Mass spectrometry-based analysis of glycoproteins and its clinical applications in cancer biomarker discovery

Farid Abu Shammala

Department of Pharmacology. University of Palestine. Gaza. Email: drfaridshammala@hotmail.com.

Abstract. Most proteins are glycosylated, glycosylation is one of the most important posttranslational modifications of proteins and plays essential roles in various biological processes. Aberration in the glycan moieties of glycoproteins is associated with many diseases. It is especially critical to develop the rapid and sensitive methods for analysis of aberrant glycoproteins associated with diseases. 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. It is especially critical to develop the rapid and sensitive methods for analysis of aberrant glycoproteins associated with diseases. Mass spectrometry (MS) has become a powerful tool for mapping glycoprotein glycosylation and detailed glycan structural determination. Especially, tandem mass spectrometry can provide highly informative fragments for structural identification of glycoproteins. This review provides an overview of the development of MS technologies and their applications in identification of abnormal glycoproteins and glycans in human serum to screen cancer biomarkers in recent years.

ORCID
0000-0001-8897-5184
Farid Abu Shammala

Introduction

Glycoproteomic analysis is complicated not only by the variety of types of carbohydrates, but also by the complex linkage of the glycan to the protein. Glycosylation can occur at several different amino acid residues in the protein sequence. The most common and widely studied forms are N-linked and O-linked glycosylation. O-linked glycans are linked to the hydroxyl group on serine or threonine residues. N-linked glycans are attached to the amide group of asparagine residues in a consensus Asn-X-Ser/Thr sequence (X can be any amino acid except proline) (Bause, 1983). Other known, but less well studied forms of glycosylation include glycosylphosphatidylinositol anchors attached to protein carboxyl terminus, C-glycosylation that occurs on tryptophan residues (Weim, 2009), and S-linked glycosylation through a sulfur atom on cysteine or methionine (Lote and Weiss, 1971; Vijayakrishnan et al., 2009). Our following work is focused on glycoproteomic analysis of the most common N-linked and O-linked glycoproteins. The two major types of glycosylation, are both involved in the maintenance of protein conformation, protein activity, protein protection from proteolytic degradation and protein intracellular trafficking and secretion

(Aebersold and Mann, 2003). N-glycan moieties also play a vital role in the folding, processing, and secretion of proteins from the endoplasmic reticulum (ER) and the Golgi apparatus (Aebersold and Mann, 2003).

In the past few decades, there has been growing attention to protein glycosylation in the biomedical field, since glycosylation alteration has been associated with a variety of diseases, including cancer, inflammatory and degenerative diseases. Indeed, 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 body environment require a concerted approach drawing from a variety of techniques.

Thus, the study of glycans and their localization and structural pattern will shed light on their role in protein regulation and function, both under normal conditions and in pathologies. In biotechnology, too, there is a need for an in depth understanding of glycosylation, since the efficient production of recombinant proteins, including glycoproteins (Aoki-Kinoshita, 2008), is becoming increasingly important. The structural aspects of glycans have been studied by mass spectrometry (MS) for many years, during the last decade the development of matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) MS for this class of compounds has accelerated substantially the acceptance of MS-based methodologies. In the structural analysis of complex glycans originating from various isolated glycoproteins

(Carlson, 1968; Dell, 1987; Hellerqvist, 1990; Varki, 1993; Tarentino and Plummer, 1994; Tretter et al., 1991; Faid et al., 2006; Ohtsubo and Marth, 2006; Faid et al., 2007; Morelle et al., 2005a; Morelle et al., 2005b; Morelle et al., 2006b;

Morelle et al., 2006c), MALDI-MS in conjunction with exoglycosidase digestion and a tandem (MS/MS) operation have become particularly popular (Küster, Wheeler, Hunter, Dwek and Harvey, 1997).

Although MALDI-MS structural analysis of most glycans can basically be performed in their native forms, there are several reasons for conversion of such compounds into their methylated derivatives. These include an easy determination of branching, interglycosidic linkages and the presence of configurational and conformational isomers. Permethylation also stabilizes the sialic acid residues in acidic oligosaccharides, yielding more predictable ion products when subjected to MS/MS experiments. Moreover, methylated carbohydrate deem to ionize more efficiently than their native counterparts. In conjunction with ESI and collision-induced dissociation (CID), permethylated carbohydrate also yield a most detailed structural information (Patel, 1993; Colangelo and Orlando, 1999; Huang et al., 2001; Wheeler and Harvey, 2001; Huang et al., 2002; Zaia, 2004; Taylor et al., 2006).

Most permethylation procedures, employed over a number of years in carbohydrate analysis, are derived from two successful methodologies. The first, originally described by Hakomori (Morelle et al., 2006c), utilizes the dimethyl sulfoxide anion (DMSO⁻, commonly referred to as 'dimsyl anion') to remove protons from the sample analyte molecules prior to their replacement with methyl groups. The second, and currently more widespread approach, introduced in 1984 by Ciucanu and Kerek (Powell and Harvey, 1996), is based on the addition of methyl iodide to DMSO containing powdered sodium hydroxide (NaOH). The original procedure was modified more recently (Ciucanu and Kerek, 1984). The current popularity of the modified procedure stems from its rapidity, experimental simplicity, 'cleaner' reaction products and the effectiveness for replacing protons at both oxygen and nitrogen sites in oligosaccharides. This methylation procedure has now been used for

derivatization of carbohydrates (Ciucanu and Costello, 2003; Kui Wong et al, 2003; Kang et al., 2005), other polyols, and fatty acids, including their hydroxyl derivatives (Morelle et al., 2004; An et al., 2009). However, it has gained a particular popularity with complex oligosaccharides and glycolipids. However, due to the challenges mentioned above, purification, enrichment and fractionation approaches are of the essence prior to MS analysis. Several comprehensive reviews about sample pretreatment of glycoproteins have been published in recent years (Fanayan et al., 2012; Ongay et al., 2012; Lazar et al., 2013), so we will not discuss them here. The current review focuses on the development of MS technologies and their application in the analysis of abnormal glycoproteins in biological samples to screen cancer biomarkers in recent years.

Experimental

Chemicals

Pancreatic bovine ribonuclease B, maltoheptaose, fetuin from fetal calf serum, human α_1-acid glycoprotein, proteomicsgrade trypsin, PNGase F, tris-(hydroxymethyl)amino-methane hydrochloride (Tris-HCl), ethylenediaminetetraacetic acid (EDTA), and ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N'*,*N'*-tetraacetic acid (EGTA) were obtained from Sigma. The MALDI matrix, 2,5-dihydroxybenzoic acid (DHB), and NaOH were purchased from Aldrich. Chloroform, methyl iodide, and sodium chloride were received from EM Science. Dithiothreitol (DTT), iodoacetamide, and 3-([3-cholamidopropyl]dimethylammonio)-1-propane-sulfonate (CHAPS) were purchased from Bio-Rad Laboratories. Ammonium bicarbonate was received from Mallinckrodt Chemical Company and sodium pyrophosphate was obtained from J. T. Baker, Inc. Acetonitrile (ACN) and hydrochloric acid solution N/10 were purchased from Fisher Scientific.

Experimental conditions

Reduction, alkylation, and proteolytic digestion One milligram of ribonuclease B (Sigma, St. Louis, MO) was reconstituted in 100 μL of solution containing 20 mM Tris-HCl pH 7.8, 6 M guanidine-HCl and 10 mM DTT. The mixture was incubated for 30 min at 50 °C, and 50 μL of 0.2 M iodoacetamide and 50 μL of 0.2 M ammonium bicarbonate pH 7.8 were added. The mixture was incubated at room temperature in the dark for two hours. The alkylated solution was dialyzed against 20 mM ammonium bicarbonate pH 7.8 overnight at 4° C using a Slide-A-Lyzer® MINI Dialysis Unit. The dialyzed protein was digested by adding a 1:60 ratio (enzyme:sample) of a proteolytic enzyme (Promega, Madison, WI) at a concentration of 0.52 mg/mL in 50 mM ammonium bicarbonate pH 7.8 and incubated overnight at 37 °C.

Sample analysis

Samples were introduced into the LTQ using the NanoMate™ 100 mounted in front of the LTQ and 5 μL samples (at 1 pmol/μL in 50% methanol; 0.1% formic acid) were infused at a flow rate of 100 nL/min. Generally, glycosylated protein analysis by MS is typically achieved by three main approaches.

In the first method, conventional glycan analysis is performed using chemical or enzymatic treatment to liberate the glycans from the protein, followed by derivatization prior to MS analysis.

In the second method, the intact protein is analyzed with no sample pre-treatment before analysis.

The third method, the analysis of glycopeptides is derived by proteolytic digestion of the glycosylated protein.

MS spectrometry

Mass Spectrometer: LTQ run in tune mode

Ionization Mode: Nano-electrospray

Ion Polarity: Positive
Spray voltage: 1.55 kV
Spray pressure: 0.2 psi
Capillary temperature: 150 °C
Normalized Collision Energy™: 20-25% for MS^n experiments
Maximum scan time: 50 ms
Number of micro scans summed for each scan: 2-3

Permethylation

For both the spin-column and capillary derivatization versions, the amount of NaOH used was the same, aside from the experiments involving the optimization of the capillary length. All permethylation conditions, including sample amount, amount of methyl iodide, and water/chloroform sample extraction, were maintained constant to ensure a valid quantitative comparison. In-solution permethylation, maltoheptaose and all N-glycans derived from glycoproteins were permethylated according to the procedure of Ciucanu and Costello. Briefly, methyl iodide, a trace of water, and NaOH powder were suspended in DMSO and mixed for 10 min at room temperature. Typically, 1–10 μg sample were suspended in 30 μL of DMSO, to which 3.6 mg of NaOH powder, 0.3 μL of water and 5.6 μL of methyl iodide were added.

Identification of a glycosylation site is important, since such knowledge could provide an indication of the function of that glycan. Moreover, assignment of a certain structure to specific site can also shed light on the protein's glycosylation profile and microheterogeneity and hence, plausibly, its activity. As mentioned above, removal of an entire N-glycan moiety by PNGase F or A results in the conversion of Asn to Asp and a shift of one mass unit for each N-glycosylation site. Thus, when a deglycosylated protein is further digested with trypsin, the peptides that are bound to the glycan moiety will be 1 Da heavier than the expected theoretical mass. By subjecting these peptides to MS/MS, each peptide that possesses Asp (instead of Asn) is identified as formerly attached to the glycan moiety. A similar approach involves glycan removal with PNGase F in the presence of the deglycosylated Asn would be labeled with, thus its mass, altered by 3 Da.

Results and Discussion

There are two general MS-based strategies for glycoprotein analysis. In the first one the intact glycoproteins are directly subjected to MS and tandem MS analysis to provide the protein sequencing ladders and in situ localization of complex glycans without extensive separation or digestion. Although most proteomics and glycomics analyses are performed in a positive-ion mode, the glycan moieties composition is diverse, and if the particular moiety contains N-acetylation and acidic residues, such as sialic acid, ionization may be prevented. In that case, a negative-ion mode could be employed, but further MS/MS and fragmentation analysis (performed in a positive mode) is usually not applicable. Positive ionization can be improved by adopting one of the following two strategies. The glycan moiety can be desialylated by sialidase that cleaves the terminal sialic acid. Although desialylation enables better ionization, some information is lost: in a pool of released oligosaccharides it is practically impossible to identify which oligosaccharide possessed the sialic acid. Alternatively, permethylation, in which all the glycan-free OH groups are methylated, will stabilize oligosaccharides containing sialic acid residues. Moreover, permethylation masks highly polar groups and confers a slight hydrophobicity; thus, the oligosaccharides are more easily separated from contaminants (e.g., salts) that may interfere with the analysis, and they also become more uniformly and efficiently ionized. Indeed, fragment ions obtained from permethylated derivatives are easily assignable to glycan sequences.

In our work, glycan site mapping is not always straightforward, and some challenges may be encountered. One challenge is to identify a glycopeptide in a large pool with an absolute majority of nonglycosylated peptides. Moreover, the

glycan moiety tends to suppress ionization compared to unmodified peptides. Therefore, in such a case glycopeptide enrichment should be employed. A vast variety of commercial resin-bound lectins are available for column chromatography. Another challenge involves the analysis of modified peptides, specifically glycopeptides; in this case, the commonly used CID and HCD fragmentation methods result in intensive fragmentation of the glycan moiety, while the peptide remains intact. Two recently developed fragmentation methods, electron capture dissociation (ECD) and electron transfer dissociation (ETD), may solve this problem. Both these methods, based on electronic excitation energy, result in peptide fragmentation that leaves the modification intact and attached to the amino acid. Thus, enriched glycopeptides can be subjected to CID-ETD fragmentation, in which the CID fragments the glycan moiety and the ETD sequences the glycopeptide with the attached glycan moiety.

Optimum chemical derivatization of analytes should modify quantitatively the required functional groups, yielding preferably a single and stable reaction product. Analyte modification should be fast, simple, and applicable to minute sample quantities. The last requirement is particularly crucial for the glycans derived from trace quantities of glycoproteins. The optimum conditions utilized for in-solution permethylation were used initially and optimized for the different approaches. Although each parameter was varied while the other parameters were kept constant, no major changes were observed for the studied parameter upon changing the values of the other parameters (data not shown). The effect of the amount of methyl iodide on permethylation efficiency was evaluated, the efficiency slightly decreased as the amount of methyl iodide increased. However, this decrease was within the error of measurement variation, indicating no substantial effect for a linear oligosaccharide.

Figure 1 shows *N*-linked glycans were released by subjecting monoclonal antibody mAb S057 to enzymatic treatment using PNGase F using method 1. The glycans were isolated and permethylated prior to MALDI-MS analysis. Three *N*-linked glycan masses were observed at m/z 1836, 2041, and 2245 that are assigned the glycan compositions corresponding to G0, G1 and G2 of S057 as shown in Figure 2. The glycan structure assignments on this figure were made based on knowledge of the core structure of *N*-glycans of antibodies and probable galactose additions.

Figure 3 represents deconvoluted MS data of intact reduced heavy chain of S057 using method 2. The theoretical molecular mass of intact protein was derived from known modifications and peptide sequence of immunoglobulin antibodies. The observed masses on the spectrum are consistent with the expected theoretical masses of the glycosylated S057. The differences in mass of these ions to the deglycosylated mass of the heavy chain (48,953 Da) provide a distribution of glycan moieties present in S057. The glycoform distribution closely correlates to the distribution observed in released glycans observed from the MALDI data.

Figure 4 shows LC-MS data from three glycopeptides that correspond to the three glycoforms (m/z 1318, 1399 and 1480) expected in S057 as described previously using method 3. The compositions of the glycan moieties of these glycopeptides are consistent to the data observed in Figure 2 and 3. MS/MS data confirmed both the peptide sequence and the glycan composition, i.e., the correct combination of hexoses (galactose and mannose), HexNAcs (*N*-acetylhexosamine) and fucoses assigned to these ions, thereby providing site-specific information. In this case the site of glycosylation was identified as Asn (N) in the tryptic peptide EEQYNSTYR. The glycan masses can either be obtained from (1) the difference between the mass of this peptide and the observed glycopeptide mass or (2) by adding the series of neutral glycan masses observed during MS/MS. The glycoform distribution in this figure also correlates well with the distribution observed in Figures 1 and 2.
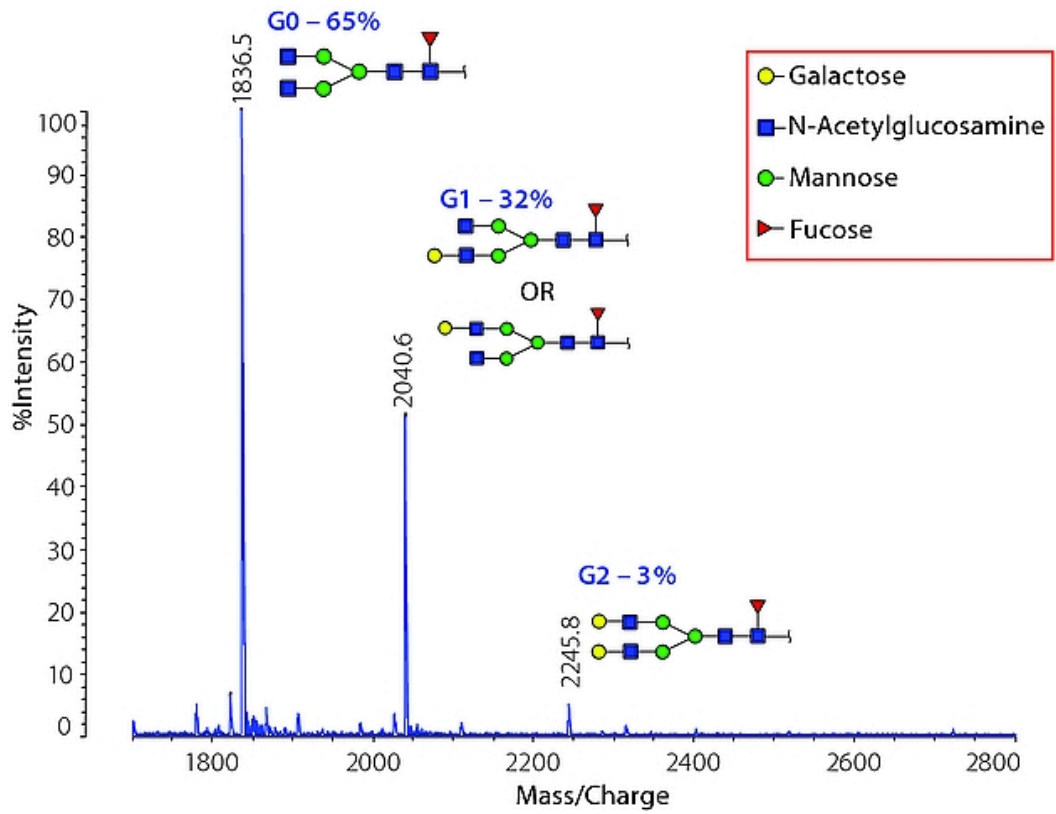
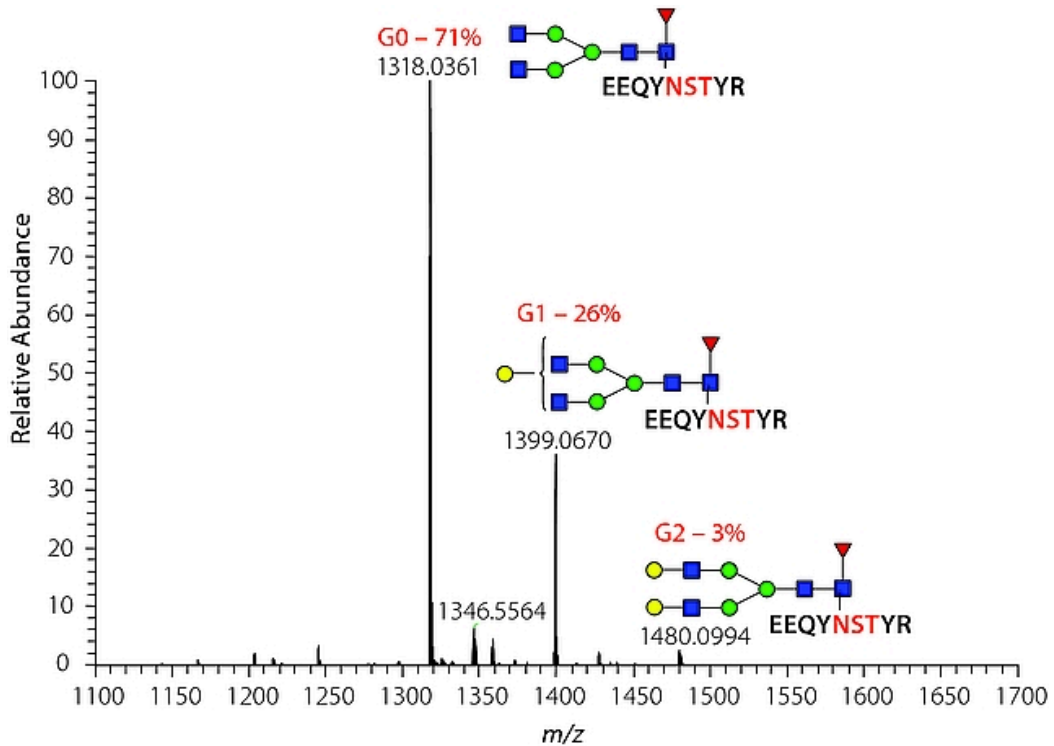**Figure 1**. MALDI-TOF MS of released glycans from S057 using Method 1.

Figure 2. ESI-LTQ/FT MS spectrum of glycopeptides identified from S057.
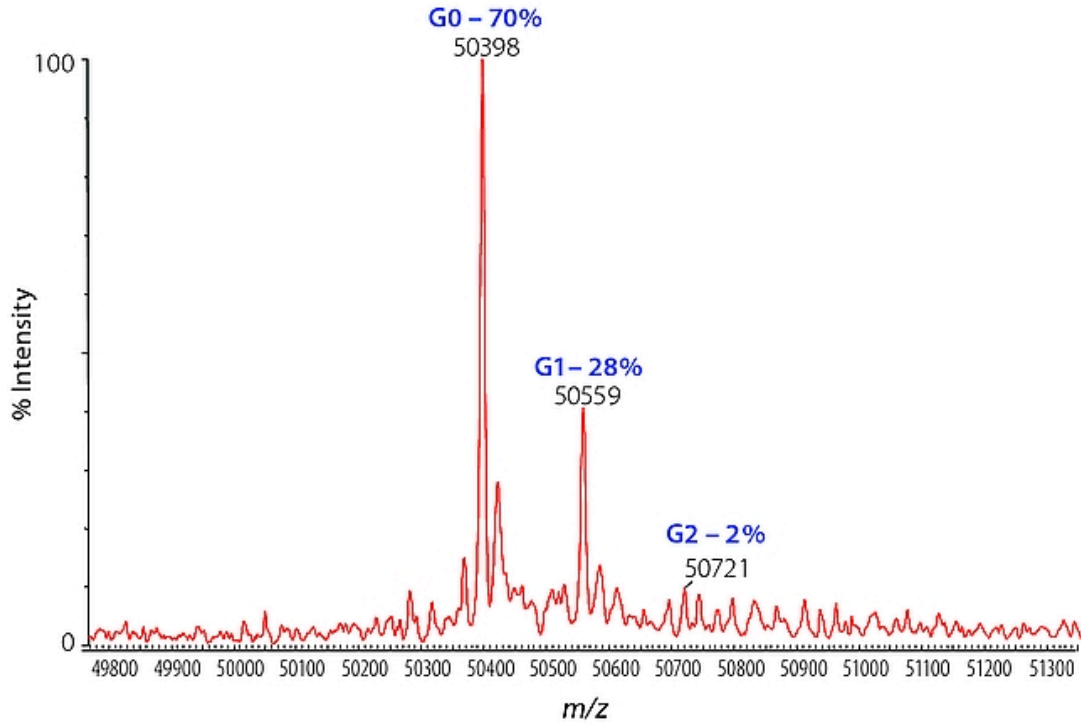
Figure 3. Deconvoluted MS spectrum of intact reduced heavy chain of S057.
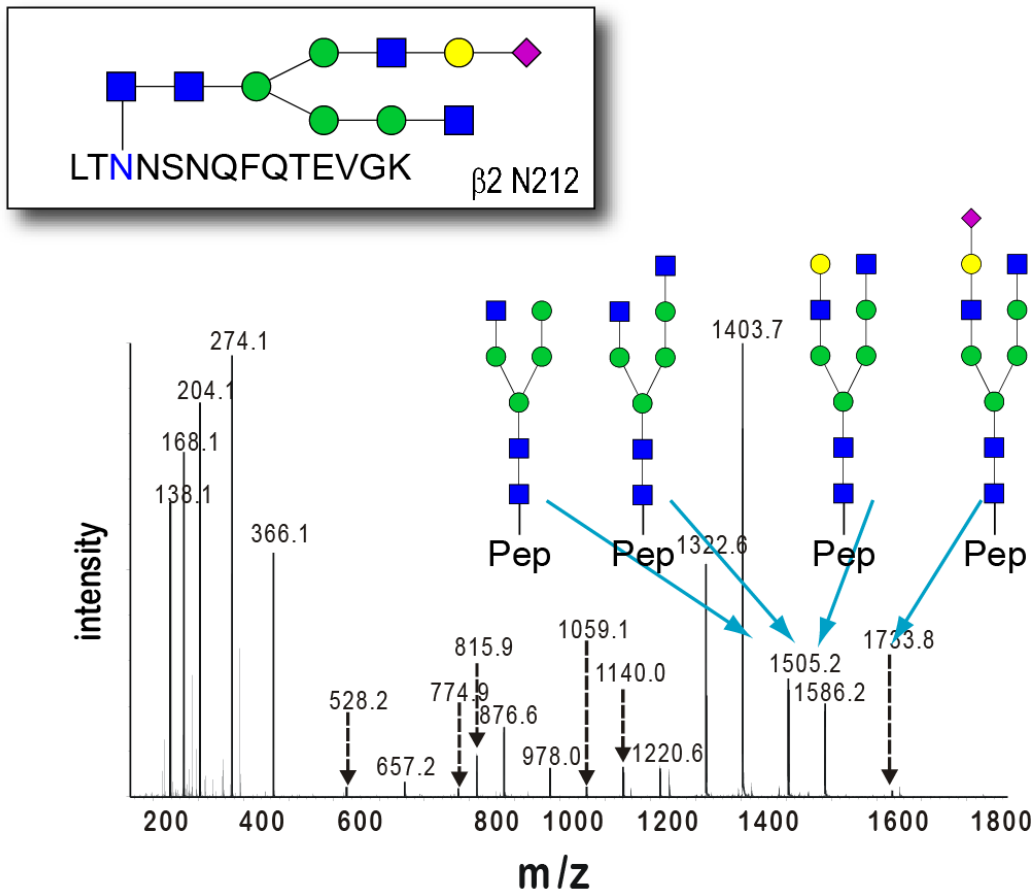

Figure 4. Glycan and glycopeptide sequencing by MALDI-TOF MS.

The results of conventional glycan analysis, using chemical or enzymatic treatment to liberate the glycans from the protein, followed by derivatization prior to MS analysis is shown in Figure 4. In this method N-linked glycans were released by subjecting monoclonal glycan to enzymatic treatment using PNGase F using method 1 and then the glycopeptide sequencing were sequenced by MALDI-TOF MS.

Table 1 shows List of results of some of the glycoprotein cancer biomarkers using method 1. CA: cancer antigen, FDP: fibrin degradation protein, sPIgR: secreted chain of the polymeric immunoglobulin receptor. Glycosylation, links the glycan chains to protein, is the most common and complex post-translational modifications of proteins. One of the important functions of glycans of glycoprotein is to maintain the ordered social life of each cell in multicellular organisms. Glycosylation is sensitive to the biochemical environment leading to the alternation of glycan chains under the physiological condition and specific diseases such as cancer. But cancer and other lethal diseases could not be diagnosed at early stage because of the traditional marker's low sensitivity and less specificity. Recent years, the development of glycan profiling technologies has offered great technology support for understanding the complex glycan structures and has increased great interest in applying the technology for cancer studies. Figure 5 shows some common N- and O-linked glycans altered carbohydrate structures expressed in various cancers.

Table 1. List of some of the glycoprotein cancer biomarkers. CA: cancer antigen, FDP: fibrin degradation protein, sPIgR: secreted chain of the polymeric immunoglobulin receptor.

Biomarker(a)	Glycosylation	Source	Disease
CA15.3	Yes	Serum	Breast
CA27-29	Yes	Serum	Breast cancer
HER2/NEU	Yes	Serum	Breast cancer
Fibrin/FDP	Yes	Urine	Bladder
CEA (carcinoembryonic antigen)	Yes	Serum	Colon cancer
Carcinoembryonic antigen (CEA)	Yes	Serum	Colon, breast, lung, pancreatic
Epidermal growth factor receptor	Yes	Tissue	Colon cancer
CA19-9	Yes	Serum	Gastrointestinal
KIT	Yes	Tissue	Gastrointestinal tumor
α-fetoprotein(AFP)	Yes	Serum	Hepatoma, testicular cancer
Human chorionic gonadotropin-β	Yes	Serum	Testicular cancer
Thyroglobulin	Yes	Serum	Thyroid cancer
CA125	Yes	Serum	Ovarian
PSA (prostate-specific antigen)	Yes	Serum	Prostate

In our work, the glycoproteins are digested by trypsin to generate nonglycosylation peptides and glycopeptides. Glycopeptides, deglycosylation peptides or nonglycosylation peptides derived from targeted glycoproteins in cancer samples can be analyzed by MS. This strategy is excellent to detect altered glycoproteins at glycoprotein level. For N-glycoprotein analysis, N-glycoproteins were enriched using hydrazide chemistry and lectin capture to remove nonglycoproteins, digested by trypsin and PNGase F treatment. Generated glycopeptides or deglycosylation peptides or nonglycosylation peptides were identified and quantified by LC-ESI-MS/MS. The profile of O-glycopeptides using MS technologies facilitates the analysis of altered O-glycoproteins in serum for clinical research. Two main strategies based MS, i.e. MALDI-MS in linear TOF mode and on-line LC-ESI-MS, were used to analyze cysteine-alkylated tryptic hinge O-glycopeptides isolated from solution digests. Data obtained from most of these laboratories were remarkably consistent despite the use of a variety of sample handling procedures and MS instruments. However, there were some shortcomings of these methods. In the MALDI mass spectra, sialylated glycopeptides had the low efficiency of the ionization in the positive and negative ion mode and sialyl residues were lost on profile of oligosaccharides or glycopeptides. Jacalin lectin chromatography was used to isolate the glycopeptides. However, the result was insufficient for a comprehensive MS study due to the volume (2 mL) of jacalin was too great to efficiently recover the glycopeptides from the lectin column for smaller scale analysis.

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.
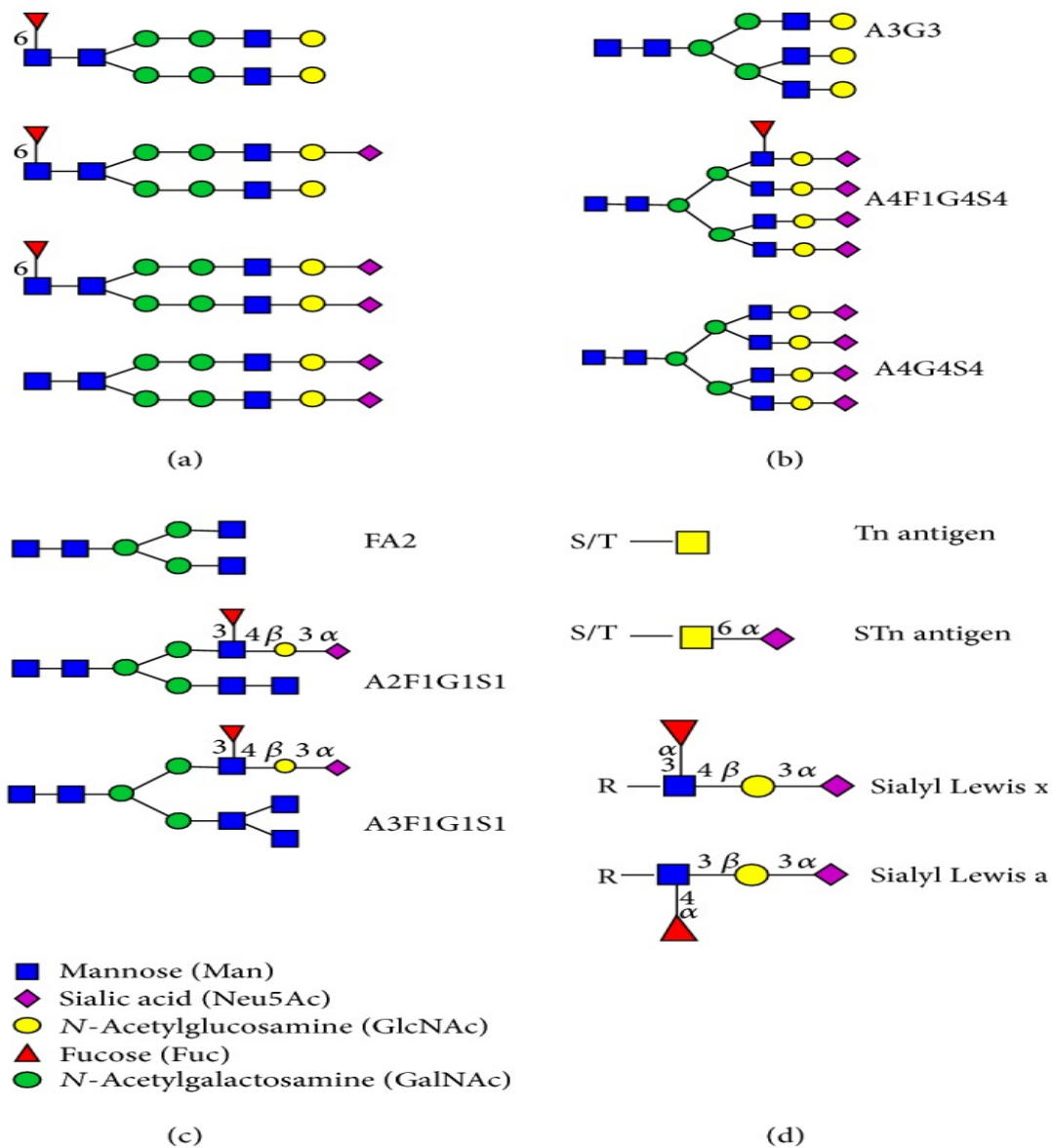
Figure 5. Common N- and O-linked glycans altered carbohydrate structures expressed in various cancers. (a) N-linked oligosaccharides expressed on AFP in HCC patients, the majority of which have core fucosylation. (b) N-linked oligosaccharide structures which change oligosaccharide structures which change in abundance as the cancer progresses. (c) N-linked oligosaccharide structures that are upregulated in lymph node metastasis positive breast cancers. (d) Tumour associated carbohydrate structures.

N-linked glycosylated-proteins, have been shown to be increasingly important in biomarker analyses. 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 (Tian and Zhang, 2010; Kay, Gabrielson and Zhang, 2012). Although there has been a great deal of progress in treating cancer 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. 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).

Conclusions

Although there is no universal method for comprehensive analysis and identification of glycoproteins, mass spectrometry has the great advantages in structural determination of glycoproteins. Mass spectrometry-based strategies are extensively used to detect the altered glycan and glycoprotein expression in human cancer samples to find cancer biomarkers. At present, despite MS-based strategy is dominated in glycoprotein analysis the technique still has some limitations. However, the characterization of glycoproteins is not always comprehensive

that either glycans or glycosylation peptides are detected without intact glycoprotein analysis. With the improvement of MS instruments, we believe MS-based approach will become a convenient, sensitive and rapid way to directly analyze the glycoprotein without time-consuming digestion and separation in the future. There is another approach in proteomics analysis that proteins are digested to larger peptides for MS analysis. This method could be complementary to the identification of glycoproteins. In addition, the improvements of purification, enrichment and fractionation will be great helpful for the characterization and quantitation of glycoproteins. MS-based strategies provide the valuable insights to better understand the cancer development and progression. Despite the 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.

Conflict of interest statement

Authors declare that they have no conflict of interests.

References

Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. **Nature**, v. 422, p.198-207, 2003.

An, H. J.; Froehlich, J. W.; Lebrilla, C. B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. **Curr. Opin. Chem. Bio.**, v. 13, no. 4, p. 421-426, 2009.

Aoki-Kinoshita. K. F. An introduction to bioinformatics for glycomics research. **PLoS Comput. Biol.**, v. 4, no. 4, p. 75-79, 2008.

Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. **Biochem. J.**, v. 209, p. 331-336, 1983.

Carlson, D. M. Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins. **J. Biol. Chem.**, v. 243, p. 616-626, 1968.

Ciucanu, I.; Costello C. E. Elimination of oxidative degradation during the per-O-methylation of carbohydrates. **J. Am. Chem. Soc.**, v. 125, p. 16213-16219, 2003.

Ciucanu, I.; Kerek, F. A simple and rapid method for the permethylation of carbohydrates. **Carbohydr. Res.**, v. 131, p. 209-217, 1984.

Colangelo, J.; Orlando, R. On-target exoglycosidase digestions/MALDI-MS for determining the primary structures of carbohydrate chains. **Anal. Chem.**, v. 71, p. 1479-1482, 1999.

Dell, A. FAB mass spectrometry of carbohydrates. **Adv. Carbohydr. Chem. Biochem.**, v. 45, p. 19-72, 1987.

Faid, V.; Chirat. F.; Seta. N.; Foulquier. F.; Morelle. W. A rapid mass spectrometric strategy for the characterization of N- and O-glycan chains in the diagnosis of defects in glycan biosynthesis. **Proteomics**, v. 7, p. 1800-1813, 2007.

Faid, V.; Evjen, G.; Tollersrud, O. K.; Michaslki, J. C.; Morelle, W. Site specific glycosylation analysis of the bovine lysosomal alpha mannosidase. **Glycobiology**, v. 16, p. 440-461, 2006.

Fanayan, S.; Hincapie, M.; Hancock, W. S. Using lectins to harvest the plasma/serum glycoproteome. **Electrophoresis**, v. 33, no. 12, p. 1746-1754, 2012.

Floyd, N.; Vijayakrishnan. B; Koeppe. J. R.; Davis. B. G. Thiyl glycosylation of olefinic proteins: S-linked glycoconjugate synthesis. **Angew Chem. Int. Ed. Engl. Davis,** v. 48, p. 7798-7802, 2009.

Hellerqvist. C. G. Linkage analysis using Lindberg method. **Methods Enzymol.**, v.193, p. 554-573, 1990.

Huang, Y.; Konse, T.; Mechref, Y.; Novotny, M .V. Matrix-assisted laser desorption/ionization mass spectrometry compatible beta-elimination of O-linked oligosaccharides. **Rapid Commun. Mass Spectrom**, v. 16, p. 1199-1204, 2002.

Huang, Y.; Mechref, Y.; Novotny, M. V. Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis. **Anal. Chem.**, v. 73, p. 6063-6069, 2001.

Kang, P.; Mechref, Y.; Klouckova, I.; Novotny, M. V. Solid-phase permethylation of glycans for massspectrometric analysis. **Rapid Commun. Mass Spectrom**, v. 19, p. 3421-3428, 2005.

Kay, L. Q.; Gabrielson, E. H.; Zhang, H. Application of glycoproteomics for the discovery of biomarkers in lung cancer. **Proteomics Clin. Appl.**, v. 11, p. 244-256, 2012.

Kui Wong, N.; Easton, R. L.; Panico, M.; Sutton-Smith, M.; Morrison, J. C.; Lattanzio, F. A.; Morris, H. R.; Clark, G. F.; Dell, A.; Patankar, M. S. Characterization of the oligosaccharides associated with the human ovarian tumor marker CA125. **J. Biol. Chem.**, v. 278, p. 28619-28634, 2003. https://doi.org/10.1074/jbc.M302741200

Küster, B..; Wheeler, S. F.; Hunter, A. P.; Dwek, R. A.; Harvey, D. J. Sequencing of N-linked oligosaccharides directly from protein gels: in-gel deglycosylation followed by matrix-assisted laser desorption/ionization mass spectrometry and normal-phase high-performance liquid chromatography. **Anal. Biochem.**, v. 250, p. 82-101, 1997.

Lazar, I. M.; Lee, W.; Lazar, A. C. Glycoproteomics on the rise: established methods, advanced techniques, sophisticated biological applications. **Electrophoresis**, v. 34, no. 1, p. 113-125, 2013.

Lote, C. J.; Weiss, J. B. Identification in urine of a low-molecular-weight highly polar glycopeptide containing cysteinyl-galactose. **Biochem. J.**, v. 123, p. 25-29, 1971.

Morelle, W.; Bernard, M.; Debeaupuis, J.-P.; Buitrago, M.; Tabouret, M.; Latgé, J.-P. Galactomannoproteins of *Aspergillus fumigatus*. **Eukaryotic Cell**, v. 4, no. 7, p. 1308-1316, 2005a. https://dx.doi.org/10.1128/EC.4.7.1308-1316.2005

Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. C. The use of mass spectrometry for the proteomic analysis of glycosylation. **Proteomics**, v. 6, p. 3993-4015, 2006a.

Morelle, W.; Donadio, S.; Ronin, C.; Michalski, J. C. Characterization of N-glycans of recombinant human thyrotropin using mass spectrometry. **Rapid Commun. Mass Spectrom.**, v. 20, p. 331-345, 2006b.

Morelle, W.; Faid, V.; Michalski, J C. Structural analysis of permethylated oligosaccharides using electrospray ionization quadrupole time-of-flight tandem mass spectrometry and deutero-reduction. **Rapid Commun. Mass Spectrom.**, v. 18, p. 2451-2464, 2004.

Morelle, W.; Flahaut, C.; Michalski, J. C.; Louvet, A.; Mathurin, P.; Klein, A. Mass spectrometric approach for screening modifications of total serum N-glycome in human diseases: application to cirrhosis. **Glycobiology**, v. 16, no. 4, p. 281-293, 2006c. https://doi.org/10.1093/glycob/cwj067

Morelle, W.; Jimenez. J. C.; Cieniewski-Bernard. C.; Dei-Cas. E.; Michalski. J. C. Characterization of the N-linked glycans of *Giardia intestinalis*. **Glycobiology**, v. 15, p. 549-559, 2005b.

Ohtsubo, K.; Marth. J. D. Glycosylation in cellular mechanisms of health and disease. **Cell**, v. 126, p. 855-867, 2006.

Ongay, S.; Boichenko, A.; Govorukhina, N.; Bischoff, R. Glycopeptide enrichment and separation for protein glycosylation analysis. **J. Sep. Sci.**, v. 35, no. 18, p. 2341-2372, 2012.

Patel. T.; Bruce, J.; Merry, A.; Bigge, C.; Wormald, M.; Parekh, R.; Jaques, A. Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. **Biochemistry**, v. 32, no. 2, p. 679-693, 1993. https://dx.doi.org/10.1021/bi00053a037

Powell, A. K.; Harvey. D. J. Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption/ionization mass spectrometry. **Rapid Commun. Mass Spectrom.**, v. 10, p. 1027-1032, 1996.

Tarentino, A. L.; Plummer. T. H. Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from *Flavobacterium meningosepticum*. **Methods Enzymol.**, v. 230, p. 44–57, 1994.

Taylor, A. M.; Holst, O.; Thomas-Oates, J. Mass spectrometric profiling of O-linked glycans released directly from glycoproteins in gels using in-gel reductive β-elimination. **Proteomics**, v. 5, p. 2936-2946, 2006.

Tian, Y.; Zhang, H. Glycoproteomics and clinical applications. **Proteomics Clin. Appl.**, v. 11, p. 124-132, 2010.

Tretter, V.; Altmann, F.; Marz. L. Peptide-N4-(N-acetyl-β-glucosaminyl) asparagine amidase F cannot release glycans with fucose attached α1-3 to the asparagine linked N-acetylglucosamine residue. **Eur. J. Biochem.**, v. 199, p. 647-652, 1991.

Varki. A. Biological roles of oligosaccharides: all of the theories are correct. **Glycobiology**, v. 3, p. 97-130, 1993.

Weim, X. L. Comparative glycoproteomics: approaches and applications. **Brief. Funct. Genomic. Proteomic**, v. 8, p. 104-113, 2009.

Wheeler, S. F.; Harvey. D. J. Extension of the in-gel release method for structural analysis of neutral and sialylated N-linked glycans to the analysis of sulfated glycans: application to the glycans from bovine thyroid-stimulating hormone. **Anal. Biochem.**, v. 296, p. 92-100, 2001.

Zaia. J. Mass spectrometry of oligosaccharides. **Mass Spectrom Rev.**, v. 23, p. 161-227, 2004.