary
    structure and post-translational modifications. J. Biomet. Tech. 16, 429–440
100. Kubota, K., Sato, Y., Suzuki, Y., Goto-Inoue, N., Toda, T., Suzuki, M.,
    using lectin affinity chromatography with MALDI-TOF mass spectromet-
    ry. Anal. Chem. 80, 3683–3698
    Iida, J., and Suzuki, A. (2006) Structural characterization of glycopep-
    tides by N-terminal protein ladder sequencing. Anal. Chem. 78, 2239–2243
102. Kjeldsen, F., Haselmann, K. F., Budnik, B. A., Sorensen, E. S., and
    Zubarev, R. A. (2003) Complete characterization of posttranslational modi-
    fications in the bovine milk protein αS1 casein using tandem mass spec-
    trometry with electron capture dissociation as the last stage. Anal.
    Chem. 75, 3256–3262
    (2005) New features of site-specific horseradish peroxidase (HRP) gly-
    cosylation uncovered by nano-LC-MS with repeated ion-isolation/frag-
104. Adamson, J. T., and Håkansson, K. (2006) Infrared multiphoton dissoci-
    ation and electron capture dissociation of high-mannose type glycopep-
    tides. J. Proteome Res. 5, 493–501
    tion of O-glycosylation sites by MALDI-LIFT-TOF mass spectrometry.
    Glycoconjugate J. 21, 6097–6101
    proteomics Methods

Molecular & Cellular Proteomics 10.1

10.1074/mcp.R110.003251–11
Glycoproteomics Methods


tion of mucin gene products in paraffin embedded sections. Glycoconj. J. 17, 691–703


bases using mass spectrometry data. Electrophoresis 20, 3551–3567


tion of proteomic data generated by tandem mass spectrometry. Nat. Methods 4, 787–797

etry and sequence database searching. Methods Mol. Biol. 367, 87–119


syst. 5, 1298–1302

155. Roth, J. (2002) Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality con-
trol, and cell interactions. Chem. Rev. 102, 285–303

10.1074/mcp.R110.003251-12

Molecular & Cellular Proteomics 10.1


