Published in final edited form as:

Electrophoresis. 2011 January ; 32(1): 14–29. doi:10.1002/elps.201000394.

Recent Advances in the Mass Spectrometric Analysis of Glycoproteins: Capillary and Microfluidic Workflows

Diego F. Cortes¹, Jarod L. Kabulski¹, Alexandru C. Lazar³, and Iulia M. Lazar^{1,2,*}

¹Virginia Bioinformatics Institute Virginia Polytechnic Institute and State University Washington St. Bio II/283, Blacksburg, VA 24061

²Department of Biological Sciences, Virginia Polytechnic Institute and State University Washington St. Bio II/283, Blacksburg, VA 24061

³ImmunoGen, Inc., 830 Winter St., Waltham, MA 02451

Abstract

Recent developments in bioanalytical instrumentation, mass spectrometry (MS) detection and computational data analysis approaches have provided researchers with capabilities for interrogating the complex cellular glycoproteome, to help gain a better insight into the cellular and physiological processes that are associated with a disease, and to facilitate the efforts centered on identifying disease-specific biomarkers. This review describes the progress achieved in the characterization of protein glycosylation by using advanced capillary and microfluidic-MS technologies. The major steps involved in large-scale glycoproteomic analysis approaches are discussed, with special emphasis given to workflows that have evolved around complex MS detection functions. In addition, quantitative analysis strategies are assessed, and the bioinformatics aspects of glycoproteomic data processing are summarized. The developments in commercial and custom fabricated microfluidic front-end platforms to ESI- and MALDI-MS instrumentation, for addressing major challenges in carbohydrate analysis such as sensitivity, throughput and ability to perform structural characterization, are further evaluated and illustrated with relevant examples.

1. Introduction

The advent of mass spectrometry instrumentation has contributed to the initiation of a vast number of studies with focus on the elucidation of the complex biological glycoproteome. Although several approaches for the large-scale characterization of glycoproteins by MS have been explored, two main strategies have evolved. In the first strategy, the glycans are released from glycoproteins by enzymatic or chemical means, purified, and subjected to mass spectrometric analysis for the structural characterization of the released glycans [1, 2]. These so called “glycogenic” approaches ignore the identities of the involved proteins and focus on identifying the classes, compositions and structures of the released oligosaccharides [3]. In the second strategy, the glycoproteins are digested enzymatically, and the generated glycopeptides are analyzed by MS or tandem MS after enrichment and glycan release [4, 5]. These so called “proteogenic” approaches ignore the analysis of the glycan moieties, and sometimes even the identification of the glycosylation site. Overall, due to the complex nature of the analysis, most glycoproteomic studies focus either on the structural characterization of the released glycans, or on the identification of the glycosylated proteins, but not both. Recent advances in the capabilities of mass spectrometry

*To whom correspondence should be addressed. lazar@vt.edu.

instrumentation to perform sophisticated scanning functions (neutral losses, precursor ion scans, multi-stage MS), and make use of alternative ion activation/fragmentation techniques such as electron capture dissociation (ECD) and electron transfer dissociation (ETD), have opened the doors to analysis strategies that facilitate the comprehensive sequencing of the glycoproteome as a whole. Electrospray ionization (ESI) or matrix assisted laser desorption ionization (MALDI), in positive or negative ion operation modes, have been used extensively with virtually all types of available MS instrumentation, i.e., ion trap (IT), linear trap quadrupole (LTQ), time-of-flight (TOF), quadrupole/triple quadrupole (Q), Orbitrap and Fourier-transform ion cyclotron resonance (FTICR) mass analysis platforms.

Parallel to the developments in MS, the miniaturization of separation methodologies such as liquid chromatography (LC) and capillary electrophoresis (CE) represents a developing trend in proteomic and glycoproteomic analysis [6–21]. The advantages of using miniaturized systems include improved efficiency in sample processing, throughput, response time, and automation. Ideally, all steps including sample handling, (pre)treatment, chemical reaction, analytical separation, isolation and analyte detection could be integrated on a single miniaturized device. In terms of throughput, analysis times of only 40–55 s/sample are achievable with automated chip-infusion technology [22], at nL/min flow rates, and using sample volumes as small as 1–10 pL [16, 22]. Multiplexed devices with as many as 96–768 processing lines have been already devised [23–29], including devices that feature MALDI-MS detection for proteomic applications [27]. The use of miniaturized instrumentation in the form of microchip electrophoresis was first explored for the analysis of proteins and the separation of monosaccharides ten years ago [30], and later demonstrated for the analysis of *N*- and *O*-glycosylated peptides, as well [7, 11, 13, 31]. Recent developments have advanced MS detection as a glycan profiling tool for microfluidic devices, with the implementation of CE and LC separation techniques and direct hyphenation to MS in a high-throughput manner [14, 16]. These advances have been driven by the fact that often times the proteomic sample is very limited in size, and that components of interest such as glycoproteins are sometimes present only at the sub-microgram level. The ability to integrate multiple steps such as sample preparation, purification, separation and detection into a small analytical unit, such as a microchip, minimizes the number of handling and transfer steps during analysis, reduces protein losses due to adsorption on the instrumentation surface before detection and measurement, improves separation efficiencies, and ultimately significantly enhances the sensitivity and performance of the analytical method.

This review is aimed at providing an overview of some of the most advanced analytical and technological strategies that have been recently implemented for the analysis of glycoproteins. Innovative workflows for the selection, sequencing, and identification of glycoconjugates are described. Details about the MS experimental conditions that enabled the elucidation of peptide/glycan structures are emphasized. Applications involving the use of commercial microfluidic platforms, such as the NanoMate™ 100 [32] and the ChipCube [7], as well as of custom fabricated miniaturized devices with ESI and MALDI-MS detection, are summarized. The impact of miniaturization on improving workflow efficiency is underscored.

2. A generic strategy for the analysis of glycoproteins

An outline of the major steps involved in a generic glycoproteomic approach, as applied to the analysis of complex biological samples, is provided in Figure 1. The sample is typically subjected to an initial pre-treatment step (tissue delipidation, removal of highly abundant proteins from plasma, etc.), followed by a glycoprotein enrichment step via non-covalent/covalent chemistries (lectin/multilectin chromatography, hydrazide resin bonding, etc.) [33–

39], and, occasionally, by an additional 1D/2D-gel electrophoresis step with immunoblot or lectin-blot visualization for the fractionation/isolation of a narrower class of glycoproteins [40–42]. 1D/2D-gel separations tend to be, however, omitted from high-throughput analysis platforms due to poor detection of low abundance proteins and low protein recovery from the gels. Further MS analyses of the enriched glycoprotein samples tend to be divided into two major categories, i.e., (a) the identification of glycoproteins and detection of glycosylation sites, and (b) the characterization of the free glycans. The analysis of glycoproteins involves proteolytic digestion with an enzyme or a combination of enzymes (trypsin, pepsin, thermolysin, proteinase K, pronase, etc. [5, 43, 44]), glycopeptide enrichment (lectin chromatography, hydrazide chemistry, hydrophilic interaction chromatography (HILIC), etc. [45–48]), optional peptide fractionation/clean-up on reversed phase or strong cation exchange chromatography (SCX) columns [49], and nano-LC-MS/MS via a variety of tandem MS technologies including CID, ETD, ECD and ion mapping [8, 44, 50, 51]. The mixture of peptides/glycopeptides can be analyzed with or without the release of glycans, enzymatic release being commonly performed with PNGase F [45]. Conventional CID-MS/MS will generally perform well for the identification of non-glycosylated or glycan-free peptides, while ETD/ECD-MS/MS will facilitate the analysis of glycopeptides by enabling the identification of glycosylated sites. For simple mixtures, or for isolated proteins, peptide mass fingerprinting by MALDI-MS prior and after glycan release can be performed.

For the characterization of the glycan moiety, glycans are released from proteins and peptides by chemical or by enzymatic means. Chemical release is performed by β -elimination, hydrazinolysis or other milder hydrolysis approaches [52–54]. For enzymatic release, depending on the type of glycosylation that is pursued, a variety of enzymes can be used (*N*-glycosidases, *O*-glycosidases, sialidases, etc. [53, 55–57]). Glycans released from proteins or peptides are further isolated or fractionated by solid phase extraction on graphitized carbon cartridges [54], or by chromatographic techniques such as reversed phase liquid chromatography (RPLC), normal phase liquid chromatography (NPLC), HILIC or anion exchange chromatography (AXC). Derivatization with fluorescent tags enables detection and conversion of glycan retention times to glucose units (GU) [58]. Next, glycans are permethylated to facilitate MS analysis [59, 60], and ultimately analyzed by MALDI-MS or ESI-MS/MS with or without prior nano-LC separation.

3. Advanced MS workflows for the analysis of glycoproteins

Based on the generic strategy that was described in Figure 1, a broad variety of approaches for the analysis of glycoproteins have been developed. As it is not feasible to discuss all contributions to the field, in the followings, we review a few select examples that capitalize on the complexity of these approaches at the sample preparation and mass spectrometric analysis levels. For the MS savvy reader, emphasis is placed on the sophistication of MS functions that enabled, ultimately, the characterization of the glycoproteome, and on the specific sample and data analysis features that resulted in the confident identification and quantitation of a large number of glycosylated proteins and sites (*N*-linked, *O*-linked and sialylated). The outcome of each experiment is provided. Alternative techniques that enabled the validation of results are also described. As aberrant protein glycosylation seems to play a major role in numerous diseases, the specific biological or clinical applications that benefited from the described workflows are highlighted. Theoretical considerations that led to the development of these glycoproteomic workflows are discussed in the preceding manuscript [61].

3.1. MS workflows for characterizing complex glycoprotein samples

Alvarez-Manila *et al.* have described an elegant approach for the analysis of an embryonic stem cell protein extract [pluripotent murine embryonic stems cells (ES) vs. ES cells differentiated into embryoid bodies, $\sim 1 \times 10^7$ cells] by tryptic digestion, *N*-linked glycopeptide enrichment by lectin chromatography (ConA), PNGase F release of glycans in H_2^{18}O , and nano-LC-MS/MS on an LTQ-MS [45]. Glycan release in the presence of H_2^{18}O enabled glyco-site localization on a low-resolution MS instrument. The released glycans were permethylated with $^{12}\text{C}/^{13}\text{C}$ methyl iodide and analyzed by MALDI-TOF-MS detection for generating differential global MALDI-MS profiles. The identified glycans were annotated based on inferences made from the monosaccharide compositions derived from the masses of the probable structures that were consistent with ESI-MS² analysis. This study enabled the identification of 180 glycoproteins and 293 unique *N*-linked glycopeptide sequences. Some of these *N*-linked glycoproteins represent candidate biomarkers characteristic of certain developmental stages during embryonic stem cell differentiation.

Abbott *et al.* have characterized glycoproteins isolated from ovarian tumor and normal tissues by using a method involving multilectin affinity chromatography and nano-LC-MS/MS on an LTQ mass spectrometer [4]. *N*-linked glycosylated proteins were isolated from the biological sample (~ 50 mg tissue/20 mg delipidated protein powder) with the aid of biotinylated lectins (E-PHA, AAL and DSL) and captured on paramagnetic streptavidin beads. After release from the beads, the glycoproteins were reduced, alkylated, digested with trypsin and subjected to nano-LC-MS/MS. While the MS analysis set-up could not confirm the presence of specific glycosylated peptides/glycosylation sites, the presence of certain glycoproteins was confirmed by immunoprecipitation with antibodies against selected glycoproteins, SDS-PAGE and Western/lectin blotting. After multilectin enrichment, a total of 504 unique proteins were identified in the ovarian tumor and 315 unique proteins in the ovarian normal tissues. Relevant glycoproteins were assigned to functional categories that play a significant role in the progression and metastatic spread of ovarian cancer. Using a similar strategy involving the isolation of *N*-linked glycoproteins via biotinylated L-PHA/streptavidin bead capture, Abbott *et al.* [50] have characterized tumor and normal breast cancer tissues (~ 100 mg) by using complementary ESI-MS/MS methods, i.e., nano-LC-MS/MS for the analysis of tryptic glycopeptides, and infusion-MS/MS ion mapping for the analysis of permethylated glycans released with PNGase F from tryptic/chymotryptic peptides generated from the same sample.

Chen *et al.* [5] used hydrazide chemistry and multiple enzyme digestion to analyze *N*-linked glycopeptides from human liver tissue. As glycosylated tryptic peptides are large in mass and may not fall in the preferred detection range of most mass spectrometers, in addition to trypsin, the liver protein extract (~ 5 mg) was digested separately with pepsin and thermolysin. Glycopeptides were captured on an affinity hydrazide gel, glycans were released from the captured peptides with PNGase F, and the liberated peptides were analyzed by a 2D-SCX/nano-LC protocol with linear ion trap tandem MS detection. Approximately $\sim 85\%$ of the identified glycopeptides contained the Asn-Xxx-Ser/Thr glyco-motif. Tryptic digestion alone enabled the identification of 622 *N*-glycosites, while the use of pepsin and thermolysin resulted in the identification of additional 317 glycosites. Combined results matched these peptides to 523 non-redundant glycoproteins.

A “glycoproteomic reactor” was developed by Zhou *et al.* for the analysis of small volumes of human plasma [62]. After enrichment on a ConA column, glycoproteins were analyzed on the glycoproteomic reactor that performed protein concentration/purification, disulfide bond reduction, peptide-*N*-glycosidase-mediated ^{18}O -labeling and deglycosylation, alkylation, tryptic digestion and pH-based fractionation. The glycoproteomic reactor consisted of a $700\ \mu\text{m}$ i.d. $\times$ 8 cm long fused silica capillary that was packed with SCX particles ($10\ \mu\text{m}$). All

sample processing steps were performed by loading various reagents and eluents on the reactor, while proteins continued to be retained on the SCX particles until elution with varying pH buffers (pH~3–12). Using as little as 5 μ L of plasma, 82 unique glycopeptides representing 41 unique glycoproteins were identified. The total analysis time using the glycoproteomic reactor was reduced to <1.5 h, and reagent consumption was estimated to be reduced ~1000-fold.

The analysis of *O*-linked glycoproteins is somewhat hindered by the lack of broad specificity enzymes such as the *N*-glycanase hydrolases that are so frequently used for the analysis of *N*-linked glycoproteins. To address this limitation, additional sample preparation/processing approaches have been developed. For example, Schwientek *et al.* [41] have analyzed the *O*-glycoproteome of *Drosophila melanogaster* S2 cells by using a serial lectin affinity enrichment approach (Jacalin, PNA and HPA), followed by 2D electrophoresis (2DE), in-gel tryptic digestion and MALDI-MS. The presence of glycoproteins was validated by either 1D-SDS-PAGE or mAbs against the anti-Tn, anti-T, anti-sialyl-Tn, antimucin type A and antimucin type H carbohydrate epitopes, or by nano-LC-Q-TOF-MS analysis via precursor ion scanning for the diagnostic sugar oxonium-ion fragments HexNAc⁺ (*m/z* 204). Further, sugars released by β -elimination were permethylated and analyzed by gas chromatography (GC)-MS, or monosaccharide compositional analysis was performed by methanolysis, re-*N*-acetylation/trimethylsilylation and GC-MS. Twenty one secreted and intracellular mucin-type *O*-glycoproteins were identified. Alternatively, Nandi *et al.* [63] have demonstrated the analysis of *O*-GlcNAc modified proteins by a tagging-via-substrate (TAS) approach. The method relies on the metabolic labeling of cells by the incorporation of (peracetylated) *O*-GlcNAc azide analogues during cell culture. The azide handle is used for chemoselective conjugation to a biotinylated phosphine reagent, which further facilitates the enrichment/detection of these proteins on avidin/streptavidin columns. The *O*-GlcNAc modified proteins are next digested with trypsin and analyzed by nano-LC-MS/MS on an ion trap instrument. While MS analysis was not used for identifying the glycosylated peptides, confirmation of the azido-GlcNAc modification was accomplished by immunoprecipitation with specific antibodies and Western blotting. The study led to the identification of 199 putative *O*-GlcNAc modified proteins in HeLa cells, of which, 23 were confirmed by immunoprecipitation.

3.2. MALDI-MS fingerprinting

While not involving capillary separations, it is worth summarizing the success accomplished by MALDI-MS fingerprinting in generating glycomic maps with utility in studying the dynamics of a broad range of biological processes and diseases. Overall, such methods greatly benefit from simplicity at the MS detection level. Common sample analysis strategies include the following steps: glycoprotein enrichment, enzymatic digestion, PNGase F release of *N*-glycans, separation of *N*-linked glycans from their peptide counterparts and *O*-linked glycans, permethylation and MALDI-MS analysis. Similarly, the *O*-linked glycans that are preserved on peptides after PNGase F treatment, can be further released by reductive elimination and prepared for MALDI-MS. Structure assignment for the identified ions is accomplished with the aid of bioinformatics tools available via GlycoWorkBench, NetNGlc and NetOGlc (see next section on bioinformatics tools). Structure confirmation is typically performed via MALDI-TOF/TOF-MS or ESI-MS/MS [42, 64]. Using such an approach, Pang *et al.* [64] were able to detect high mannose, bi-/tri-/tetraantennary core-fucosylated complex type *N*-glycans with antennae terminated with Lewis^X/Lewis^Y sequences or capped with sialic acids, and Core1/Core2 *O*-glycans, and study human innate and adaptive immune response in seminal plasma. Novotny *et al.* have used extensively MALDI-MS profiling of *N*- and *O*-glycans to demonstrate the utility of the method for the diagnosis of cancer (breast, prostate, liver, hepatocellular and esophageal

cancers) [65–69]. Data processing via PCA (principal component analysis) or ROC (receiver operating characteristic) curve analysis enabled the selection of putative biomarkers from as little as 10 μ L of serum [69]. For example, hepatocellular carcinoma putative biomarkers were identified by using a combined experimental/computational approach that involved the analysis of 203 serum samples (78 hepatocellular carcinoma, 52 chronic liver disease and 78 healthy cases) by extraction of glycoproteins, *N*-glycan release, permethylation, MALDI-TOF/TOF-MS analysis of released glycans, and peak selection by multiple random sampling strategies with recursive feature selection based on support vector machines [68]. Ten peak sets were generated with seven peaks being shared by each of the ten peak sets, of which four peaks (i.e., glycans with elucidated structure) could be assigned to proteins with relevance to cancer.

As alteration of sialylation patterns has been shown to be involved in cancer progression, efforts have been also focused on characterizing sialic acid linkages and structural features. Most methods typically involve glycoprotein enrichment, enzymatic digestion, glycan release, glycan separation into neutral and acidic (sialic acid containing) fractions on a graphite column, permethylation and MALDI-MS mapping of the glycan pool [55]. Glycan sequencing to determine the sialylation sites and linkages is accomplished via MALDI-TOF/TOF-MS. Nano-LC-ESI-MS/MS via precursor ion scanning for specific oxonium ions (m/z 204, 366) of selected gel separated proteins may complement MALDI-MS in identifying the protein carriers of implicated glycotopes [70]. Sialic acid amidation/lactonization can be used to distinguish between α 2,3- and α 2,6-linked sialic acids [55]. Prior to MALDI-MS/MS analysis, Ghesquière *et al.* have incorporated COFRADIC (combined fractional diagonal chromatography) technology combined with neuraminidase/PNGase F treatment and $^{16}\text{O}/^{18}\text{O}$ labeling of samples to map sialylated *N*-glycosites in mice serum. Sialylation events on peptides were detected either by a retention time shift and Asn-to-Asp conversion in the Asn-X-Ser/Thr motif after neuraminidase treatment, or, by the appearance of a sialylated peptide as an isotope single (^{18}O), chromatographic shift and Asn-to-Asp conversion [71].

3.3. MS quantitation of glycoproteins

To take full advantage of the complex information that mass spectrometry can generate in glycoproteomics research, technologies must be developed not just for the structural characterization of glycans and their associated proteins, but also for assessing quantitative changes in the glycosylation patterns. The quantitation of glycoproteins has followed the same trends as the quantitation of proteins, i.e., stable isotope label-based and label-free approaches have been developed. Among the isotope labeling techniques, $^{16}\text{O}/^{18}\text{O}$, $^{12}\text{C}/^{13}\text{C}$, SILAC (stable isotope labeling with amino acids in cell culture), iTRAQ (isobaric tags for relative and absolute quantitation), ICAT (isotope coded affinity tags) and other tagging technologies have been implemented to label either specific amino acids or the glycan moiety within glycoproteins and glycopeptides. Among the label-free approaches, spectral counting and ion intensity based quantitation methods have been explored. Some of the most recent and intricate developments are summarized in the followings.

For peptide-level labeling, Liu *et al.* have developed an $^{16}\text{O}/^{18}\text{O}$ labeling approach capable of incorporating three heavy ^{18}O in a glycopeptide: two heavy ^{18}O were incorporated into the carboxyl terminus of all peptides during tryptic digestion, and one heavy ^{18}O was incorporated at the *N*-glycosylation site during PNGase F release of the glycan [72]. Quantitation was accomplished by comparing intensity ratios of samples processed in H_2^{16}O and H_2^{18}O ; mass shifts of $\Delta m=6$ Da and $\Delta m=4$ Da were used for selecting peptides for quantitation, and mass shifts of $\Delta m=6$ Da vs. $\Delta m=4$ Da were used as indicators of the presence of an *N*-glycosylation site. Using this strategy, labeled tryptic glycopeptides from serum were enriched via a hydrophilic affinity extraction protocol and analyzed by nano-

LC-MS/MS (LTQ-Orbitrap) and MALDI-TOF/TOF-MS. The method was used for the parallel quantitation of glyco- and non-glyco peptides in serum samples originating from an ovarian cancer patient and a healthy individual. iTRAQ based peptide labeling approaches have been also proven to be successful for quantifying glycoproteins in biological samples. For example, the comparative analysis of glycoproteins from healthy/hepatocellular cancer serum samples was accomplished by processing the samples via tryptic digestion, iTRAQ labeling, PNGase F glycan release, multi lectin affinity chromatography enrichment and nano-LC-Q-TOF-MS/MS [73]. Alternatively, glycoproteins from tear fluid were quantified by using a strategy that involved enrichment via hydrazide chemistry, tryptic removal of non-glycosylated peptides, PNGase F release of N-linked glycopeptides, iTRAQ labeling and 2D-SCX-nano-LC-Q-TOF-MS/MS [49].

For protein-level labeling, the use of ExacTag kits for the 10-plex labeling of Cys residues in serum proteins, followed by AAL/LCA lectin enrichment in fucosylated N-linked glycoproteins, tryptic digestion and nano-LC-MS/MS analysis on an LTQ mass spectrometer, enabled the relative quantitation of fucosylation levels in hepatocellular carcinoma and cirrhosis samples. The results were confirmed by lectin and antibody arrays [74]. An interesting ICAT approach involving the preliminary synthesis of biotinylated/cleavable ICAT reagents, that incorporated recognition elements for cellulases and xylanases, was developed for comparing the relative expression and activity levels of glycosidases in proteomic extracts [75]. Somewhat less common, Ueda *et al.* have used the $^{12}\text{C}_6$ -/ $^{13}\text{C}_6$ -NBS (2-nitrobenzensulfonyl) stable isotope labeling kit (labels Trp residues) for serum samples that were enriched on LCA lectin columns in N-linked glycoproteins with α 1,6-fucosylation on the Asn-linked GlcNAc. Fucose-glycoprotein enriched and labeled samples were digested with trypsin, enriched in Trp containing peptides, and analyzed by 2D-SCX-LC with spotting on a target plate for MALDI-MS analysis. Changes in α 1,6-fucosylation for selected glycoproteins were confirmed by immunoprecipitation and lectin blotting [76].

For metabolic labeling, a variation of SILAC was developed for facilitating glycoproteomic quantitation and the interpretation of glycan fragmentation. The method was coined IDAWG (isotopic detection of aminosugars with glutamine), and involved the use of ^{15}N -Gln for metabolic labeling of cellular aminosugars by the incorporation of the ^{15}N -labeled Gln into N-linked and O-linked glycans via GlcNAc, GalNAc and sialic acid residue biosynthesis. This type of labeling facilitates MS quantitation by increasing the mass of the labeled glycan by 1 Da per aminosugar incorporation [77].

The labeling of the glycan moieties was typically accomplished by permethylation. For example, released glycans can be permethylated with $^{12}\text{C}/^{13}\text{C}$ methyl iodide and analyzed by MALDI-TOF-MS for differential glycan profiling [45]. Alternatively, a method of quantitation by isobaric labeling (QUIBL) via permethylation with $^{13}\text{CH}_3\text{I}$ and $^{12}\text{CH}_2\text{DI}$ that generates a pair of isobaric tags with the same nominal mass, but with a small mass difference of 0.002922 Da, was demonstrated [78]. Multiple methylation sites on a glycan result in a total mass difference that enables quantitation on mass spectrometers that display a resolution of $\sim 30,000$. The method also enables the quantitation of isomeric glycans in the MS^n mode, but overall has the disadvantage of being applicable only with the use of high-resolution mass spectrometers. Other methods for differential quantitation of carbohydrates were developed by introducing stable isotope tags to the reducing end of the glycans by reductive amination protocols. The tags were synthesized to include a UV absorbing moiety (2-anthranilic acid) for enabling UV detection, and a D_0 -to- D_4 deuterated hydrophobic moiety for generating a suitable MS fragmentation pattern of glycans (stable, amide bond containing labels, that do not fragment under the same conditions as the glycan does) [79, 80].

In the case of label-free approaches, spectral counting was used for the exploration of differential expression of glycoproteins in breast cancer cell extracts enriched on ConA and WGA lectin columns [81], and intensity-based measurements were used to evaluate saccharide compositions and site-occupancy in polyclonal antibody-purified *O*-linked glycoproteins from serum [82]. For the intensity measurement method, it was shown that glycopeptide ion intensities can be used for assessing attachment site (ESI-ETD MS² scans), occupancy (ESI-ETD MS² scans), and saccharide compositions (MALDI or ESI-MS scans), and that the method is robust enough to be suitable for a detailed analysis of abnormalities in mucin-type *O*-glycan structures.

4. Bioinformatics tools for the interpretation of glycoproteomic data

With the recent developments in analytical technologies, especially MS, the rapid increase in the amount of experimental glycoproteomic data places a challenge on researchers that seek fast analysis and meaningful data interpretation. Due to the structural complexity of carbohydrates, most of the bioinformatics tools and applications developed for the analysis of DNA and protein sequences cannot be directly applied to carbohydrates. Glycomics data processing can be very time-consuming for the human analyst and requires support in the form of data repositories and software tools [83]. The development of automated methods for the interpretation, structural assignment and quantification of high-throughput glycomics data is of high priority, and while progress has been achieved, much work still awaits to be done [84]. Several paths of research currently in progress in glycobioinformatics are summarized in two excellent reviews [84, 85]. The efforts are focused on the development of: (a) data repositories with an explicit description of glycan structures; (b) bioinformatics tools for determining glycan structure, identifying glycan-protein binding sites and structural motifs to which glycans may be attached; (c) bioinformatics tools for predicting structural features of complex carbohydrates, carbohydrate-protein interactions and structure-function relationships; and (d) bioinformatics tools for the identification of enzymes that build or modify oligosaccharides and lectins that bind specific sugar epitopes. With respect to MS detection, additional tools are needed for the interpretation of carbohydrate and glycopeptide fragmentation tandem mass spectra.

The availability of databases and tools to store, retrieve and analyze glycomics data in an efficient way is of fundamental importance to glycobiology. To this end, there has been a substantial increase in both the development and use of databases. Starting with the Complex Carbohydrate Structure Database (CCSD; often referred as CarbBank), several initiatives have resulted in the further development of independent data repositories with partially overlapping goals (Table 1) [84, 85]. However, these glycomics resources rarely provide communication or data exchange mechanisms that would permit taking full advantage of the stored data. For each database, different file formats, encoding schemes for residues and sequence topology were designed. Therefore, while such experimental data repositories exist, it is virtually impossible to obtain an overview of all deposited structures or to compare or combine the contents across various databases [86, 87]. In response, the Complex Carbohydrate Research Center (CCRC) at the University of Georgia developed the GLYDE-II XML [88] as a standard format for data exchange. More recently, the German Cancer Research Center Heidelberg developed *GlycomeDB* (www.glycome-db.org) to integrate the seven major open access carbohydrate structure databases [89]. The existence of a centralized glycan structure database will substantially increase the ability to annotate and cross-reference local data with a variety of external bioinformatics resources. Another important development for glycoproteomic research is the integration of glyco-related pathways into the Kyoto Encyclopedia of Genes and Genomes (KEGG), which has led to the development of KEGG GLYCAN [90, 91]. This particular KEGG module establishes the connection between glycan structures and the current knowledge of enzymatic reactions

that build the glycans. Additionally, the KEGG research group made significant progress in applying algorithms to the tree-like structures of glycans for comparison and alignment, to develop similarity scores, and to establish a global view of all glycans belonging to related pathways [90, 91].

In addition to data repositories, several software tools have been developed to aid the arduous process of data analysis, glycan structure annotation, and glycosylation prediction [84, 85]. Currently, automatic and reliable data processing represents a major bottleneck for the practical implementation of high-throughput glycomics projects, and, not surprisingly, the development of algorithms and tools to interpret MS data constitutes one of the most active areas of software development in glycomics. One of the most promising applications for glycomic mass spectrometric data interpretation is *Cartoonist*, an algorithm designed to annotate *N*-linked glycan spectra with the help of an archetype based glycan library [92]. *Cartoonist* has been adopted by the Consortium for the Functional Glycomics (CFG) for the annotation of MS data, mostly of that generated by MALDI-MS [93–95]. Other recently available applications such as *GlycoWorkbench* [96] and *Glyco-Peakfinder* [97], developed by the EuroCarbDB initiative, and *GlycoPepDB* [98] at University of Kansas, are among the closest to a complete MS glycan analysis tool. *GlycoWorkbench*, a semiautomatic annotation tool for mass spectrometric data interpretation and rapid drawing of glycan structures, is compatible with a broad spectrum of MS experimental techniques and provides an integrated environment with easy to use graphical interface. On the other hand, *Glyco-Peakfinder*, aids the *de novo* interpretation of glycan-derived MS signals regardless of the MS source (ESI, MALDI, *etc.*) and fragmentation technique. The capabilities of *GlycoWorkbench* and *Glyco-Peakfinder* can be interfaced with carbohydrate structure databases and other software applications such *Cartoonist*. Future developments of this integration may result in a complete workflow for glycomics, with utility spanning from raw data processing to complete annotation of mass spectra. A variety of additional informatics tools have been developed for glycobiology research, including: (a) theoretical and experimental generation of glycan MS and MSⁿ libraries [94, 99–102]; (b) identification of glycopeptides from protein digest MS data [103]; (c) glycan structure annotation of glycopeptides from tandem MS [92, 93, 104]; and (d) glycosylation prediction programs, such as *NetNGly* [105], *NetOGly* [106] and *NetCGly* [107], for *N*-glycosylation, *O*-glycosylation and *C*-glycosylation, respectively, among others [108–110]. A set of glyco/glycoproteomic informatic tools has been made available through the ExPASy Proteomics server (<http://expasy.org/tools/#proteome>). Further discussion of the available tools is beyond the scope of this review, and for a detailed overview of the current status and needs of informatics in glycomics research the reader is referred to references [84, 85, 111–113].

5. Microfluidic platforms with MS detection for the analysis of glycoproteins

While research on glycoproteomic topics involving microfluidic instrumentation did not reach yet the intricacy of the advanced workflows that have been implemented on conventional analysis platforms, over the years, substantial progress has been achieved [14]. The use of miniaturized devices is advantageous for analytical and economical reasons. The integration of microfluidics with MS detection reduces sample and reagent consumption, eliminates the possibility of cross contamination, improves detection limits and reproducibility, and facilitates automation and high-throughput analysis. In addition, microfluidic devices possess the potential to integrate multiple sample processing steps, a critical feature for the success of a complex analysis approach such as needed in glycoproteomics (i.e., sample isolation, enrichment, purification, derivatization and separation). Ultimately, the applicability of such platforms for prospective biomarker discovery applications, further spurs the interest in their development both in the academic

and industrial setting. The interfacing of microfluidic devices to ESI and MALDI-MS detection has been extensively reviewed [26, 114–116], and the challenges and accomplishments of microfluidic-MS technologies in the area of glycoprotein analysis have been described [12, 14, 117]. Liquid sheath, liquid junction, nano-ESI and MALDI sources have been developed for a variety of glass, polymeric or silicon microfabricated chips. The intent of the following section is to provide the reader with an update on the advancement of microfluidics-MS in glycoproteomics research and on the most recent high-throughput chip-based applications. Commercial and custom fabricated devices with integrated or inserted ESI emitters, as well MALDI-chip and high-throughput array technologies will be discussed. The essential features that emerged from these applications, i.e., the complexity of processes that can be handled on the chip, the detection limits, the short analysis times and improved throughput, clearly establish microfluidics-MS as a promising technology in the challenging area of the glycoproteome.

5.1. Microfluidic devices with ESI-MS detection-commercial platforms

Two commercial platforms have gained, so far, popularity for microfluidic applications: the NanoMate platform (Advion Biosciences, Ithaca, NY, USA), that has been often used for fast infusion ESI-MS explorations from small sample volumes, and the ChipCube (Agilent Technologies, Inc. Palo Alto, CA, USA), that has shown applicability to a broader range of sample processing operations. The NanoMate™ 100, a versatile microfluidic platform capable of high-throughput screening and automated analysis, has been frequently used in the analysis of glycoproteins (Figure 2A). In this system, the microchip functions as a nano-ESI ion source, transferring samples at 50–100 nL/min for mass spectrometric analysis. The microchip is a 10×10 array of nano-ESI nozzles etched into the planar surface of a silicon wafer. The system enables the aspiration and delivery of samples ($<5 \mu\text{L}$) from a 96 well plate to the microchip. Electrospray, at an optimal flow rate of $\sim 100 \text{ nL/min}$, is initiated by the application of voltage and head pressure to the sample-delivery ESI conductive tip on the NanoMate™ platform [117]. To demonstrate the capability of the NanoMate™ 100, the microchip was used in tandem with a linear ion trap mass spectrometer to analyze two standard synthetic glycans using multi-stage fragmentation (MS^5) in $+/-$ ionization mode (Figure 3), while consuming only a small amount of sample ($5 \mu\text{L/min}$; $1 \text{ pmol}/\mu\text{L}$) [32]. In a later study, Zhang *et al.* explored the limits of detection for the NanoMate™ 100 via precursor ion scanning MS on a quadrupole linear ion trap mass spectrometer, discovering that the system could be used to map the *N*-glycosylation site of RNase B and obtain glycan structures from sample concentrations as low as $100 \text{ fmol}/\mu\text{L}$ in analysis times $<5 \text{ min}$ [118]. In a different microchip-MS arrangement, the NanoMate™ 100 was combined with QTOF-MS in negative ion mode to map and sequence glycoconjugates from complex mixtures of oligosaccharides and glycopeptides. Data were collected for just over a min from samples containing *O*-glycosylated sialylated peptides ($\sim 3 \text{ pmol}/\mu\text{L}$) and a native mixture of glycans and *N*- and *O*-glycosylated sialylated peptides ($5 \text{ pmol}/\mu\text{L}$). Interpretable mass spectra could be acquired in less than 30 sec acquisition time [119]. In addition to carbohydrate analysis, the automated NanoMate™ 100 QTOF-MS system was used to characterize a complex ganglioside mixture from human cerebellar tissue via MS and MS/MS analysis. Three major ganglioside species, GM1, GD1, and GT1 (with mono-, di- and tri- sialic acid moieties) were detected from a working ganglioside sample ($15 \text{ pmol}/\mu\text{L}$) in $\sim 1 \text{ min}$ analysis time [120]. To test the compatibility and sensitivity of off-line capillary electrophoresis coupling to MS, the NanoMate™ 100 used in conjunction with a Q-TOF-MS system was used to sample *O*-glycosylated amino acids and peptides ($5 \mu\text{g}/\mu\text{L}$) from urine. The samples were injected in a bench-top CE system and collected in $15 \mu\text{L}$ buffer solution. A limit of detection of less than $1 \text{ pmol}/\mu\text{L}$ in the collected fractions was demonstrated, and informative tandem mass spectra could be generated from the consumption of as little as 12 pmol of an *O*-glycosylated peptide [121].

Another commercially available microfluidic platform applied to glycoproteomics is the Agilent ChipCube LC/MS interface [7, 8]. The ChipCube incorporates valve rotors, trapping and separation columns, flow capillaries and a nanospray emitter on a polyimide chip, to decrease dead volumes and peak dispersion while improving chromatographic performance and ESI-MS analysis. The ChipCube platform can be designed with an assortment of packing materials, such as activated graphitized carbon and reversed phase C18 particles for custom LC applications. In one example, the ChipCube (one enzyme chip, one C8 chip and one HPLC chip) was used to characterize IgG *N*-glycans. Interestingly, analysis times were reduced from hours/days to min when using the ChipCube to perform the following functionalities: glycan release with immobilized PNGase F, removal of deglycosylated proteins, capture of the released glycans, glycan separation on graphitized carbon chromatography columns, and nano-ESI-TOF-MS (Figure 2B) [7]. The ChipCube LC/MS platform was also employed to evaluate the performance of activated graphitized carbon and C18 particles for the separation of tryptic glycopeptides from model glycoproteins [8]. It was found that: (a) the hydrophilic glycopeptides were preferentially retained on the graphitized carbon chip, enabling thus the identification of hydrophilic glycopeptides that were not retained on the more hydrophobic C18 chip; (b) the reproducibility for two model glycopeptides - SRNLTK-GlcNAc₂Man₅ and NVGLNR-GlcNAc₂(Fuc)Man₃(Xyl) - was excellent, i.e., 19.2±0.1 min and 12.1±0.1 min, respectively; and (c) the characterization of the glycosylation site on glycoproteins was feasible from <1 pmol sample [8]. In a comparison of activated graphitized carbon packed HPLC chips (Agilent), a low-capacity “Glycan-Chip” (40 nL enrichment column and 43 × 0.075 mm² i.d. analytical separation column) and a high-capacity HPLC-Chip (160 nL enrichment column, 150 × 0.075 mm² i.d. analytical separation column) were interfaced with TOF-MS to assess the diversity of *N*-linked oligosaccharides found in human serum without derivatization (Figure 4) [122]. Overall, both chips offered good sensitivity and reproducibility, the high-capacity chip showing somewhat less retention time shifts, while the Glycan-Chip showing somewhat better peak area and height reproducibility. The glycan chip enabled the identification of ~200 glycans, including complex, high mannose and hybrid oligosaccharides [122]. The CubeChip interface has been also utilized in the quest to discover potentially new breast cancer biomarkers by identifying permethylated *N*-linked glycans in serum samples. Glycoproteins from human blood serum (cancer and cancer-free patients) were digested with trypsin, and the *N*-glycans were released with PNGase F, reduced to alditols, and subjected to solid phase permethylation and chip-HPLC-MS analysis. Both the trapping and separation columns on the chip were packed with C18 particles. This strategy resulted in the identification of approximately 20 different *N*-linked glycan structures from only 0.25 µL of blood serum. Statistical analysis revealed five glycans with *p*<0.05, of which two were implicated as potential indicators of late-stage breast cancer [123]. The Agilent HPLC chip packed with amide-silica HILIC particles was also used for the isolation and identification of complex oligosaccharides from glycosaminoglycans (GAGs), enabling the analysis of heparin and heparan sulfate oligosaccharides via LC/MS in <25 min [124]. The addition of a makeup flow (MUF) to the amide-HILIC chip increased the reproducibility of the platform, as the MUF facilitated ESI in negative ion mode from highly aqueous solutions (up to 100%) [125].

5.2. Microfluidic devices with ESI-MS detection-custom designs

To further advance the analysis of glycoconjugates on miniaturized platforms, a variety of custom microfluidic designs with ESIMS detection have been explored. A polyimide chip, enclosed with a polyethylene-polyethylene terephthalate laminate, was described by Zamfir *et al.* [126]. The chip microchannels (120 µm × 45 µm) were constructed in the polyimide substrate by plasma etching, had gold-coated electrodes placed at the bottom, and had the channel edges shaped into a nano-ESI tip. The polyimide chip microsprayer, in tandem with

Q-TOF-MS, was used for the mapping, sequencing and structural elucidation of glycoconjugates derived from human body-fluid samples. The sensitivity of the chip was tested with *O*-GalNAc glycosylated amino acids and peptides isolated from human urine. At a stable flow rate of 200 nL/min (1.25 pmol/ μ L), 13 different components were identified from only ~0.16 pmol sample [126]. Using similar experimental conditions, the same polyimide chip microsyringe (Figure 5A) was utilized in conjunction with an FTICR mass spectrometer operated in negative ionization mode for the identification of singly sialylated *O*-glycosylated amino acids and peptides from urine (6–12 pmol/ μ L) [127]. To confirm and complement the data collected with the custom polyimide microsyringe system, the NanoMate 100 platform interfaced with FTICR-MS was used [128].

In addition to infusion chips, more complex microfluidic designs that incorporate several functional elements (CE separation systems, microreactors, MS interfaces), have been pursued, as well. For example, Zheng *et al.* have designed a CE glass chip with a 5.7 cm separation channel that accommodated the insertion of a 35 cm (75 μ m i.d. $\times$ 150 μ m o.d.) capillary for sample transfer to the ESI source of a Q-TOF-MS. A concentric flow of sheath liquid, delivered by a pump, and sheath gas were used to stabilize the spray. Using MS/MS analysis, and only 20 pmol of sample, the platform facilitated the identification of a glycopeptide and its β -elimination product, as well as the identification of the glycosylation site on the peptide [129]. Another sheath flow glass microchip-CE design, incorporating a sample transfer capillary (10 cm long, 50 μ m i.d. $\times$ 195 μ m o.d.), was used for the analysis of glycopeptides and glycoforms of RNase B. Both chip channel and transfer capillary were coated with 3-aminopropyltriethoxysilane, the capillary acting as an extension of the separation channel on the chip. CE separations were performed by applying the separation voltage between a reservoir on the chip and a T-union used for liquid sheath delivery to the ESI source. With a sheath flow of 100–200 nL/min, glycopeptides at low fmol levels were detected on a quadrupole ion trap mass spectrometer [130]. Yet another microfluidic platform, fabricated in glass, incorporated three reaction chambers that were used to examine three sequential glycosyltransfer reactions that mimicked the generation of glycoconjugates in the Golgi apparatus. The reaction chambers were filled with bead-attached glycosyltransferases, and the microchip was interfaced to an LC-ion trap-TOF-MS system to monitor the reaction products (Figure 5B). The results demonstrated that in comparison to batch reactions, the glycosylation reactions were enhanced on the microfluidic platform, and the use of sequential enzymatic reactions enabled for the first time the synthesis of a tetrasaccharide [10].

While not involving MS detection, it is worth describing briefly some microchip CE designs that have been used successfully for the analysis of glycoproteins, and that show great potential for future interfacing to mass spectrometry. For example, Zhuang *et al.* have described a spiral CE microfluidic chip that was used for the electrophoretic separation of *N*-glycans, and that generated separation efficiencies of 400,000–655,000 plates on a 22 cm long channel, at 750 V/cm, in less than 3 min analysis time. The microfluidic system was used to analyze *N*-glycans derived from RNase B, linkage isomers from asialofetuin, and *N*-glycans released from blood [11]. Alternatively, a microchip integrating a monolithic lectin affinity column (PSA-pisum sativum agglutinin) was used in the separation of turkey ovalbumin, chicken ovalbumin and ovomucoid. Only 300 pg of glycoprotein were required for laser-induced fluorescent detection, and the integration of microfluidics and lectin-affinity separation reduced analysis times from 4 h to only 400 sec [31]. The future implementation of polymeric or silica monolithic LC/CEC separation columns [131, 132] on a chip will further expand the functionality and capabilities of microfluidic platforms for glycoproteomic applications.

5.3. Microfabricated devices with MALDI-MS detection

With the goal of promoting high-throughput chip-MS technologies, the development of microfluidic devices with MALDI-MS detection has centered on designing platforms and interfaces that enable off-chip or on-chip MALDI-MS analysis, as well as on advancing the microfabrication of silicon/polymeric MALDI target plates [26, 115]. The implementation of microchip-MALDI-MS technologies for glycoprotein analysis involves the development of boronic acid functionalized gold-coated MALDI target plates or Si wafers. For example, truncated icosahedral Au nanoparticles were sintered to stainless steel MALDI targets, and functionalized with 4-mercaptophenylboronic acid. Functionalization with boronic acid, a compound that forms reversible covalent bonds with cis-diol groups common to Man, Gal and Glu, provided the foundation for glycopeptide enrichment (Figure 6) [133]. Detection from low 2.5×10^{-9} M solutions of horse radish peroxidase (HRP) was achieved. Alternatively, boronic acid functionalized gold-coated Si wafers have been used as MALDI plates to improve sample handling, enrichment and reduce non-specific binding of glycopeptides. The integration of glycopeptide capture/enrichment and purification steps with subsequent MS analysis, resulted in a decrease in sample losses and led to increased sensitivity, enabling the detection of HRP, asialofetuin and bovine serum albumin from ~ 0.4 fmol/ μ L solutions. The enrichment process resulted in ~ 2 orders of magnitude improvement in detection limits over conventional analysis methods [134].

As an alternative to the previously discussed enrichment platforms, MALDI-MS was implemented as a novel detection strategy for lectin microarray chips, as well. Lectin microarray is a high-throughput, sensitive, multiplexed lectin-glycan interaction profiling technology that can be successfully used for the structural elucidation of glycans, the analysis of differential glycan expression, and the detection of glycan alterations. Due to the wealth of information that can be generated, the use of lectin microarrays in diagnostic applications is very promising. For example, MALDI-MS detection was incorporated into a 2D high-throughput lectin array analysis format to perform enzymatic sequencing of glycopeptides. A series of biotinylated lectins (24 μ M) were immobilized to a streptavidin functionalized MALDI target plate. The glycoprotein sample was digested with trypsin, divided in aliquots, and the generated peptides were further subjected to sequential exoglycosidase digestions. Tryptic peptides, and tryptic peptide samples treated with various enzymes or enzyme mixtures, were deposited on every lectin. The nonglycosylated peptides were removed by washing, while glycopeptides immobilized on their specific lectin spot were analyzed by MALDI-MS. Rapid and simultaneous determinations of *N*- and *O*-linked glycan sequences and glycosylation sites could be achieved with this 2D MALDIMS array [135]. In another fairly complex approach, Ito *et al.* were able to identify glycan alterations on alpha-fetoprotein (AFP) from two hepatic cell lines [136]. The authors used a combination of multiplex quantitative PCR for profiling glycogen expression patterns, lectin-microarrays for profiling glycoproteins from cell extracts, and tandem MALDI-MS detection. Lubman *et al.* have completed several studies that used lectin microarrays for the discovery of cell surface glycan expression patterns in human adherent and stem-like GBM cell lines [137], or for the detection of unique protein glycosylation patterns in the serum [138] and plasma [139] of pancreatic and colorectal cancer patients, respectively. The lectin microarrays were used to complement or validate LC-MS/MS results [137, 139], or to act as direct targets subjected to MALDI-MS analysis [138]. The commercially available lectin-coupled ProteinChip array (Jacalin or SNA-coupled chip) was used in tandem with surface enhanced laser desorption ionization (SELDI)-MS to detect glycoproteins with abnormal sialylation patterns, characteristic of lung cancer associated glycosylation disorders. The analysis of serum from cancer and control groups yielded 41 protein peaks showing significant differences ($p < 0.05$) in sialylation pattern [76].

6. Concluding remarks

Glycoproteomics research continues to evolve rapidly due to the successful integration of superior methods of analysis with advanced bioanalytical instrumentation and ultrasensitive MS detection. This review provided a summary of the diverse strategies that have been developed for the detection of glycoproteins, the identification of glycosylation sites, and the structural characterization of *N*- and *O*-linked glycans. While the progress made by mass spectrometry in elucidating the structure of glycoproteins is not achievable by other methods of detection, the full capabilities of this technology need to be further harvested to explore the subtle differences that arise among various protein glycoforms characteristic of healthy or diseased cell states. The use of mass spectrometry for the identification of disease-specific glycoproteins will be invaluable to diagnostics and the development of novel therapeutic drugs.

The incorporation of microfluidic devices into glycoproteomics research, highlighting the potential applicability of this technology in biomarker discovery, has been reviewed, as well. While custom manufactured devices illustrate capabilities for diverse proof-of-concept applications, commercial chip devices advance applications that facilitate high-throughput MS detection and the implementation of rapid, high-performance, standardized analysis methods. The development of array technologies and their integration with microfluidics will create further opportunities for large-scale glycoproteomics research. In the long term, it is expected that the merger of microfluidics with MS detection will contribute extensively to overcoming the setbacks imposed by the inner complexity of glycoproteins and proteomics at large. There is no doubt that the knowledge generated from the complete characterization of glycoproteins will enhance our understanding of the myriad of cellular processes, facilitating promising new developments towards disease diagnostics and management.

Acknowledgments

This work was supported in part by grants from NSF (Career BES-0448840) and NIH/NCI (R21CA126669).

References

- [1]. Harvey DJ. Expert Review of Proteomics. 2005; 2:87–101. [PubMed: 15966855]
- [2]. Zaia J. Chemistry & Biology. 2008; 15:881–892. [PubMed: 18804025]
- [3]. Dods ED, Seipert RR, Clowers BH, German JB, Lebrilla CB. J Proteome Res. 2009; 8:502–512. [PubMed: 19072223]
- [4]. Abbott KL, Lim JM, Wells L, Benigno BB, et al. Proteomics. 2010; 10:470–481. [PubMed: 19953551]
- [5]. Chen R, Jiang X, Sun D, Han G, et al. J Proteome Res. 2009; 8:651–661. [PubMed: 19159218]
- [6]. Tanaka K, Fujii Y, Tokimoto H, Mori Y, et al. Chemistry-an Asian Journal. 2009; 4:574–580.
- [7]. Bynum MA, Yin HF, Felts K, Lee YM, et al. Anal. Chem. 2009; 81:8818–8825. [PubMed: 19807107]
- [8]. Alley WR, Mechref Y, Novotny MV. Rapid Commun. Mass Spectrom. 2009; 23:495–505. [PubMed: 19145579]
- [9]. Yamamoto S, Hirakawa S, Suzuki S. Anal. Chem. 2008; 80:8224–8230. [PubMed: 18841941]
- [10]. Ono Y, Kitajima M, Daikoku S, Shiroya T, et al. Lab on a Chip. 2008; 8:2168–2173. [PubMed: 19023483]
- [11]. Zhuang Z, Starkey JA, Mechref Y, Novotny MV, Jacobson SC. Anal. Chem. 2007; 79:7170–7175. [PubMed: 17685584]
- [12]. Mechref Y, Novotny MV. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences. 2006; 841:65–78.
- [13]. Bedair M, Oleschuk RD. Analyst. 2006; 131:1316–1321. [PubMed: 17124539]

- [14]. Bindila L, Peter-Katalinic J. Mass Spectrometry Reviews. 2009; 28:223–253. [PubMed: 19145581]
- [15]. Mao XL, Wang K, Du YG, Lin BC. Electrophoresis. 2003; 24:3273–3278. [PubMed: 14518056]
- [16]. Vanderschaeghe D, Festjens N, Delanghe J, Callewaert N. Biol. Chem. 391:149–161. [PubMed: 20128687]
- [17]. Jacobson SC, Culbertson CT, Daler JE, Ramsey JM. Anal. Chem. 1998; 70:3476–3480.
- [18]. Jacobson SC, Hergenroder R, Koutny LB, Ramsey JM. Anal. Chem. 1994; 66:1114–1118.
- [19]. Xie J, Miao Y, Shih J, Tai YC, Lee TD. Anal Chem. 2005; 77:6947–6953. [PubMed: 16255594]
- [20]. Yin H, Killeen K, Brennen R, Sobek D, et al. Anal Chem. 2005; 77:527–533. [PubMed: 15649049]
- [21]. Lazar IM, Trisiripisal P, Sarvaiya HA. Anal Chem. 2006; 78:5513–5524. [PubMed: 16878890]
- [22]. Kapron JT, Pace E, Van Pelt CK, Henion J. Rapid Commun Mass Spectrom. 2003; 17:2019–2026. [PubMed: 12955729]
- [23]. Chen CJ, Li FA, Her GR. Electrophoresis. 2008; 29:1997–2003. [PubMed: 18409161]
- [24]. Krishnamoorthy G, Carlen ET, Bomer JG, Wijnperle D, et al. Lab Chip. 10:986–990. [PubMed: 20358104]
- [25]. Javanmard M, Talasaz AH, Nemat-Gorgani M, Pease F, et al. Lab Chip. 2009; 9:1429–1434. [PubMed: 19417910]
- [26]. Sikanen T, Franssila S, Kauppila TJ, Kostianen R, et al. Mass Spectrom Rev. 2010; 29:351–391. [PubMed: 19514079]
- [27]. Gustafsson M, Hirschberg D, Palmberg C, Jornvall H, Bergman T. Anal Chem. 2004; 76:345–350. [PubMed: 14719881]
- [28]. Emrich CA, Tian H, Medintz IL, Mathies RA. Anal Chem. 2002; 74:5076–5083. [PubMed: 12380833]
- [29]. Aborn JH, El-Difrawy SA, Novotny M, Gismondi EA, et al. Lab Chip. 2005; 5:669–674. [PubMed: 15915260]
- [30]. Suzuki S, Shimotsu N, Honda S, Arai A, Nakanishi H. Electrophoresis. 2001; 22:4023–4031. [PubMed: 11700736]
- [31]. Mao XL, Luo Y, Dai ZP, Wang KY, et al. Anal. Chem. 2004; 76:6941–6947. [PubMed: 15571345]
- [32]. Zhang S, Chelius D. J Biomol Tech. 2004; 15:120–133. [PubMed: 15190085]
- [33]. Dai Z, Zhou J, Qiu SJ, Liu YK, Fan J. Electrophoresis. 2009; 30:2957–2966. [PubMed: 19711376]
- [34]. Wu AM, Lisowska E, Duk M, Yang Z. Glycoconj J. 2009; 26:899–913. [PubMed: 18368479]
- [35]. Yang Z, Harris LE, Palmer-Toy DE, Hancock WS. Clin Chem. 2006; 52:1897–1905. [PubMed: 16916992]
- [36]. Kullolli M, Hancock WS, Hincapie M. Anal Chem. 82:115–120. [PubMed: 19957969]
- [37]. Madera M, Mechref Y, Klouckova I, Novotny MV. J Chromatogr B Analyt Technol Biomed Life Sci. 2007; 845:121–137.
- [38]. Arcinas A, Yen TY, Kebebew E, Macher BA. J Proteome Res. 2009; 8:3958–3968. [PubMed: 19530676]
- [39]. Zhang H, Li XJ, Martin DB, Aebersold R. Nat Biotechnol. 2003; 21:660–666. [PubMed: 12754519]
- [40]. Arnold JN, Saldova R, Hamid UM, Rudd PM. Proteomics. 2008; 8:3284–3293. [PubMed: 18646009]
- [41]. Schwientek T, Mandel U, Roth U, Muller S, Hanisch FG. Proteomics. 2007; 7:3264–3277. [PubMed: 17708590]
- [42]. Xu Z, Zhou X, Lu H, Wu N, et al. Proteomics. 2007; 7:2358–2370. [PubMed: 17623300]
- [43]. Larsen MR, Hojrup P, Roepstorff P. Mol Cell Proteomics. 2005; 4:107–119. [PubMed: 15561728]
- [44]. Wu SL, Huhmer AF, Hao Z, Karger BL. J Proteome Res. 2007; 6:4230–4244. [PubMed: 17900180]

- [45]. Alvarez-Manilla G, Warren NL, Atwood J 3rd, Orlando R, et al. J Proteome Res. 2010; 9:2062–2075. [PubMed: 19545112]
- [46]. Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. J Proteome Res. 2004; 3:556–566. [PubMed: 15253437]
- [47]. Omaetxebarria MJ, Hagglund P, Elortza F, Hooper NM, et al. Anal Chem. 2006; 78:3335–3341. [PubMed: 16689534]
- [48]. Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H. Nat Protoc. 2007; 2:334–339. [PubMed: 17406594]
- [49]. Zhou L, Beuerman R, Chew AP, Koh SK, et al. J Proteome Res. 2009
- [50]. Abbott KL, Aoki K, Lim JM, Porterfield M, et al. J Proteome Res. 2008; 7:1470–1480. [PubMed: 18271524]
- [51]. Perdivara I, Petrovich R, Allinquant B, Deterding LJ, et al. J Proteome Res. 2009; 8:631–642. [PubMed: 19093876]
- [52]. Goetz JA, Novotny MV, Mechref Y. Anal Chem. 2009; 81:9546–9552. [PubMed: 19874002]
- [53]. Hanisch FG, Teitz S, Schwientek T, Muller S. Proteomics. 2009; 9:710–719. [PubMed: 19132687]
- [54]. Williams TI, Saggese DA, Muddiman DC. J Proteome Res. 2008; 7:2562–2568. [PubMed: 18422354]
- [55]. Alley WR, Novotny MV. J Proteome Res. 2010; 9:10.
- [56]. Hagglund P, Matthiesen R, Elortza F, Hojrup P, et al. J Proteome Res. 2007; 6:3021–3031. [PubMed: 17636988]
- [57]. Lim JM, Aoki K, Angel P, Garrison D, et al. J Proteome Res. 2009; 8:673–680. [PubMed: 19072240]
- [58]. Knezevic A, Polasek O, Gornik O, Rudan I, et al. J Proteome Res. 2009; 8:694–701. [PubMed: 19035662]
- [59]. Ciucanu I, Kerek F. Carbohydrate Research. 1984; 131:209–217.
- [60]. Kang P, Mechref Y, Klouckova I, Novotny MV. Rapid Commun Mass Spectrom. 2005; 19:3421–3428. [PubMed: 16252310]
- [61]. Lazar IM, Lazar AC, Cortes DF, Kabulski JL. Electrophoresis. 2011; 32 in press.
- [62]. Zhou H, Hou W, Denis NJ, Vasilescu J, et al. J Proteome Res. 2009; 8:556–566. [PubMed: 19067587]
- [63]. Nandi A, Sprung R, Barma DK, Zhao Y, et al. Anal Chem. 2006; 78:452–458. [PubMed: 16408927]
- [64]. Pang PC, Tissot B, Drobnis EZ, Morris HR, et al. J Proteome Res. 2009; 8:4906–4915. [PubMed: 19606896]
- [65]. Goetz JA, Mechref Y, Kang P, Jeng MH, Novotny MV. Glycoconj J. 2009; 26:117–131. [PubMed: 18752066]
- [66]. Goldman R, Ressom HW, Varghese RS, Goldman L, et al. Clin Cancer Res. 2009; 15:1808–1813. [PubMed: 19223512]
- [67]. Kyselova Z, Mechref Y, Al Bataineh MM, Dobrolecki LE, et al. J Proteome Res. 2007; 6:1822–1832. [PubMed: 17432893]
- [68]. Tang Z, Varghese RS, Bekesova S, Loffredo CA, et al. J Proteome Res. 2010; 9:104–112. [PubMed: 19764807]
- [69]. Mechref Y, Hussein A, Bekesova S, Pungpapong V, et al. J Proteome Res. 2009; 8:2656–2666. [PubMed: 19441788]
- [70]. Lin SY, Chen YY, Fan YY, Lin CW, et al. J Proteome Res. 2008; 7:3293–3303. [PubMed: 18549263]
- [71]. Ghesquiere B, Buyl L, Demol H, Van Damme J, et al. J Proteome Res. 2007; 6:4304–4312. [PubMed: 17918875]
- [72]. Liu Z, Cao J, He Y, Qiao L, et al. J Proteome Res. 2010; 9:227–236. [PubMed: 19921957]
- [73]. Lee HJ, Na K, Choi EY, Kim KS, et al. J Proteome Res. 2010; 9:308–318. [PubMed: 19899825]
- [74]. Liu Y, He J, Li C, Benitez R, et al. J Proteome Res. 2010; 9:798–805. [PubMed: 19961239]

- [75]. Hekmat O, He S, Warren RA, Withers SG. J Proteome Res. 2008; 7:3282–3292. [PubMed: 18563928]
- [76]. Ueda K, Fukase Y, Katagiri T, Ishikawa N, et al. Proteomics. 2009; 9:2182–2192. [PubMed: 19322776]
- [77]. Orlando R, Lim JM, Atwood JA 3rd, Angel PM, et al. J Proteome Res. 2009; 8:3816–3823. [PubMed: 19449840]
- [78]. Atwood JA 3rd, Cheng L, Alvarez-Manilla G, Warren NL, et al. J Proteome Res. 2008; 7:367–374. [PubMed: 18047270]
- [79]. Bowman MJ, Zaia J. Anal Chem. 2007; 79:5777–5784. [PubMed: 17605469]
- [80]. Bowman MJ, Zaia J. Anal Chem. 2010; 82:3023–3031. [PubMed: 20230064]
- [81]. Wang Y, Ao X, Vuong H, Konanur M, et al. J Proteome Res. 2008; 7:4313–4325. [PubMed: 18729497]
- [82]. Wada Y, Tajiri M, Ohshima S. J Proteome Res. 2010; 9:1367–1373. [PubMed: 20104905]
- [83]. North SJ, Hitchen PG, Haslam SM, Dell A. Current Opinion in Structural Biology. 2009; 19:498–506. [PubMed: 19577919]
- [84]. Aoki-Kinoshita KF. PLoS Comput Biol. 2008; 4:e1000075. [PubMed: 18516240]
- [85]. von der Lieth CW, Lutteke T, Frank M. Biochim Biophys Acta. 2006; 1760:568–577. [PubMed: 16459020]
- [86]. Taniguchi N, Paulson JC. Proteomics. 2007; 7:1360–1363. [PubMed: 17436269]
- [87]. Packer NH, von der Lieth CW, Aoki-Kinoshita KF, Lebrilla CB, et al. Proteomics. 2008; 8:8–20. [PubMed: 18095367]
- [88]. Sahoo SS, Thomas C, Sheth A, Henson C, York WS. Carbohydr Res. 2005; 340:2802–2807. [PubMed: 16242678]
- [89]. Ranzinger R, Herget S, Wetter T, von der Lieth CW. BMC Bioinformatics. 2008; 9:384. [PubMed: 18803830]
- [90]. Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, et al. Glycobiology. 2006; 16:63R–70R.
- [91]. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, et al. Nucleic Acids Res. 2006; 34:D354–357. [PubMed: 16381885]
- [92]. Goldberg D, Bern M, North SJ, Haslam SM, Dell A. Bioinformatics. 2009; 25:365–371. [PubMed: 19073587]
- [93]. Goldberg D, Sutton-Smith M, Paulson J, Dell A. Proteomics. 2005; 5:865–875. [PubMed: 15693066]
- [94]. Kronewitter SR, An HJ, de Leoz ML, Lebrilla CB, et al. Proteomics. 2009; 9:2986–2994. [PubMed: 19452454]
- [95]. Babu P, North SJ, Jang-Lee J, Chalabi S, et al. Glycoconj J. 2009; 26:975–986. [PubMed: 18587645]
- [96]. Ceroni A, Maass K, Geyer H, Geyer R, et al. J Proteome Res. 2008; 7:1650–1659. [PubMed: 18311910]
- [97]. Maass K, Ranzinger R, Geyer H, von der Lieth CW, Geyer R. Proteomics. 2007; 7:4435–4444. [PubMed: 18072204]
- [98]. Go EP, Rebecchi KR, Dalpathado DS, Bandu ML, et al. Anal Chem. 2007; 79:1708–1713. [PubMed: 17297977]
- [99]. Meitei NS, Banerjee S. Proteomics. 2007; 7:2530–2540. [PubMed: 17623275]
- [100]. Lohmann KK, von der Lieth CW. Proteomics. 2003; 3:2028–2035. [PubMed: 14625865]
- [101]. Kameyama A, Kikuchi N, Nakaya S, Ito H, et al. Anal Chem. 2005; 77:4719–4725. [PubMed: 16053281]
- [102]. Kameyama A, Nakaya S, Ito H, Kikuchi N, et al. J Proteome Res. 2006; 5:808–814. [PubMed: 16602687]
- [103]. Deshpande N, Jensen PH, Packer NH, Kolarich D. J Proteome Res. 9:1063–1075. [PubMed: 20030399]
- [104]. Ren JM, Rejtar T, Li L, Karger BL. J Proteome Res. 2007; 6:3162–3173. [PubMed: 17625816]
- [105]. Gupta R, Brunak S. Pac Symp Biocomput. 2002:310–322. [PubMed: 11928486]

- [106]. Hansen JE, Lund O, Tolstrup N, Gooley AA, et al. *Glycoconj J*. 1998; 15:115–130. [PubMed: 9557871]
- [107]. Julenius K. *Glycobiology*. 2007; 17:868–876. [PubMed: 17494086]
- [108]. Eisenhaber B, Bork P, Yuan Y, Löffler G, Eisenhaber F. *Trends Biochem Sci*. 2000; 25:340–341. [PubMed: 10871885]
- [109]. Bohne-Lang A, von der Lieth CW. *Nucleic Acids Res*. 2005; 33:W214–219. [PubMed: 15980456]
- [110]. Fankhauser N, Maser P. *Bioinformatics*. 2005; 21:1846–1852. [PubMed: 15691858]
- [111]. Lutteke T. *Chembiochem*. 2008; 9:2155–2160. [PubMed: 18693281]
- [112]. Karlsson, NG.; Packer, NH. *Bioinformatics for Glycobiology and Glycomics: An Introduction*. von der Lieth, CW.; Luetteke, T.; Frank, M., editors. Wiley; 2009. p. 223-256.
- [113]. Maass, K.; Ceroni, A. *Bioinformatics for Glycobiology and Glycomics: An Introduction*. von der Lieth, CW.; Luetteke, T.; Frank, M., editors. Wiley; 2009. p. 195-202.
- [114]. Zamfir AD. *J Chromatogr A*. 2007; 1159:2–13. [PubMed: 17428492]
- [115]. Lazar, IM. *Handbook of Capillary and Microchip Electrophoresis and Associated Microtechniques*. Landers, JP., editor. CRC Press; 2007. p. 1459-1506.
- [116]. Lazar IM, Grym J, Foret F. *Mass Spectrom Rev*. 2006; 25:573–594. [PubMed: 16508917]
- [117]. Zamfir AD, Bindila L, Lion N, Allen M, et al. *Electrophoresis*. 2005; 26:3650–3673. [PubMed: 16152660]
- [118]. Zhang S, Williamson BL. *J Biomol Tech*. 2005; 16:209–219. [PubMed: 16461944]
- [119]. Zamfir A, Vakhrushev S, Sterling A, Niebel HJ, et al. *Anal Chem*. 2004; 76:2046–2054. [PubMed: 15053670]
- [120]. Zamfir A, Vukelic Z, Bindila L, Peter-Katalinic J, et al. *J Am Soc Mass Spectrom*. 2004; 15:1649–1657. [PubMed: 15519233]
- [121]. Bindila L, Almeida R, Sterling A, Allen M, et al. *J. Mass Spectrom*. 2004; 39:1190–1201.
- [122]. Chu CS, Ninonuevo MR, Clowers BH, Perkins PD, et al. *Proteomics*. 2009; 9:1939–1951. [PubMed: 19288519]
- [123]. Alley WR Jr, Madera M, Mechref Y, Novotny MV. *Anal Chem*. 2010; 82:5095–5106. [PubMed: 20491449]
- [124]. Staples GO, Bowman MJ, Costello CE, Hitchcock AM, et al. *Proteomics*. 2009; 9:686–695. [PubMed: 19137549]
- [125]. Staples GO, Naimy H, Yin H, Kileen K, et al. *Anal Chem*. 2010; 82:516–522. [PubMed: 20000724]
- [126]. Zamfir AD, Lion N, Vukelic Z, Bindila L, et al. *Lab Chip*. 2005; 5:298–307. [PubMed: 15726206]
- [127]. Bindila L, Froesch M, Lion N, Vukelic Z, et al. *Rapid Commun Mass Spectrom*. 2004; 18:2913–2920. [PubMed: 15529419]
- [128]. Froesch M, Bindila LM, Baykut G, Allen M, et al. *Rapid Commun Mass Spectrom*. 2004; 18:3084–3092. [PubMed: 15562445]
- [129]. Zheng Y, Li H, Guo Z, Lin JM, Cai Z. *Electrophoresis*. 2007; 28:1305–1311. [PubMed: 17377943]
- [130]. Mao X, Chu IK, Lin B. *Electrophoresis*. 2006; 27:5059–5067. [PubMed: 17117389]
- [131]. Zhong H, El Rassi Z. *J Sep Sci*. 2009; 32:1642–1653. [PubMed: 19391172]
- [132]. Zhong H, El Rassi Z. *J Sep Sci*. 2009; 32:10–20. [PubMed: 19058161]
- [133]. Tang J, Liu Y, Qi D, Yao G, et al. *Proteomics*. 2009; 9:5046–5055. [PubMed: 19834891]
- [134]. Xu Y, Zhang L, Lu H, Yang P. *Proteomics*. 2010; 10:1079–1086. [PubMed: 20166148]
- [135]. Tzur Y, Markovich A, Lichtenstein RG. *J Proteome Res*. 2008; 7:1188–1198. [PubMed: 18205330]
- [136]. Ito H, Kuno A, Sawaki H, Sogabe M, et al. *Journal of Proteome Research*. 2009; 8:1358–1367. [PubMed: 19178301]
- [137]. He J, Liu Y, Xie X, Zhu T, et al. *J Proteome Res*. 2010; 9:2565–2572. [PubMed: 20235609]

- [138]. Li C, Simeone DM, Brenner DE, Anderson MA, et al. *Journal of Proteome Research*. 2009; 8:483–492. [PubMed: 19072160]
- [139]. Qiu YH, Patwa TH, Xu L, Shedden K, et al. *Journal of Proteome Research*. 2008; 7:1693–1703. [PubMed: 18311904]
- [140]. Doubet S, Albersheim P. *Glycobiology*. 1992; 2:505. [PubMed: 1472756]
- [141]. Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. *Nat Methods*. 2005; 2:817–824. [PubMed: 16278650]
- [142]. Lutteke T, Bohne-Lang A, Loss A, Goetz T, et al. *Glycobiology*. 2006; 16:71R–81R.
- [143]. Cooper CA, Joshi HJ, Harrison MJ, Wilkins MR, Packer NH. *Nucleic Acids Res*. 2003; 31:511–513. [PubMed: 12520065]
- [144]. Cooper CA, Harrison MJ, Wilkins MR, Packer NH. *Nucleic Acids Res*. 2001; 29:332–335. [PubMed: 11125129]

Figure 1.
Schematic diagram of a generic approach for glycoprotein analysis by mass spectrometry.

Figure 2. Commercial microfluidic designs with ESI-MS detection used for the analysis of glycoproteins. **(A)** The NanoMate™ 100: silicon nozzle sprayer and analysis platform (reprinted with permission from reference [117]); **(B)**. The ChipCube: representation of three chips used for enzymatic deglycosylation, protein removal, glycan concentration and glycan separation (reprinted with permission from reference [7]).

Figure 3. Glycopeptide CID-MS³ spectrum generated from an infusion experiment with the NanoMate interfaced to a linear ion trap (reprinted with permission from reference [32]).

Figure 4.

Base peak chromatogram illustrating the HPLC-chip separation of neutral (top), neutral and anionic (middle) and anionic (bottom) oligosaccharides from human serum (reprinted with permission from reference [122]).

Figure 6.

Illustration of the procedure used for glycopeptide enrichment on MALDI target plates via boronic acid functionalization (reprinted with permission from reference [133]).

Table 1

Major carbohydrate structure databases.

Database Name	Description	URL	Reference
Bacterial Carbohydrate Structure DB	Structural, bibliographic, taxonomic and related information on nearly all known bacterial glycan structures published up to 2007.	http://www.glyco.ac.ru/bcsdb3/	
CCSD CarbBank	Containing 50000 records relating to approximately 15000 different carbohydrate structures, CCSD is still one of the largest repositories. Although an online version is available, it is not longer updated.	http://www.boc.chem.uu.nl/sugabase/carbbank.html	[140]
CFG Glycan Database	Developed by CFG. It contains N- and O-linked glycans from CarbBank, Glycominds, and glycans experimentally identified or synthesized by the CFG Core.	http://www.functionalglycomics.org/static/index.shtml	[141]
GlycomeDB	Meta-database to search various databases via one interface. Access to more than 33000 unique sequences.	http://www.glycome-db.org/	[89]
Glycosciences.de	Database of the GLYCOSCIENCES.de portal based on LInear Notation for Unique description of Carbohydrate Sequences (LINUCS) notation.	http://glycosciences.de/	[142]
GlycoSuiteDB	Available through ExPASy. This curated and annotated glycan database contains 9436 entries, sourced from 864 references, representing 245 species.	http://glycosuitedb.expasy.org/glycosuite/glycodb	[143,144]
KEGG Glycan	Database of the KEGG Glycan portal, contains experimentally determined glycan structures.	http://www.genome.jp/ligand/kcam/	[90]