New Technologies for Glycomic Analysis: Toward a Systematic Understanding of the Glycome

John F. Rakus and Lara K. Mahal

Department of Chemistry, New York University, New York, New York 10003; email: jr2720@nyu.edu, lkmahal@nyu.edu

Annu. Rev. Anal. Chem. 2011. 4:367–92

First published online as a Review in Advance on April 1, 2011

The *Annual Review of Analytical Chemistry* is online at anchem.annualreviews.org

This article's doi: 10.1146/annurev-anchem-061010-113951

1936-1327/11/0719-0367$20.00

Keywords

carbohydrate, lectin microarray, mass spectrometry, glycosylation, glycan, oligosaccharide

Abstract

Carbohydrates are the most difficult class of biological molecules to study by high-throughput methods owing to the chemical similarities between the constituent monosaccharide building blocks, template-less biosynthesis, and the lack of clearly identifiable consensus sequences for the glycan modification of cohorts of glycoproteins. These molecules are crucial for a wide variety of cellular processes ranging from cell-cell communication to immunity, and they are altered in disease states such as cancer and inflammation. Thus, there has been a dedicated effort to develop glycan analysis into a high-throughput analytical field termed glycomics. Herein we highlight major advances in applying separation, mass spectrometry, and microarray methods to the fields of glycomics and glycoproteomics. These new analytical techniques are rapidly advancing our understanding of the importance of glycosylation in biology and disease.

1. INTRODUCTION

Glycosylation: the process of enzymatically attaching carbohydrate molecules to protein or lipid substrates

Glycan: a carbohydrate structure that contains more than one monosaccharide unit

Saccharide: a synonym for carbohydrate

Reducing sugar: any sugar in which the cyclic acetal or ketal can open to its corresponding aldehyde or ketone form when in solution

Endoplasmic reticulum: a membrane-bound eukaryotic organelle responsible for the synthesis of many biological macromolecules, most notably proteins

Golgi apparatus: a membrane-bound eukaryotic organelle responsible for protein trafficking and modification

Carbohydrates are a major class of biological macromolecules that are important in many cellular processes, including adhesion, motility, immune response, and pathogen interaction (1–4). Glycosylation, which occurs on both proteins and lipids, is more prevalent than phosphorylation, methylation, or acetylation (5). It is estimated that at least 50% of all proteins are glycosylated (6). Aberrant glycosylation has been implicated in disease, most strikingly in a class of diverse diseases collectively referred to as congenital disorders of glycosylation. These diseases involve genetic mutations in *N*-linked glycosylation pathways that result, in most cases, in some form of reduced glycosylation in a tissue. Symptoms are generally severe and can include seizure events, growth abnormalities, and mild to severe mental retardation (7). Alterations in glycosylation have also been observed in numerous other pathological states, including inflammation and cancer (8). For example, a recent study of messenger RNA expression patterns in human cancer patients suggested that the expression of a specific α-2,6-sialyltransferase increases metastasis from the breast to the brain by increasing cellular adhesion to brain tissue. These observations, confirmed by testing in mouse models (9), illustrate the importance of studying glycosylation.

The difficulty associated with carbohydrate analysis has long been an impediment to such research. In this review, we highlight the latest advances addressing these issues, focusing on methods that are high throughput and are therefore breaking down barriers to the analysis of whole glycomes. We begin with a brief overview of glycan biosynthesis and structure, emphasizing issues that confound analysis of these complex molecules. We then cover a broad range of technologies for glycan analysis, including current applications of chromatography, mass spectrometry (MS), and microarray methodologies, that are advancing our understanding of glycosylation.

2. BACKGROUND

2.1. Glycan Biosynthesis

Carbohydrates can be found in a variety of contexts. Composed of 10 common monosaccharide units, carbohydrates can be expressed either as free reducing sugars, such as breast milk oligosaccharides (10), or as glycoconjugates to proteins or lipids (**Figure 1**). In proteins, glycans are found on asparagines (*N*-linked), serines or threonines (Ser/Thr-linked), or in rare instances, tryptophan residues (*C*-linked) (11). *C*-linked glycosylation is poorly understood and is not described further in this review.

2.2. *N*-Linked Glycans

N-linked oligosaccharides are the most-studied class of glycans. All *N*-linked glycans are elaborated from a core oligosaccharide of $Man_3GlcNAc_2$ (**Figure 2*a***), which is derived from the initiation of the *N*-linked biosynthetic pathway in the endoplasmic reticulum by the wholesale transfer of $Glc_3Man_9GlcNAc_2$ to an asparagine (Asn) found in a consensus sequence of Asn-!Pro-Ser/Thr (where !Pro refers to any residue except proline). Recent work has demonstrated other possible consensus sequences (12). Trimming of the glucose residues is a prerequisite for the exit of glycosylated proteins from the endoplasmic reticulum. Glycoproteins that contain the glucose residues are recognized by the chaperone proteins calnexin and calreticulin, which help to fold the proteins and thereby enable removal of the glucose units. Properly folded glycosylated proteins then exit from the endoplasmic reticulum to the Golgi apparatus, where additional modification of the glycans takes place. Glycan structures are altered in the Golgi apparatus to create the three major classes of *N*-linked glycans: high-mannose, complex, and hybrid (**Figure 2*a***).

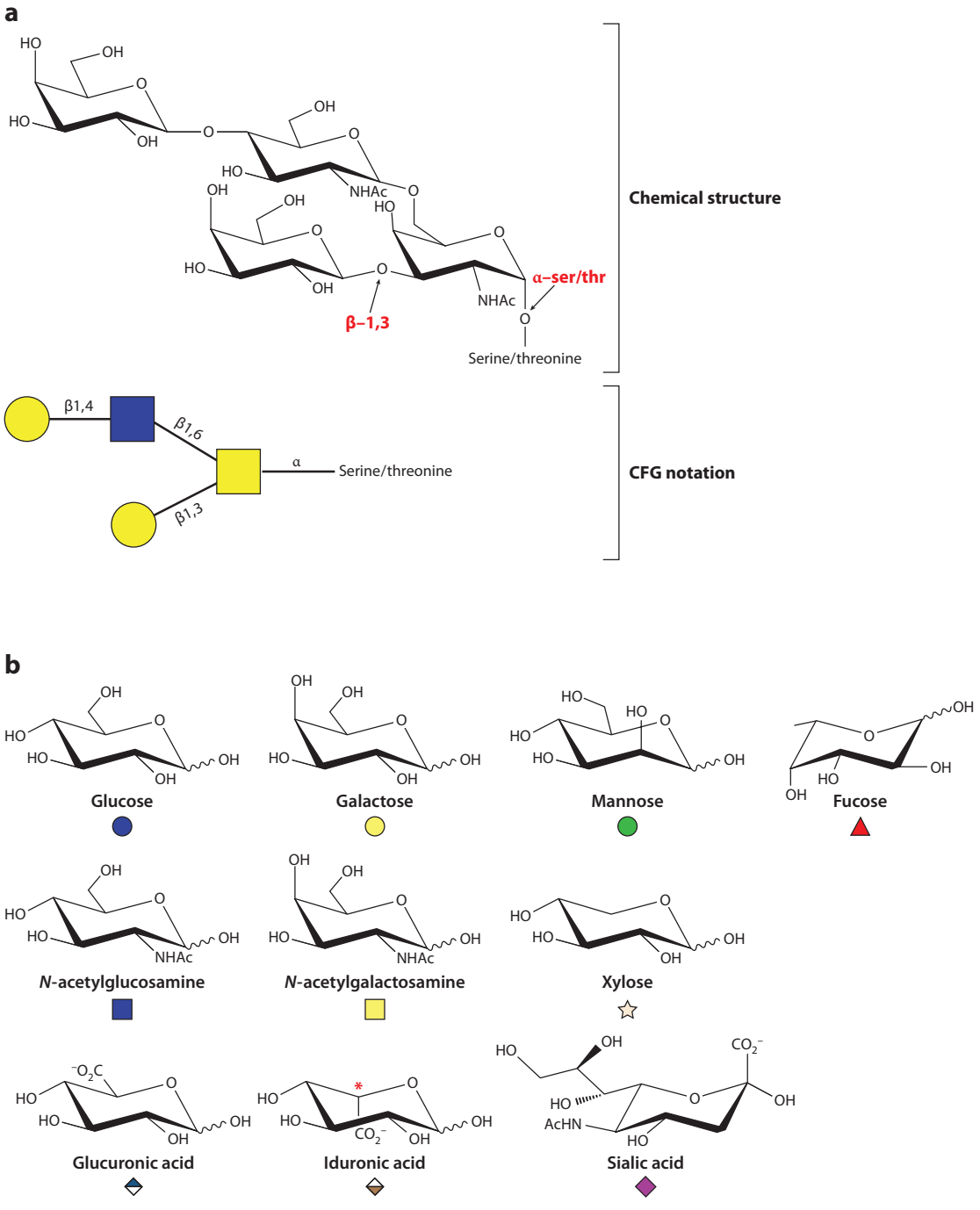

Figure 1

(*a*) Structures of human monosaccharides with Consortium for Functional Glycomics (CFG) notation. (*b*) Depiction of a glycan in CFG notation and the corresponding chemical structure. Both α- and β-linkages are indicated. Note that iduronic acid is made within oligosaccharide polymers from the epimerization of glucuronic acid by an epimerase enzyme and is not found as the free sugar. All monosaccharides are shown as their free reducing sugars.

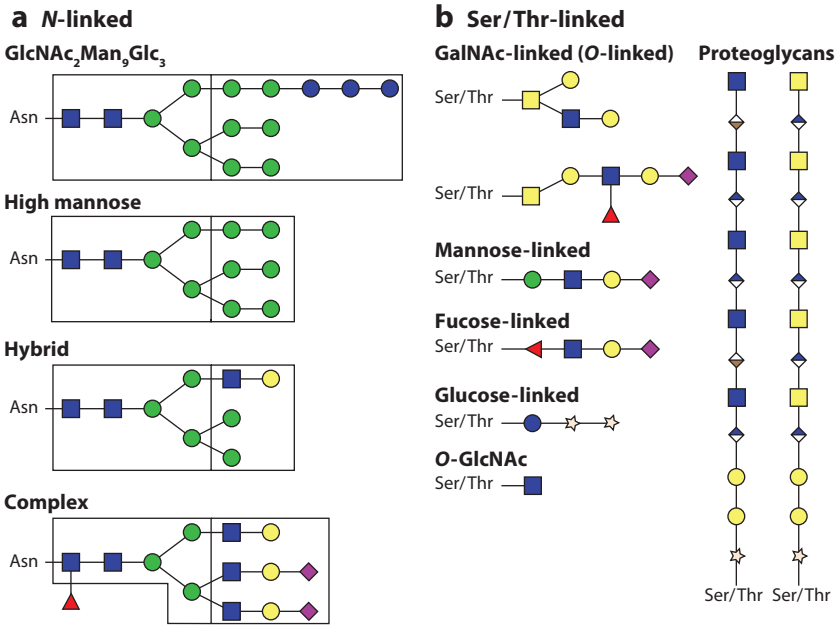

Figure 2

Representative structures of (*a*) asparagine (Asn)-linked, (*b*) serine/threonine (Ser/Thr)-linked, and (*c*) lipid-linked glycans in Consortium for Functional Glycomics notation. The conserved *N*-linked core structures are indicated by black boxes.

2.3. Serine/Threonine-Linked Glycans

The synthesis of Ser- or Thr-linked glycans differs from that of *N*-linked glycans in several key ways. In contrast to *N*-linked glycans, sugars attached to Ser/Thr are not bound by a conserved consensus sequence or a by single, conserved structural core. The canonical *O*-linked glycans constitute a set of structures in which Ser/Thr is conjugated to an α-linked GalNAc (**Figure 2*b***). These glycans are then modified to form several possible core structures that can be further altered to form a diverse array of *O*-linked glycans. The initial step in *O*-glycan synthesis is the transfer of GalNAc from uridine diphosphate (UDP)-GalNAc directly to a protein substrate by a polypeptide GalNAc transferase. Up to 24 such enzymes have been proposed to exist in the human genome (13), and 14 have been characterized to date. The piecemeal synthesis of *O*-linked glycans, combined with multiple initiator enzymes and the lack of a consensus sequence, makes this structural class difficult to predict and analyze.

Hypomannosylation: a state of decreased mannosylation, compared with normal levels

Muscular dystrophy: refers to a class of hereditary diseases in which muscle tissue is weakened

Several other subcategories of Ser/Thr-linked glycans exist. These include glycans formed via *O*-mannosylation, *O*-fucosylation, and *O*-glucosylation, as well as the proteoglycans heparan sulfate and chondroitin sulfate. In *O*-mannosylation, mannose that is α-linked to Ser/Thr forms a core residue that is then further modified (14). These glycans are essential for proper tissue development, as evidenced by the fact that hypomannosylation of the protein α-dystroglycan is a key feature in a number of muscular dystrophies (15). Both *O*-fucosylation, which consists of a α-Fuc-Ser/Thr core, and *O*-glucosylation, in which glucose is β-linked to Ser/Thr, are also core modifications that can be further altered and are important for the proper development of embryos (16). The proteoglycans heparan sulfate and chondroitin sulfate initiate with xylose β-linked to serine. This core residue is then extended in a linear fashion to create large, highly charged

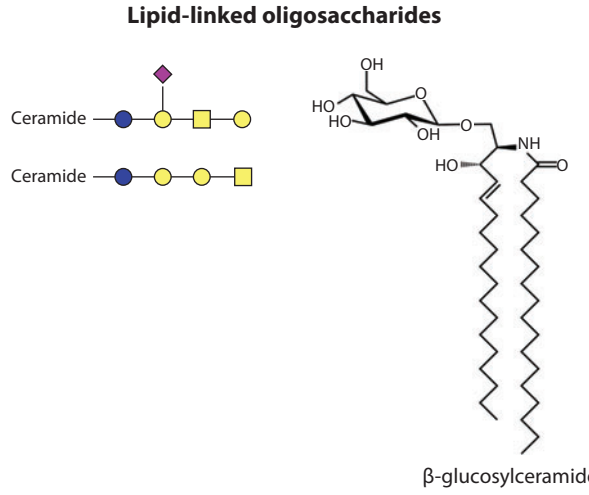

Figure 3

Representative structures of lipid-linked glycans in Consortium for Functional Glycomics notation and the chemical structure of a glucosylceramide. The length and saturation of the hydrophobic tail are variable.

structures that can contain complex patterns of epimerization (i.e. the conversion of glucuronic acid to iduronic acid), sulfation, and acetylation (11). These molecules form part of the extracellular matrix and are involved in a wide variety of biological functions, including neuronal guidance and wound healing (17). In contrast to the more elaborate Ser/Thr-linked modifications displayed on the extracellular surface, a simple monosaccharide, β-*O*-GlcNAc, has been found on Ser/Thr residues within the cell. This transitory modification shares hallmarks with phosphorylation and is thought to be a key signaling molecule (18–19).

2.4. Glycolipids

At present, there are ~500 known glycolipids that mediate cell-cell interactions and play roles in membrane microdomain association. Most mammalian glycolipids are synthesized in a stepwise fashion, beginning with addition of a monosaccharide (glucose or galactose) to the sphingosine ceramide by a glycolipid-specific glycosyltransferase. The synthesis of the core globoside, ganglioside, lactoside, and neolactoside structures are subsequently built from a glucosylceramide core (**Figure 3**). Galactose ceramide can also be generated, but galactosphingolipids are rarely extended beyond one or two additional monosaccharide units. A second class of glycolipids, the glycoglycerolipids, are found predominantly in plants and bacteria. They consist of sugars bound mainly to diacylglycerol cores and are abundant in cyanobacteria and plants (20).

Epimerization: refers, in this context, to the multiple stereogenic centers characteristic of carbohydrates

Ceramide: a type of lipid composed of a sphingoid base with an amide-linked fatty acid on its secondary amine; it has two large hydrophobic tails and a primary alcohol as the head group for coupling to other molecules such as phosphates or carbohydrates

2.5. Complex Structural Context

Several factors contribute to the complexity of glycan structural analysis. First, unlike the proteome, there is no DNA template that predicts the glycome. Rather, glycans are constructed biosynthetically from many reactions that involve redundant and overlapping enzymes, sugar nucleotide transporters, and other cellular machinery. Thus, standard genetic methods cannot be applied to glycan structural analysis because genetically inactivating a single glycosyltransferase often has unexpected consequences for glycan structure. For example, mice deficient in

Endoglycosidases: a class of enzymes that cleave glycosidic bonds in internal oligosaccharide residues

β-1,4-galactosyltransferase-1, the main enzyme responsible for the synthesis of galactosylated *N*-linked glycans in mouse, do not display a decrease in galactosylated *N*-linked glycans in their tissues. Rather, a shift of the structures from β-1,4-linked to β-1,3-linked galactose, as a result of compensation by related galactosyltransferases, is observed (21). Second, a specific modification site on a protein is not always associated with the same glycan structure; this condition is known as microheterogeneity. Thus, two molecules of the same protein from the same tissue sample may have different carbohydrate structures attached at a specific amino acid residue. Investigators formerly believed that the site occupancy—that is, whether or not a glycan would be attached—was also variable, but recent work argues that the vast majority of *N*-linked carbohydrate sites are fully occupied (12). Third, many of the monosaccharides that compose oligosaccharide structures are structural isomers, which complicates mass-based analysis (**Figure 1**). Fourth, glycans are often branched polymers that can be connected via either α- or β-stereochemistry at multiple potential linkage points between monosaccharide building blocks (**Figure 1*b***). Such connections generate an astronomical number of theoretical glycan structures [estimated to be 10^{12} for a hexasaccharide (22)]. However, the number of actual structures is probably significantly lower. The best estimates suggest that the number of possible structural determinants—that is, unique glycan components that can be recognized by a protein (two to six linear monosaccharides)—is on the order of 7,000 (23). These estimates do not take into account some of the structural variations observed in sialic acid, nor do they include combinations of terminal structures in branched structures, which may constitute new recognition elements. Therefore, the total number of unique glycan structures in the glycome is probably significantly larger than this value because multiple structures can be formed by combinations of glycan determinants. Compared with the 40,000 known genes of the human genome, the number of glycan determinants would seem to be a tractable analytical problem; however, the complexities of carbohydrate structure make the total characterization of a comprehensive glycome an elusive goal.

3. METHODS FOR GLYCOMIC ANALYSIS

Advances in glycomic analysis are beginning to enable systems-level characterization of whole glycomes. Many of these advances are rooted in traditional methodology and take advantage of miniaturization and advances in technology. In this section, we first review the reagents that have been commonly used for decades in glycan analysis, then the separation technologies for glycomics, including both stand-alone glycan-identification systems and coupled methods for MS. Finally, we review the latest MS- and array-based technologies for glycan analysis. We conclude this section by discussing some of the newest computer-based resources for the annotation and collection of analytical results.

The structural complexity of glycans requires isolation of more limited populations of carbohydrates for analysis. Typically, a glycan population is fractionated into its *N*-linked, Ser/Thr-linked, and glycolipid subfractions. Isolation of the *N*-linked glycome has been facilitated by use of the enzyme endoglycosidase peptide-*N*-glycosidase F (PNGaseF). This enzyme is the gold standard for detaching *N*-linked glycans because it cleaves the glycosidic bond between the Asn residue of a glycosylated protein and the core GlcNAc (**Figure 4*a***) This reaction is promiscuous, and it liberates nearly all *N*-linked glycans because of the common conserved core structure of this class (**Figure 2*a***). The drawback of PNGaseF is that it is incapable of cleavage if the core GlcNAc contains an α-1,3-linked fucose, which is a modification observed in plant and invertebrate, but not mammalian, glycan structures (**Figure 4*b***). Another endoglycosidase, PNGaseA, can cleave this modified core, but it demonstrates poorer overall efficiency. Other common enzymatic treatments for *N*-linked glycans involve use of the endoglycosidases F1–F3 and H, which cleave the

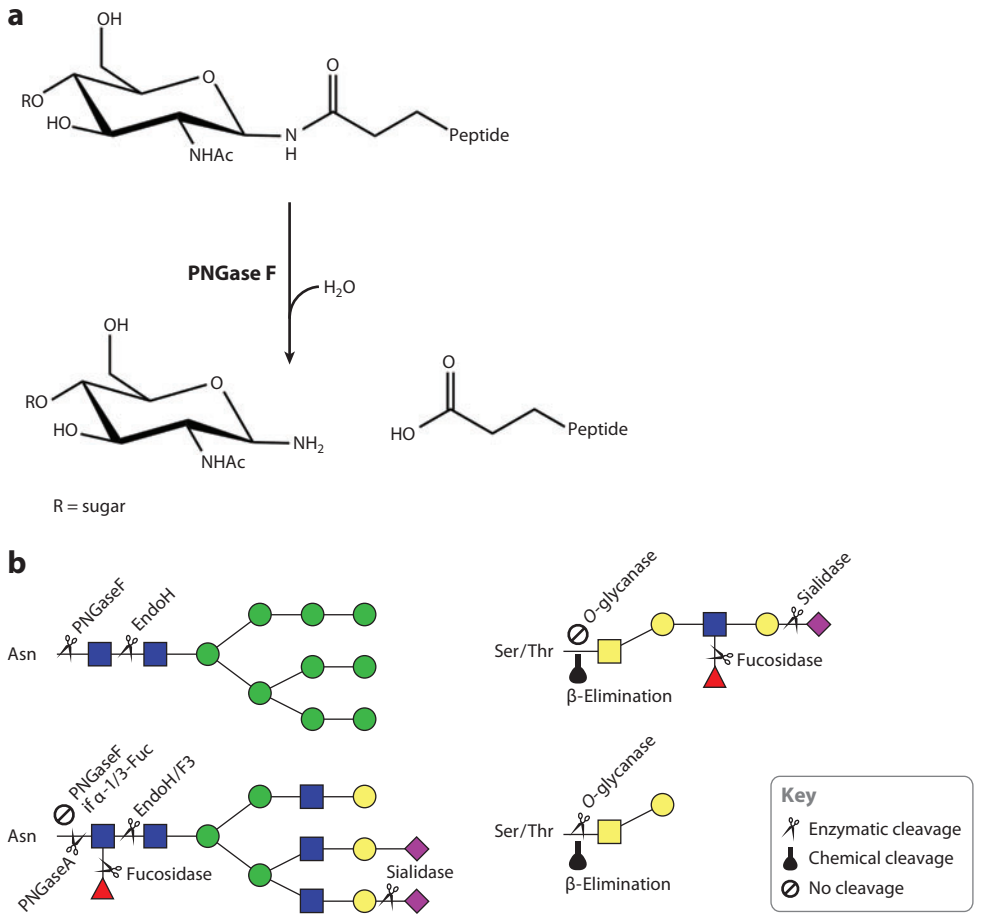

Figure 4

(*a*) Mechanism of peptide-*N*-glycosidase F (PNGase F)-catalyzed hydrolysis. (*b*) Representative examples of enzymatic and chemical methods for glycan cleavage and where they operate. Scissors represent enzymatic cleavage; beakers represent chemical cleavage; no symbol indicates no enzymatic cleavage.

GlcNAcβ-1,4-GlcNAc core of *N*-linked glycans with varying preferences for their substitution patterns. Due to the robustness of these enzymes, most glycomic analyses have been carried out on the *N*-linked subfraction.

Enzymatic cleavage of *O*-linked glycans is limited because the endoglycoside *O*-glycanase cleaves the core GalNAc-Ser/Thr bond only if the GalNAc has no modification other than β-1,3-Gal. This releases a narrow subset of *O*-linked glycans from the peptide backbone, unless structures are pretreated with exoglycosidases to trim the glycan structures to a cleavable core. More commonly, the *O*-linked glycome is released via a reductive β-elimination reaction through the use of a strong base in the presence of sodium borohydride to prevent further elimination, although other chemical strategies exist. This process is relatively inefficient and does not always provide complete subsets of *O*-linked glycans (24). In addition, unless *N*-linked glycans have been removed, the *N*-linked glycome is also susceptible to this treatment. The dearth of comprehensive

Exoglycosidases: a class of enzymes that cleave glycosidic bonds in terminal oligosaccharide residues

and efficient isolation methods for the *O*-linked glycome has caused comprehensive glycomic studies for *O*-linked glycosylation to lag behind those of *N*-linked glycosylation.

Galactosidases: a class of glycosidase enzymes that cleaves epitopes containing galactose

Fucosidases: a class of glycosidase enzymes that cleaves epitopes containing fucose

Glycolipids are often isolated by extraction with chloroform/methanol. Glycans are released from the lipid backbone by treatment with either enzymes or potentially destructive chemical cleavage agents, such as ozone or sodium periodate. The enzymes, termed endoglycoceramidases, come from the *Rhodococcus* species and cleave the glycan from the lipid (25). Although several variants of endoglycoceramidase exist, only one, EGCase II, is commercially available. EGCase II cleaves gangliosides and lactosylceramides, but it is inefficient at cleaving globosides and demonstrates no activity toward galactosides, simple glucosyl- and galactosylceramides, and glycoglycerolipids. The lack of efficient methods for glycolipid cleavage has stymied investigation of this important subclass of glycans.

The stepwise removal of terminal monosaccharides through the use of exoglycosidases provides an alternate method for obtaining structural information (**Figure 4*b***). The most commonly used enzymes are neuramidases (enzymes specific for sialic acids; also known as sialidases), galactosidases, and fucosidases. These reagents are often highly specific and can be used with either glycoconjugates or free oligosaccharides. They are often used in tandem with detection techniques such as high-performance liquid chromatography (HPLC) or MS to validate glycan structures. Recently, the sialyl Lewis X structures of α-1-acid glycoprotein were characterized by sequential enzymatic degradation by use of a series of four separate enzymatic treatments that cleave sialic acid, α-1-3/4 fucose, *N*-acetyllactosamine and β-1-4 galactose, respectively; MS was the detection technique (26).

Carbohydrate-binding proteins, known as lectins, are also commonly used for the structural characterization of glycans. Lectins, which are neither enzymes nor antibodies, are found in organisms ranging from bacteria to humans; many plant-derived lectins are commercially available. The plant lectins, such as SNA, a sialic acid binding lectin from *Sambucus nigra*, are used both to isolate and to identify glycans and glycoproteins. Many new techniques exploit the binding specificities of these reagents, both to provide structural detail and to subfractionate glycan populations for further analysis.

4. SEPARATION TECHNOLOGIES FOR GLYCAN ANALYSIS

Most of the methods used to analyze biological macromolecules, such as SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis), EMSA (electrophoretic mobility shift assay), thin-layer chromatography, and column chromatography, are rooted in the separation of unique molecular structures. Carbohydrates share special characteristics, including their hydrophilic nature, lack of a natural chromophore, and the multiplicity of closely related isomers, that require the use of specialized techniques for their separation. Three such methods are detailed below.

4.1. Capillary Electrophoresis

Capillary electrophoresis (CE) has been used to analyze a host of biological molecules, including DNA and small-molecule metabolites. For glycomic analysis, prefractionated glycans, typically the *N*-linked fraction from PNGaseF cleavage, are tagged with laser-induced fluorescence probes. These probes are typically attached to glycans by reductive amination, which is gentle enough to prevent degradation of fucosyl and sialyl moieties. Various probes allow the detection of glycans at low concentrations in neutral pH with a wide dynamic range (27). Once tagged, the ionized glycans are passed through narrow-bore channels in which the ratio of surface area to volume and electrical resistance are quite high. Samples separated in this way are subjected to very high

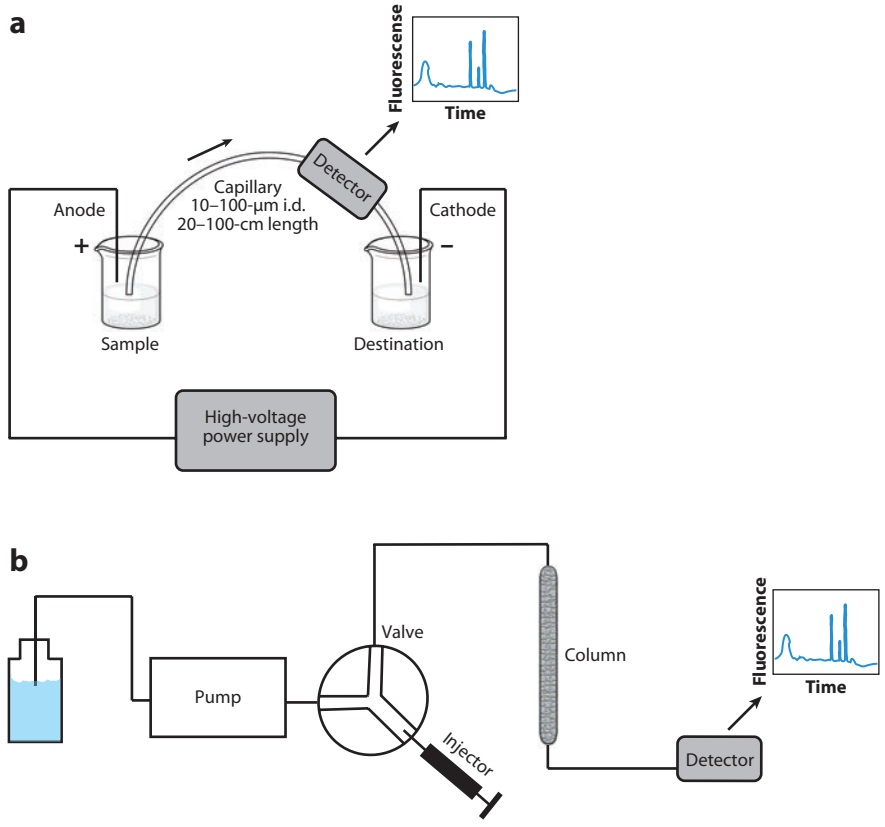

Figure 5

Schematic representation of (*a*) capillary electrophoresis and (*b*) high-performance liquid chromatography.

electrical fields, which allows for the rapid separation of structurally similar molecules in incredibly small volumes (**Figure 5*a***). Comparison between the separation time from the CE trace and that of a synthetic glycan standard enables identification of the carbohydrate structure.

CE has been used to elucidate *N*-linked carbohydrate structures on a number of glycoproteins (reviewed in Reference 28). Due to the inherently rapid analysis times, this technique can be applied in high-throughput fashion by multiplexing many capillaries to DNA sequencers, allowing nearly 200 samples per hour to be analyzed (29). Through the use of multiplexed CE, the glycosylation of influenza hemagglutinin from three flu strains cultured in six different animal cell lines has been characterized to yield fewer cross-reactive flu vaccine candidates (30). Low-abundance disaccharides in glycosaminoglycan structures have also been characterized by multiplexed CE (31). A recently developed high-throughput procedure for sample preparation enables rapid and reliable analysis of serum glycans for clinical identification of glycan biomarkers (32). The low sample volumes, high throughput, and wide detection range of this method are suitable for the generation of glycan signatures but not for detailed structural analysis. Drawbacks of this technique for determining discrete carbohydrate structures include the need for synthetic standards for every glycan analyzed, the need to prefractionate and label the glycans, the difficulty (but not impossibility) of performing MS in conjunction with CE, and the requirement that each peak in a trace belong to only a single glycan (which is not always the case).

4.2. High-Performance Anion-Exchange Chromatography

Although glycans are neutral at physiological pH (with the exception of the sialosides and structures that have been modified by phosphorylation or sulfation), the hydroxyl groups of carbohydrates are weakly acidic, and in sufficiently basic solutions (such as those at pH >12), they can be ionized. Charged carbohydrates can then be retained on strong anion-exchange resins in a process termed high-performance anion-exchange chromatography (HPAEC). A significant advantage of this approach is the detection method, which does not require fluorescent labeling. Instead, a technique known as pulsed amperimetric detection is used to electrochemically oxidize the glycan, primarily at the carbohydrate aldehyde groups (33). Because the signal depends not only on concentration of the glycan but also on the standard redox potential of the oxidized group, this method has exceptional quantitative and resolution capability. However, it is also limited by the need for carbohydrate standards to properly calibrate and identify signals in an HPAEC trace. This limitation has not prevented the use of HPAEC in glycomic characterization; the technique was utilized to identify the glycosylation on a human interferon engineered to be hyperglycosylated (34). HPAEC can be coupled to MS, but the inherently high salt concentration of the flow buffer is a detriment to analysis. In general, this technique is used primarily to study monosaccharide composition, nucleotide sugars, cellulose, and single-glycoprotein glycosylation (35).

4.3. High-Performance Liquid Chromatography

HPLC is one of the most utilitarian separation techniques available; it is frequently applied in glycomic research for the analysis of both neutral and charged carbohydrates. Similar to CE, glycan populations that are to be analyzed by HPLC are usually liberated by enzymatic means, then fluorescently tagged at the reducing end of the glycan. Carbohydrates are traditionally analyzed using normal phase HPLC columns, but reverse-phase columns are also employed. Detection of the separated, labeled glycans allows for reliable quantification (**Figure 5*b***).

A high-throughput implementation of HPLC-based glycan analysis was recently developed to investigate the *N*-linked glycome of human serum (36). This methodology utilizes a fully automated analysis platform in which (*a*) samples are immobilized in 96-well plates, (*b*) glycans are released and tagged, and (*c*) quantitative HPLC analytical traces are generated. Comparison of these traces with a database (GlycoBase) containing traces of more than 350 *N*-linked glycan standards provides a tentative structural assignment that is then confirmed with exoglycosidase digestions (annotated with autoGU, an automated program). This technique allows 96 samples to be analyzed in approximately four to five days. This system has identified serum biomarkers for a number of different conditions, including carbohydrate biomarkers in breast cancer (37), and it has been used to reveal environment-, lifestyle-, and genetics-based glycan alterations (38–40). In a recent study, gender- and stage-specific glycan biomarkers for schizophrenia were found in the low-abundance serum fraction and cerebrospinal fluid of schizophrenia patients (41). Interestingly, the observed *N*-glycome alterations recapitulate the known downregulation of certain GlcNAc- and sialyltransferases in the cerebral cortex of schizophrenia patients. This study demonstrates the power of HPLC.

Although HPLC has a much lower throughput capability than CE does, it has greater resolving power and is more amenable to multiplexing with other analytic devices, either at the front end with additional chromatographic steps or at the back end through coupling to MS. Additionally, automated structure-identification databases such as GlycoBase, GlycoExtractor, and autoGU (42, 43) may be used to rapidly assign glycan structures. The main utility of this technology is limited

to the *N*-linked glycome, and there remain problems concerning overlapping peaks and what portion of the *N*-linked glycome is represented by the current set of characterized carbohydrate standards whose HPLC traces are in the available database.

WGA: wheat germ agglutinin

RCA_{120}: *Ricinus communis* agglutinin

5. MASS SPECTROMETRY IN GLYCOMIC ANALYSIS

MS has been indispensable in the field of proteomics and is extremely useful for both glycoproteomics and glycomics. MS involves observing the mass-to-charge ratio of an ionized molecule, which can pose problems in the analysis of structural isomers such as glycans, in which many different structures have identical mass (see Section 2.5). As a result, technologies for glycoproteomics are more advanced than those for glycomics; both types of methods are discussed below.

5.1. Mass Spectrometry for Glycoproteomics

The field of glycoproteomics is dedicated to the identification of the glycosylated cohort of the proteome. By studying changes in glycoprotein expression and modification, investigators can place glycosylation in a biological context, assigning structures to specific proteins and thereby determining the role glycosylation might play in their function. MS-based glycoproteomics has been applied to identifying glycoproteins in sera, as well as biomarkers of healthy and cancerous tissue (44–46). In general, a typical glycoproteomic experiment involves the proteolytic digestion of a heterogeneous protein mixture, followed by glycopeptide enrichment, deglycosylation, and MS-based analysis of the corresponding peptide or proteome (**Figure 6**). Glycopeptides can be enriched for analysis through the use of several methods; one of the most popular uses lectin-affinity chromatography to subfractionate the glycoproteome (47). As mentioned above, lectins are nonimmunological proteins that bind carbohydrate motifs. They can have either broad or narrow specificities, and even the most promiscuous of lectins, such as the plant lectin concanavalin A (ConA) from jackbean, binds only a defined portion of the glycome, which biases the analysis in favor of the related glycoprotein fraction (**Figure 7**). To solve this problem, multiple lectins can be used to isolate a wider swath of the glycome (48). A recent variation of this technique was used by Mann and colleagues (12) to study the entire mouse *N*-glycoproteome. In this study, the authors mixed digested glycopeptides with a trio of lectins (WGA, RCA_{120}, and ConA) and filtered them through a size-exclusion membrane. The bound lectin-peptide complexes were then treated with PNGaseF to release the *N*-glycoproteome, and the mixture was again filtered. The cleaved peptides were subjected to mass spectrometric identification. Using this technology, the investigators observed 5,753 previously unidentified *N*-glycan sites, which revealed information about alternate consensus sequences for *N*-glycans and demonstrated that 98% of all *N*-linked carbohydrate sites are 100% occupied by a glycan. This study has fundamentally changed our understanding of microheterogeneity in terms of occupancy levels.

Chemical methods have also been used to isolate the glycopeptide fraction. The simplest of these techniques, hydrophilic interaction liquid chromatography (HILIC), takes advantage of the hydrogen-bonding, hydrophilic nature of glycopeptides to enrich this fraction (49, 50). HILIC is standard-phase chromatography in which water is one of the eluents and the solid phase is a polar, hydrogen-bonding matrix, such as an amide-, amine-, or diol-derivatized substrate. This technique is gaining in popularity for studies on protein glycosylation. Another method for fractionating the glycoproteome is based on the reaction of boronic acids with the 1,2-*cis*-diols found in some carbohydrates (51). Reaction of glycopeptides with boronic acids, coupled to solid support, enables isolation of the glycopeptide fraction (**Figure 7**). The peptide is then removed from the column via glycan cleavage, typically through enzymatic means. Although the vast majority of

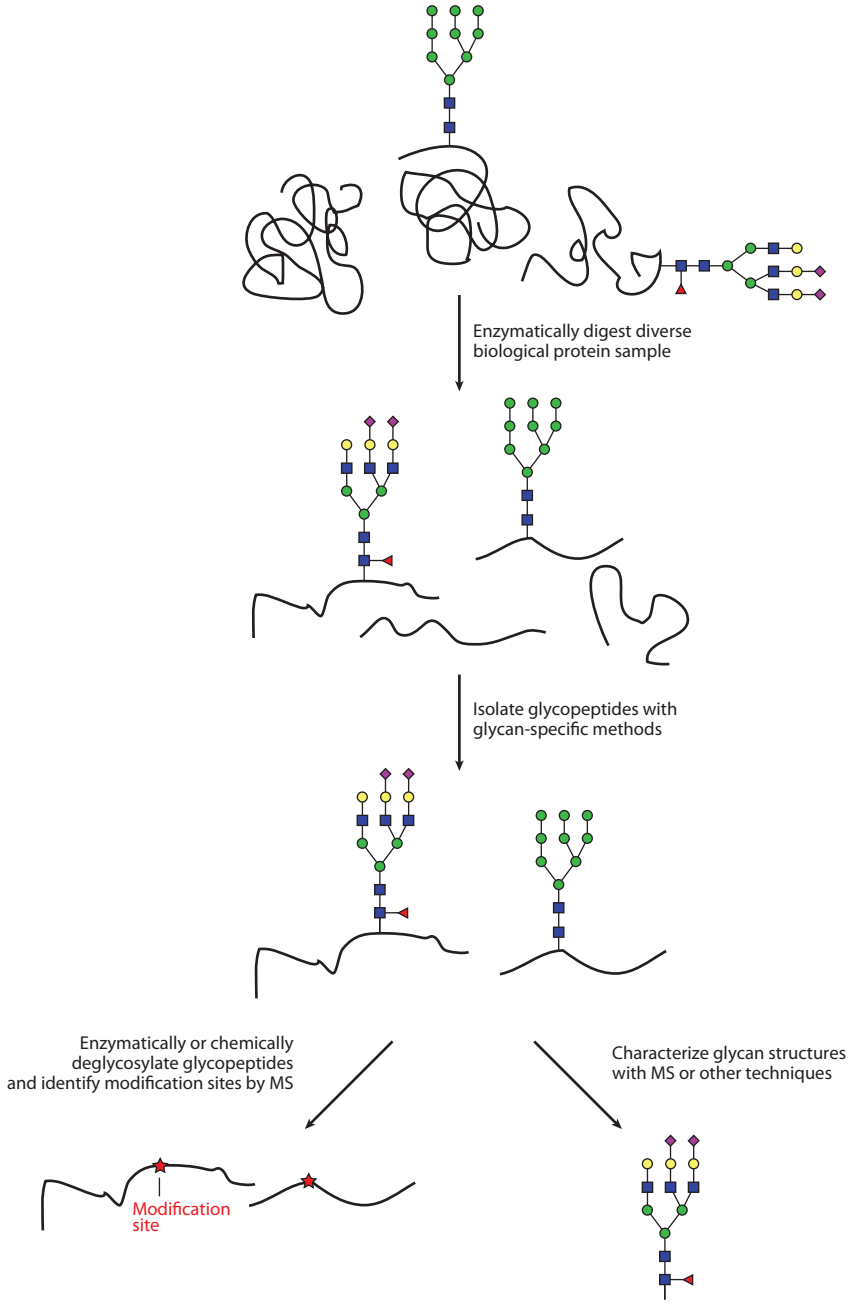

Figure 6

Experimental scheme for typical glycoproteomic or glycomic analysis by mass spectrometry (MS).

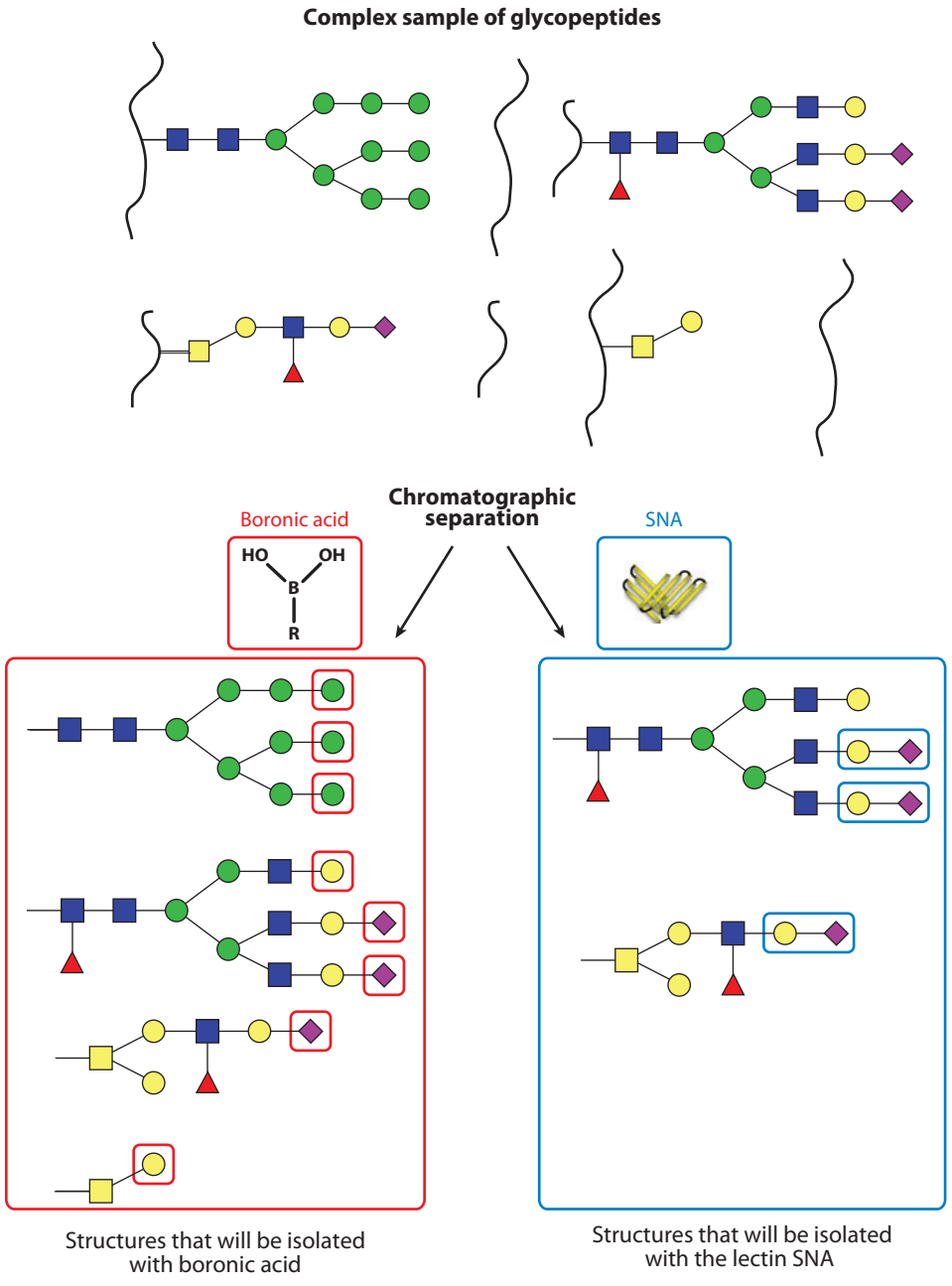

Figure 7

Separation methods for isolating pools of glycopeptides from a complex sample. Boronic acid binds to any 1,2-*cis*-diol; saccharides containing these structures are outlined in red. SNA binds to sialic acid residues (*outlined in blue*), isolating this subpopulation.

glycans are bound by such columns, the structural requirements for binding may preclude some elements of the glycome, such as those terminating in GalNAc. A third method for glycoproteome fractionation also relies on the presence of 1,2-*cis*-diols in glycans, which are oxidatively cleaved with sodium periodate to form the corresponding aldehyde-derivatized glycopeptide. Binding of the oxidized glycopeptide to either hydrazide or hydroxylamino beads enables separation of this fraction. Again, the deglycosylated peptides are typically recovered through the use of PNGaseF (52). This technique has been used to compare the *N*-glycoproteomes of T and B cells.

Once the glycopeptides are fractionated, the glycan moiety is typically released from the peptide backbone prior to peptide analysis. The methodology employed for deglycosylation often biases the observed portion of the glycoproteome because most studies use PNGaseF, which is specific for *N*-linked glycans. PNGaseF is particularly attractive for glycoproteomics because the reaction results in a mass shift on the modified asparagine of +1 Da in water or +3 Da in $H_2{}^{18}O$; this shift enables identification of the modified residues. Other methods for deglycosylation include β-elimination and hydrazinolysis, which cleave both *O*- and *N*-linked glycans. Once deglycosylated, peptides are subjected to standard proteomic analysis.

Matching a glycan structure with the site of the modification in a glycopeptide is one of the most vexing aspects of glycoproteomics, as most techniques remove the glycans prior to MS analysis. New MS tools, including electron-transfer dissociation fragmentation, are beginning to address this issue. In standard tandem mass spectrometry (MS/MS) of intact glycopeptides, the glycosidic bonds attaching the glycan to the amino acid are broken during fragmentation, obscuring identification of the glycosylation site. In contrast, electron-transfer dissociation fragments the glycopeptide at the N-C_α bond of the peptide backbone by a mechanism that, although poorly understood, maintains the glycosidic bond (53). The glycan remains attached to the residue and, by use of MS/MS-specific glycans, can be assigned to specific residues, as has been shown for the model glycoproteins horseradish peroxidase and human haptoglobin (54). This new protocol may help to bridge the gap between glycoproteomics and glycomics.

5.2. Mass Spectrometry for Glycomics

Glycomic analysis by MS is complicated by the isomeric nature of glycan structures. In MS analysis, glycans are isolated from glycoproteins or glycolipids and are either derivatized or directly ionized to yield MS spectra, which are then annotated (**Figure 6**). Specific glycan structures are often derived from higher-order MS^n experiments, in which fragmentation patterns reveal structural details. These methods require more advanced equipment and expertise (for an in-depth review on current MS techniques and technology, see Reference 55). The majority of high-throughput glycomic analysis is based on the tentative assignment of spectra from prefractionated glycans on the basis of known biosynthetic pathways and expected structures. These assignments are often confirmed through secondary techniques such as exoglycosidase digestion and others (mentioned above).

Sample preparation is a critical parameter in MS. Due to the complexity of the glycome, it is necessary to prefractionate glycans prior to MS analysis. Isolation of the *N*-, *O*-, or lipid-linked fractions uses the techniques described above in the reagent section for glycan cleavage. As with other techniques, the majority of MS analysis have also focused on *N*-linked glycans. The two main MS techniques used for glycan analysis are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (56, 57). Although both methods can detect unmodified glycans, samples are often permethylated to increase both the overall detection sensitivity and the stability of sialosides and sulfated monosaccharides to fragmentation in MALDI (55, 58). Permethylation itself can be used to aid glycomic structural identification if sequential permethylations are

performed following cleavage of either a sulfate or a sialic acid by use of isotopically labeled methyl groups (59, 60).

Advances in analysis methods and the development of databases to more rapidly annotate glycan MS spectra have led to new discoveries, especially in the field of biomarkers (61). MALDI-TOF-MS (where TOF stands for time-of-flight) was used to identify a cohort of three tri- and tetra-antennary complex *N*-glycan structures that were overexpressed in the serum samples of ~90% of patients with hepatocellular carcinoma. These glycans discriminated among patients with hepatocellular carcinoma, patients with high hepatitis C susceptibility and chronic liver disease, and normal controls (the sample population was 202) (62). Similar studies on serum glycomics have focused on the identification of breast, ovarian, prostate, and other cancer markers, as well as markers for disease states such as arthritis (63–65). Although these markers are not yet in clinical use, the identification of them is the first step toward new diagnostics in these diseases.

Glycomic studies can efficiently identify differential glycan signatures; however, MS-based glycomics generally fails to provide accurate comparative quantification of glycan structures in comparable samples. The interconversion of monosaccharides in the biosynthesis of complex oligosaccharides has prevented the use of metabolic labeling with isotope-labeled glycans; this process is a glycomics version of the traditional SILAC (stable isotope labeling by amino acids in cell culture) methodology used in proteomics, which relies on cell-culture labeling with heavy or light amino acids (66) for comparative glycomic analysis. However, recent advances in glycan labeling with mass tags have begun to allow direct comparison of carbohydrate samples by MS. Simple reagents, including sodium borohydride and borodeuteride (67), ^{13}C-aniline (68), and deuterated hydroxylamine derivatives (69) can successfully introduce differential mass tags that target the reducing ends of glycans. In a typical experiment, glycans from two samples are prepared and differentially labeled with reagents of distinct isotopic mass. This process allows the two samples to be combined for a single MS run, which enables comparative quantitation. In recent research, a technique similar to SILAC known as IDAWG (isotopic detection of aminosugars with glutamine) was developed for the isotopic labeling of glycans in cell culture. IDAWG takes advantage of the requirement for glutamine as the sole biosynthetic source of nitrogen in UDP-GlcNAc synthesis (70). UDP-GlcNAc is utilized in the biosynthesis of GalNAc and sialic acid, the other two amino sugars; thus, incubation with ^{15}N-Gln labels all the glycans that contain an *N*-acyl group. Although this method requires cell culture to incorporate the isotopic label, it results in a broad sampling of the glycome because canonical *N*- and *O*-linked oligosaccharides contain at least one sugar with an *N*-acyl group.

The development of bioinformatics tools is crucial for the rapid and accurate analysis of glycomic data generated from high-throughput MS experiments. The most commonly used programs are based on algorithms that assign theoretically generated glycan structures to MS or MS^n peaks. Cartoonist, one of the most sophisticated of these programs, applies biosynthetic criteria to the theoretical glycan pool to eliminate implausible candidates, thereby creating a short list of probable structures for each peak (71, 72). There are versions of this program for the annotation of both *N*- and *O*-linked glycans. Glyco-Peakfinder (73) and GlycoWorkbench (74) utilize a strategy similar to that of Cartoonist, but with a theoretical glycan pool generated from the EuroCarbDB database of characterized glycan structures, rather than biosynthetic constraints. A related program, SimGlycan, uses data files from an internal database of MS experiments to generate matching glycan structures, then develops a theoretical spectrum for the best-matched glycan (75). Complementary tools such as STAT, OSCAR, and StrOligo are alternative strategies for rapid annotation (76–78). As faster, higher-volume analysis of carbohydrates by MS becomes a reality, the ability to quickly and accurately annotate MS data without advanced expertise will be necessary to advance the field at a pace comparable to that of proteomics.

6. MICROARRAY-BASED METHODS

Epitope: a discrete carbohydrate structure than can be found in more than one oligosaccharide (e.g., LacNAc)

The microarray is one of the most valuable tools in the study of nucleic acid expression and dynamics. Conceptually, a microarray is a miniaturized format that enables the study of multiple probes and, simultaneously, their cognate-recognition partners. Microarray-based technology is becoming increasingly important to the field of glycomics, and it can be divided into three general subtypes: carbohydrate microarrays, lectin microarrays, and glycoprotein microarrays. Carbohydrate microarrays, in which glycans are deposited on the microarray surface, do not reveal glycosylation patterns, but rather demonstrate the binding properties of glycan-binding proteins. Because these microarrays have been extensively described in multiple excellent reviews (65, 79, 80) and do not directly analyze the glycome, we do not cover them here.

6.1. Lectin Microarrays for Glycomic Analysis

As discussed above, lectins are a class of proteins from diverse sources that bind carbohydrate epitopes. The binding specificities of lectins are well documented and range from general to highly specific (81). Lectin affinities are demonstrably weaker than antibody affinities and have dissociation constants in the micromolar range. However, many of these proteins consist of multiple binding units, and the glycans to which they bind are often presented in clustered settings, which increases the avidity of the interaction (82). Much of the glycan analysis in the literature centers around lectin-based techniques including ELLA (enzyme-linked lectinsorbent assay), lectin histology, lectin blots, lectin-affinity chromatography, and lectin-based flow cytometry (83).

In lectin microarrays, individual lectins are deposited in discrete spots on a solid support (**Figure 8*a***). These proteins can be either (*a*) covalently linked in a random orientation to supports such as epoxy or *N*-hydroxysuccinimide (NHS)-ester-modified glass slides or (*b*) adsorbed onto nitrocellulose or hydrogel slides (84–86). The lectins are printed with either contact (pin-print) or noncontact (e.g., piezoelectric) arrayers. A recent innovation in lectin microarray technology arises from the oriented deposition of recombinant bacterial lectins containing a glutathione *S*-transferase–fusion domain. Printing these lectins on a glutathione-modified NHS slide led to enhanced activity as a result of orientation, which resulted in a more sensitive lectin microarray platform (87). Lectin microarrays are used to interrogate a diversity of biological systems ranging from human immunodeficiency virus 1 (HIV-1) to whole bacteria to cancer glycomes (88).

A typical lectin microarray experiment involves using a fluorescent dye, such as cyanine-3 or -5, to label purified glycoproteins, protein lysates, or cellular membrane samples on the amines of proteins with a fluorescent dye via NHS-coupling chemistry. Depending on the sample, both *N*- and *O*-linked glycans, as well as glycolipids (for micellae), may be present and analyzed. Labeled samples are incubated with the array to allow lectin-carbohydrate binding to occur. The slides are then scanned to reveal the lectin-binding pattern (**Figure 8*a***). Two types of microarray scanners are used in lectin microarray analysis: (*a*) the traditional Cy3/Cy5 scanners used in gene microarray experiments and (*b*) an evanescent-wave scanner system. When traditional scanners are employed, the slides are washed prior to analysis, and the typical detection limits are ~10 μg ml^{-1} (89); however, in oriented lectin microarrays, glycans have been detected at concentrations of glycoprotein as low as ~12 ng ml^{-1} (~640 pM protein) (87). Evanescent-wave technology enables the detection of glycans without the need to wash the slides. In this method, only the fluorescence signal at the array surface is observed. The detection limit for this technology is 1 ng ml^{-1} of glycoprotein (~100 pM protein) (89). This more sensitive detection method has enabled the analysis of minute samples derived from tissue microarray slides (90).

In a single-color lectin microarray experiment, the types of glycan motifs in a sample (known as the glycocode) are revealed through the pattern of lectin binding (**Figure 8*a***). Such experiments

are valuable in describing the global carbohydrate expression in a sample; however, quantification between multiple complex samples is difficult because not all the lectins bind in their linear range. Therefore, some lectins show no difference in binding between samples on the array, despite clear differences in the relative levels of the cognate glycan epitopes in the samples being compared (91). Single-color lectin microarray experiments have been used to study differential glycosylation patterns in glycoproteins (92), demonstrate cell type– and developmental stage–specific glycosylation (93–95), and distinguish between pathogenic and nonpathogenic bacteria (96).

One solution to the relative quantitation conundrum is to utilize a two-color methodology for lectin microarray analysis (**Figure 8*b*,*c***). In this approach, two differentially labeled samples hybridized on the same array compete for lectin binding (91). This ratiometric analysis allows for the detection of subtle differences in glycosylation expression among many samples by comparing them with a common reference. In initial research, this method demonstrated extensive glycome alterations in the promyelocytic leukemia cell line HL-60 upon differentiation into neutrophils (**Figure 8*d***). More recently, two-color analysis revealed, strikingly, that HIV virions exiting an infected cell have a glycome similar to that of the host microvesicles (a membrane-derived particle), which suggests that the virus co-opts the natural microvesicle pathway to avoid recognition by the immune system (88). These examples demonstrate the power of lectin microarrays for the analysis of glycans in biological processes that change the glycome

Although the lectin microarray approach has multiple advantages in terms of throughput, speed, comprehensiveness of the glycomic analysis (i.e., observation of *N*-linked glycans, *O*-linked glycans, and glycolipids), and breadth of possible target samples, it also has limitations, such as an inability to analyze glycan motifs that are not represented by the lectin specificities (97). There are thought to be 7,000 estimated glycan motifs, but even the most advanced arrays have fewer than 100 binding proteins. Thus, the specific sulfation patterns of glycans, *O*-mannosylation, *O*-fucosylation, and many of the glycolipids, for example, are not represented by glycan binders. Expansion of the arrays through the addition of new recombinant lectins and antibodies, the development of synthetic lectins (98), and the evolution of novel lectins are beginning to address this problem. Another issue is that of lectin specificity. Although we have a great deal of binding data on the lectins, questions remain about the specificities of even the most commonly utilized lectins, such as the sialic acid binder MAA (*Maackia amurensis* lectin). New data from carbohydrate microarrays, frontal affinity chromatography, and other techniques are filling in the gaps of our knowledge and will advance our annotation of the array data (99, 100). The glycosylation of plant lectins is another potential complication of lectin analysis, particularly for bacteria (86). The addition of new, recombinant, nonglycosylated lectins to the arrays is one remedy for this problem. Finally, unlike advanced MS techniques, lectin microarrays do not allow discrete molecular-level analysis of the glycome. Instead, structural epitopes within complex mixtures are identified. Given that glycan binding in biological systems rarely adheres to the canonical "lock-and-key" model for protein binding to substrate and instead relies upon interactions with multiple glycan structures that bear similar epitopes, this drawback may not be an issue when these arrays are used to interrogate biological processes.

6.2. Microarray Analysis of Glycoproteins

The microarray format has also been used to identify the glycosylation patterns of the discrete glycoproteins represented on the arrays. Two forms of glycoprotein array technology exist: the glycoprotein array (101) and the antibody-lectin sandwich array (102). For both, lectins are key probes for the identification of glycan epitopes. In glycoprotein arrays, glycoproteins from a complex sample, such as immunodepleted serum, are fractionated with lectin-affinity

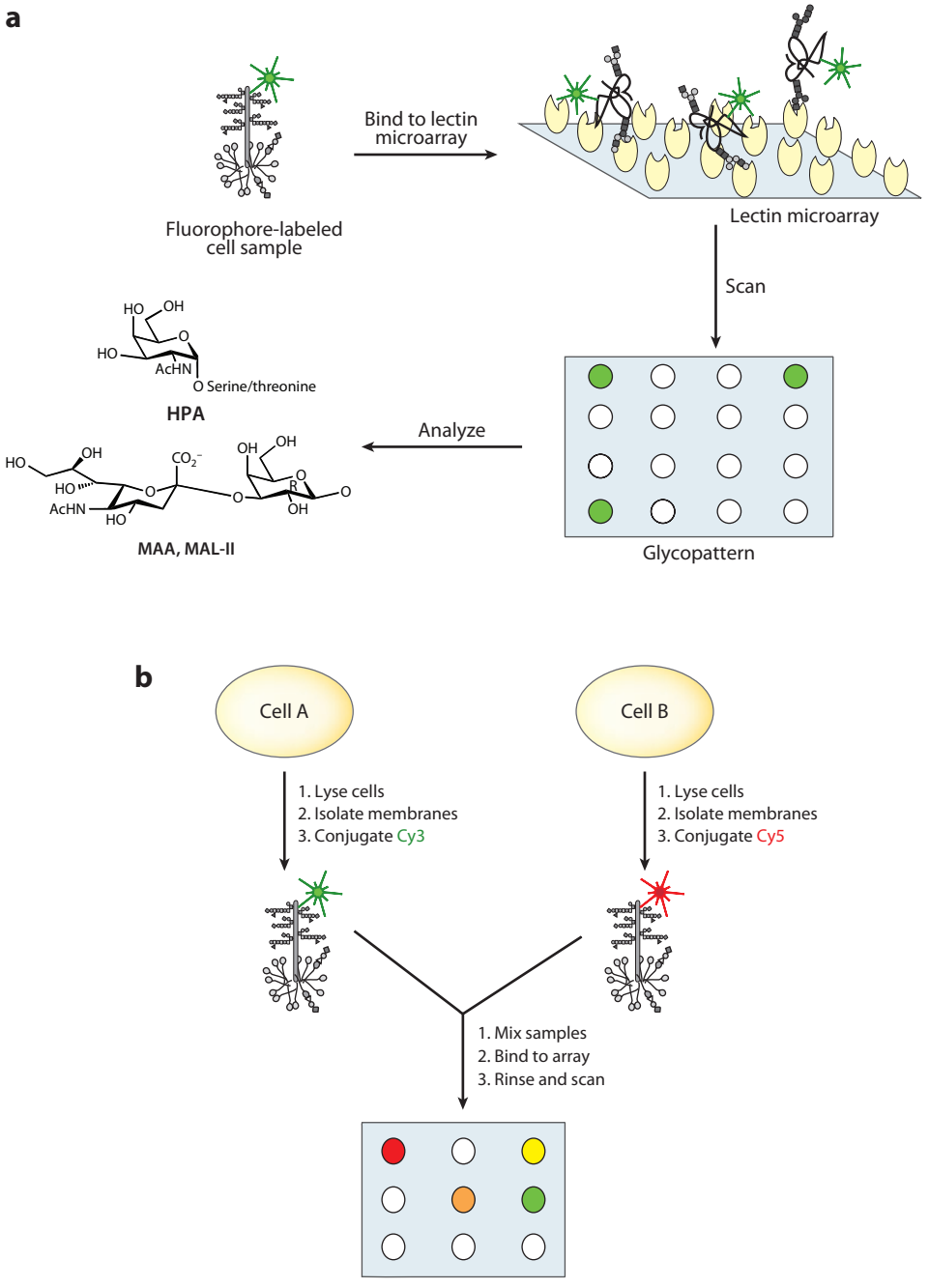

Figure 8

Experimental scheme for (*a*) single-color and (*b*) dual-color lectin microarray experiments. (*c*) A typical example of a dual-color hybridization scanned at the Cy3 and Cy5 wavelengths. (*d*) A dual-color hybridization profile represented as a heat map of the average $\log_2$-transformed binding data. Reproduced from Reference 91.

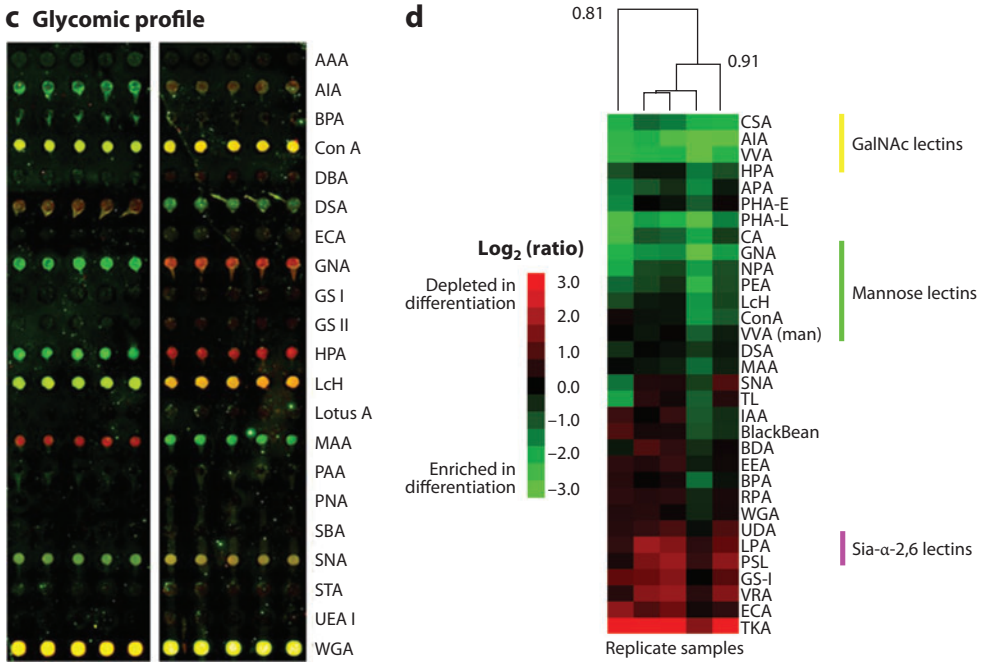

Figure 8

(*Continued*)

chromatography, followed by HPLC, and the discrete fractions are spotted onto nitrocellulose substrates. The lectins are then incubated with the array to probe the expressed glycome of the sample. By printing fractions from multiple samples and probing multiple arrays with varying lectins, one can compare the glycomes of different samples. Once differences in the glycan composition of a fraction are observed, MS can be used to identify the glycoproteins within. Lubman and coworkers (103) have used this approach to identify several potential glycoprotein biomarkers for colorectal cancer.

In antibody-lectin sandwich arrays, antibodies to known glycoproteins are arrayed on a solid support, and complex glycoprotein samples are bound to the microarray. Samples can be either crude or prefractionated to concentrate the glycoprotein fraction. The hybridized arrays are then incubated with lectins that recognize the carbohydrates on the bound glycoproteins. The lectins can be either fluorescently labeled for direct detection or biotinylated for secondary detection with labeled streptavidin, which is more common. This protocol thus identifies specific glycosylation features of discrete glycoproteins (104–106). In one of the most striking uses of this technology, Haab and colleagues (106) utilized the antibody-lectin sandwich array to analyze the sera of matched wild-type mice and littermates with intestinal tumors. The authors found that even when the mice had few or no tumors, there were clear differences in their glycoprotein glycosylation, which suggests that there may be a potential systemic response to early-onset cancer. This experiment demonstrates the power of this high-throughput method in identifying glycome dynamics at the proteomic level. Antibody-lectin sandwich technology requires relatively small amounts of samples for detection and can trace glycan structures to their specific glycoproteins; however, the system is limited in scope because it requires excellent antibodies to specific glycoproteins, and only glycoproteins represented by antibodies on the array can be observed. As investigators

get closer to creating monoclonal antibodies to the entire human proteome (107), it may become possible to study their glycoproteomic variants by using this technology.

7. GLYCOINFORMATICS

High-throughput analytical methodologies, such as those discussed above, require advanced computational and databasing infrastructure to properly analyze and place in context the large-scale data generated. Genomic field databases such as UNIPROT, GenBank, and GEO and search tools such as BLAST have rapidly advanced scientific inquiry. Complications in developing informatics for the glycome arise from the nonlinear nature of the biopolymer and its posttranslational status. Notation of glycan data in a common computer syntax that is suitable for search engines and databases has led to incompatibilities between glycoinformatics resources. A common symbolic notation for glycans has been created by the Consortium for Functional Glycomics (CFG). This representation, used in this review, is the most popular way of displaying complex glycan structures. An XML format for computational glycan notation, GLYDE-II, is a potential standard notation to facilitate data exchange between glycan databases; however, none of the major databases (CFG, KEGG Glycan, and Glycosciences.de) utilize this format (108). Instead, they use three different notations: IUPAC for CFG, a graphical representation (KEGG chemical function format, known as KCF) for KEGG Glycan, and LINUCS for Glycosciences.de. Although converters between some of these formats and GLYDE-II are available, the community must adopt a common standard and make translators so that data exchange can occur.

Glycoinformatics resources include the glycan structure databases mentioned above, new methods for the interpretation of lectin-based data, and prediction methods for the glycosylation sites of glycoproteins (for a thorough review of glycoinformatics, see Reference 109). The three databases are extensions of an older database known as CarbBank. CarbBank stores ~50,000 structures of glycans mined from publications and, although it is no longer updated, remains a useful resource (110). Notably, the CFG database contains not only structures but also experimental data generated by the CFG. Because its glycan microarrays are available to glycobiology researchers, the CFG database is the foremost resource for glycan microarray data on carbohydrate-protein-binding interactions. The microarray has been used to screen binding-recognition lectins, antibodies, glycosyltransferases, and glycosidases, as well as complex biological samples. Recently, Porter et al. (111) annotated the binding of glycan motifs by lectins and other proteins using the CFG glycan microarray data set. These authors segregated all 500 structures present on the CFG glycan microarray into 63 naturally occurring, unique carbohydrate epitopes, then characterized the binding proteins according to their ability to bind structures containing these motifs. In so doing, the authors created a database of motifs that can be cross-referenced with binding proteins, which may aid the annotation of lectin-based binding data.

An important aspect of glycoinformatics is the application of computational resources to the prediction of glycosylation on proteins and the structure of the glycans. Neural network technology has been applied to this problem, with varying levels of success. Through the use of statistical methods, and aided by a common consensus sequence, NetNGlyc predicts sites of *N*-linked glycosylation with an accuracy greater than 70% (112). Similar models, although they lack a consensus sequence, have been applied to *O*- and *C*-linked glycosylation and to *O*-GlcNAcylation (113, 114). To apply complementary predictive algorithms to glycan structure, the GlycO (Glycomics Ontology) program uses a biosynthetic model to develop glycan structures based on genetic data of glycan synthesis and modification enzymes (115). The inherent lack of a template code for glycan synthesis is still a major impediment to the systematic development of glycoinformatics tools,

but the addition of high-throughput data from multiple sources, including glycogene analysis, is beginning to enable new methods.

8. FUTURE DIRECTIONS

As progress continues toward fully automated, high-throughput methodology for glycomic studies, we expect that the many techniques used to study glycosylation as a system will unify into a single approach. A small-scale proof of such a systems biology approach has been demonstrated through comparison between the glycomes of HepG2 and HuH-7 liver carcinoma cell lines (116). Using quantitative PCR, lectin microarrays, and MALDI-MS/MS, the authors of this study identified key differences in the fucosylation of LacdiNAc on α-fetoprotein between the two cell lines. Narimatsu et al. (117) described a similar method for cancer biomarker determination by performing glycogene microarray experiments in concert with lectin microarrays on released glycans to determine changes in gene and glycan regulation while identifying specific sites of glycosylation on proteins by ^{18}O incorporation with PNGaseF. These studies represent the first step toward a comprehensive understanding of the glycome.

Several hurdles, however, must be overcome to enable the system-wide study of glycosylation. As described exhaustively above, the vast majority of methodologies for glycomics are focused on the *N*-linked glycome and the glycoproteomic cohort. Analysis of conventional *O*-linked glycans and of alternate Ser/Thr-glycan modifications, as well as of glycolipids, lags behind drastically in terms of available methodology. Several new techniques, such as lectin microarrays, can observe the broader glycome, but they too have limitations based on the probes available. Of the 7,000 potential glycan epitopes that could be recognized in the glycome, only a small subset can typically be analyzed by any method. Thus, there remain vast unexplored areas of the glycome that we must observe before we can truly comprehend the role of glycans in our complex human biology.

DISCLOSURE STATEMENT

L.K.M. is a member of the Consortium for Functional Glycomics. The authors are not aware of any other affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

L.K.M. acknowledges award number DP2OD004711 from the Office of the Director, National Institutes of Health (NIH), and the National Institute on Alcohol Abuse and Alcoholism. The content is solely the responsibility of the authors and does not represent the official views of the NIH.

LITERATURE CITED

1. Ohtsubo K, Marth JD. 2006. Glycosylation in cellular mechanisms of health and disease. *Cell* 126:855–67
2. Marth JD, Grewal PK. 2008. Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* 8:874–87
3. Sharon N. 2006. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta* 1760:527–37
4. Dube DH, Bertozzi CR. 2005. Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* 4:477–88
5. Apweiler R, Hermjakob H, Sharon N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* 1473:4–8

6. Haltiwanger RS, Lowe JB. 2004. Role of gycosylation in development. *Annu. Rev. Biochem.* 73:491–537
7. Freeze HH. 2006. Genetic defects in the human glycome. *Nat. Rev. Genet.* 7:537–51
8. Fernandes B, Sagman U, Auger M, Demetrio M, Dennis JW. 1991. β1–6 Branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer Res.* 51:718–23
9. Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, et al. 2009. Genes that mediate breast cancer metastasis to the brain. *Nature* 459:1005–9
10. Bao Y, Newburg DS. 2008. Capillary electrophoresis of acidic oligosaccharides from human milk. *Electrophoresis* 29:2508–15
11. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, et al., eds. 2008. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harb. Lab. 2nd ed.
12. Zielinska DF, Gnad F, Wiśniewski JR, Mann M. 2010. Precision mapping of an in vivo *N*-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141:897–907
13. Ten Hagen KG, Fritz TA, Tabak LA. 2003. All in the family: the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases. *Glycobiology* 13:1–16R
14. Stalnaker SH, Hashmi S, Lim JM, Aoki K, Porterfield M, et al. 2010. Site mapping and characterization of *O*-glycan structures on α-dystroglycan isolated from rabbit skeletal muscle. *J. Biol. Chem.* 285:24882–91
15. Lommel M, Strahl S. 2009. Protein *O*-mannosylation: conserved from bacteria to humans. *Glycobiology* 19:816–28
16. Luther KB, Haltiwanger RS. 2009. Role of unusual *O*-glycans in intercellular signaling. *Int. J. Biochem. Cell Biol.* 41:1011–24
17. de Wit J, Verhaagen J. 2007. Proteoglycans as modulators of axon guidance cue function. *Adv. Exp. Med. Biol.* 600:73–89
18. Hart GW, Housley MP, Slawson C. 2007. Cycling of *O*-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 446:1017–22
19. Slawson C, Housley MP, Hart GW. 2006. *O*-GlcNAc cycling: how a single sugar post-translational modification is changing the way we think about signaling networks. *J. Cell Biochem.* 97:71–83
20. Hölzl G, Dormann P. 2007. Structure and function of glycoglycerolipids in plants and bacteria. *Prog. Lipid. Res.* 46:225–43
21. Kotani N, Asano M, Iwakura Y, Takasaki S. 2001. Knockout of mouse *β1,4-galactosyltransferase-1* gene results in a dramatic shift of outer chain moities of *N*-glycans from type 2 to type 1 chains in hepatic membrane and plasma glycoproteins. *Biochem. J.* 357:827–34
22. Laine RA. 1994. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the isomer barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* 4:759–67
23. Cummings R. 2009. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* 5:1087–104
24. Duk M, Ugorski M, Lisowska E. 1997. β-Elimination of *O*-glycans from glycoproteins transferred to immobilon P membranes: method and some applications. *Anal. Biochem.* 253:98–102
25. Ito M, Yamagata T. 1989. Purification and characterization of glycosphingolipid-specific endoglycosidases (endoglycoceramidases) from a mutant strain of *Rhodococcus* sp. Evidence for three molecular species of endoglycoceramidase with different specificities. *J. Biol. Chem.* 264:9510–19
26. Kannicht C, Grunow D, Lucka L. 2008. Enzymatic sequence analysis of *N*-glycans by exoglycosidase cleavage and mass spectrometry—detection of Lewis X structures. *Methods Mol. Biol.* 446:255–66
27. Vanderschaeghe D, Festjens N, Delanghe J, Callewaert N. 2010. Glycome profiling using modern glycomics technology: technical aspects and applications. *Biol. Chem.* 391:149–61
28. Mechref Y, Novotny M. 2009. Glycomic analysis by capillary electrophoresis–mass spectrometry. *Mass Spectrom. Rev.* 28:207–22
29. Laroy W, Contreras R, Callewaert N. 2006. Glycome mapping on DNA sequencing equipment. *Nat. Protoc.* 1:397–405
30. Schwarzer J, Rapp E, Hennig R, Genzel Y, Jordan I, et al. 2009. Glycan analysis in cell culture–based influenza vaccine production: influence of host cell line and virus strain on the glycosylation pattern of viral hemagglutinin. *Vaccine* 27:4325–36

31. Vogel K, Kuhn J, Kleesiek K, Götting C. 2006. A novel ultra-sensitive method for the quantification of glycosaminoglycan disaccharides using an automated DNA sequencer. *Electrophoresis* 27:1363–67
32. Vanderschaeghe D, Szekrenyes A, Wenz C, Gassmann M, Naik N, et al. 2010. High-throughput profiling of the serum *N*-glycome on capillary electrophoresis microfluidics systems: toward clinical implementation of GlycoHepatoTest. *Anal. Chem.* 82:7408–15
33. Cataldi TR, Campa C, Angelotti M, Bufo SA. 1999. Isocratic separations of closely related mono- and disaccharides by high-performance anion-exchange chromatography with pulsed amperometric detection using dilute alkaline spiked with barium acetate. *J. Chromatogr. A* 855:539–50
34. Ceaglio N, Etcheverrigaray M, Conradt HS, Grammel N, Kratje R, et al. 2010. Highly glycosylated human α interferon: an insight into a new therapeutic candidate. *J. Biotechnol.* 146:74–83
35. Behan JL, Smith KD. 2011. The analysis of glycosylation: a continued need for high pH anion exchange chromatography. *Biomed. Chromatogr.* 25:39–46
36. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, et al. 2008. HPLC-based analysis of serum *N*-glycans on a 96-well plate platform with dedicated database software. *Anal. Biochem.* 376:1–12
37. Abd Hamid UM, Royle L, Saldova R, Radcliffe CM, Harvey DJ, et al. 2008. A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression. *Glycobiology* 18:1105–18
38. Knezević A, Polasek O, Gornik O, Rudan I, Campbell H, et al. 2009. Variability, heritability and environmental determinants of human plasma *N*-glycome. *J. Proteome Res.* 8:694–701
39. Pucic M, Pinto S, Novokmet M, Knezevic A, Gornik O, et al. 2010. Common aberrations from the normal human plasma *N*-glycan profile. *Glycobiology* 20:970–75
40. Knezevic A, Gornik O, Polasek O, Pucic M, Redzic I, et al. 2010. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma *N*-glycans. *Glycobiology* 20:959–69
41. Stanta JL, Saldova R, Struwe WB, Byrne JC, Leweke FM, et al. 2010. Identification of *N*-glycosylation changes in the CSF and serum in patients with schizophrenia. *J. Proteome Res.* 9:4476–89
42. Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. 2008. GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics* 24:1214–16
43. Artemenko NV, Campbell MP, Rudd PM. 2010. GlycoExtractor: a web-based interface for high throughput processing of HPLC-glycan data. *J. Proteome Res.* 9:2037–41
44. Vercoutter-Edouart AS, Slomianny MC, Dekeyzer-Beseme O, Haeuw JF, Michalski JC. 2008. Glycoproteomics and glycomics investigation of membrane *N*-glycosylproteins from human colon carcinoma cells. *Proteomics* 8:3236–56
45. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM. 2006. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *J. Proteome Res.* 5:1792–802
46. Dai Z, Fan J, Liu Y, Zhou J, Bai D, et al. 2007. Identification and analysis of α1,6-fucosylated proteins in human normal liver tissues by a target glycoproteomic approach. *Electrophoresis* 28:4382–91
47. Geyer H, Geyer R. 2006. Strategies for analysis of glycoprotein glycosylation. *Biochim. Biophys. Acta* 1764:1853–69
48. Yang Z, Harris LE, Palmer-Toy DE, Hancock WS. 2006. Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients. *Clin. Chem.* 52:1897–905
49. Wuhrer M, de Boer AR, Deelder AM. 2009. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrom. Rev.* 28:192–206
50. Calvano CD, Zambonin CG, Jensen ON. 2008. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J. Proteomics* 71:304–17
51. Sparbier K, Wenzel T, Kostrzewa M. 2006. Exploring the binding profiles of ConA, boronic acid and WGA by MALDI-TOF/TOF MS and magnetic particles. *J. Chromatogr. B* 840:29–36
52. Wollscheid B, Bausch-Fluck D, Henderson C, O'Brien R, Bibel M, et al. 2009. Mass-spectrometric identification and relative quantification of *N*-linked cell surface glycoproteins. *Nat. Biotechnol.* 27:378–86

53. Catalina MI, Koeleman CA, Deelder AM, Wuhrer M. 2007. Electron transfer dissociation of *N*-glycopeptides: loss of the entire *N*-glycosylated asparagine side chain. *Rapid Commun. Mass Spectrom.* 21:1053–61
54. Alley WR Jr, Mechref Y, Novotny MV. 2009. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun. Mass Spectrom.* 23:161–70
55. Zaia J. 2010. Mass spectrometry and glycomics. *OMICS* 14:401–18
56. Guillard M, Gloerich J, Wessels HJ, Morava E, Wevers RA, et al. 2009. Automated measurement of permethylated serum *N*-glycans by MALDI–linear ion trap mass spectrometry. *Carbohydr. Res.* 344:1550–57
57. Snovida SI, Perreault H. 2007. A 2,5-dihydroxybenzoic acid/*N*,*N*-dimethylaniline matrix for the analysis of oligosaccharides by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 21:3711–15
58. Ciucanu I, Kerek F. 1984. A simple rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* 131:209–17
59. Lei M, Mechref Y, Novotny MV. 2009. Structural analysis of sulfated glycans by sequential double-permethylation using methyl iodide and deuteromethyl iodide. *J. Am. Soc. Mass Spectrom.* 20:1660–71
60. Jang KS, Kim YG, Gil GC, Park SH, Kim BG. 2009. Mass spectrometric quantification of neutral and sialylated *N*-glycans from a recombinant therapeutic glycoprotein produced in the two Chinese hamster ovary cell lines. *Anal. Biochem.* 386:228–36
61. An HJ, Kronewitter SR, de Leoz ML, Lebrilla CB. 2009. Glycomics and disease markers. *Curr. Opin. Chem. Biol.* 13:601–7
62. Goldman R, Ressom HW, Varghese RS, Goldman L, Bascug G, et al. 2009. Detection of hepatocellular carcinoma using glycomic analysis. *Clin. Cancer Res.* 15:1808–13
63. Kam RK, Poon TC, Chan HL, Wong N, Hui AY, et al. 2007. High-throughput quantitative profiling of serum *N*-glycome by MALDI-TOF mass spectrometry and *N*-glycomic fingerprint of liver fibrosis. *Clin. Chem.* 53:1254–63
64. Barkauskas DA, An HJ, Kronewitter SR, de Leoz ML, Chew HK, et al. 2009. Detecting glycan cancer biomarkers in serum samples using MALDI FT-ICR mass spectrometry data. *Bioinformatics* 25:251–57
65. Liu Y, Palma AS, Feizi T. 2009. Carbohydrate microarrays: key developments in glycobiology. *Biol. Chem.* 390:647–56
66. Ong SE, Kratchmarova I, Mann M. 2003. Properties of ^{13}C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* 2:173–81
67. Xie Y, Liu J, Zhang J, Hedrick JL, Lebrilla CB. 2004. Method for the comparative glycomic analyses of *O*-linked, mucin-type oligosaccharides. *Anal. Chem.* 76:5186–97
68. Xia B, Feasley CL, Sachdev GP, Smith DF, Cummings RD. 2009. Glycan reductive isotope labeling for quantitative glycomics. *Anal. Biochem.* 387:162–70
69. Uematsu R, Furukawa J, Nakagawa H, Shinohara Y, Deguchi K, et al. 2005. High throughput quantitative glycomics and glycoform-focused proteomics of murine dermis and epidermis. *Mol. Cell Proteomics* 4:1977–89
70. Orlando R, Lim JM, Atwood JA 3rd, Angel PM, Fang M,. 2009. IDAWG: metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J. Proteome Res.* 8:3816–23
71. Goldberg D, Bern M, Li B, Lebrilla CB. 2006. Automatic determination of *O*-glycan structure from fragmentation spectra. *J. Proteome Res.* 5:1429–34
72. Goldberg D, Sutton-Smith M, Paulson J, Dell A. 2005. Automatic annotation of matrix-assisted laser desorption/ionization *N*-glycan spectra. *Proteomics* 5:865–75
73. Maass K, Ranzinger R, Geyer H, von der Lieth CW, Geyer R. 2007. "GlycO-Peakfinder"—de novo composition analysis of glycoconjugates. *Proteomics* 7:4435–44
74. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, et al. 2008. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* 7:1650–59
75. Apte A, Meitei NS. 2010. Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. *Methods Mol. Biol.* 600:269–81

76. Gaucher SP, Morrow J, Leary JA. 2000. STAT: a saccharide topology analysis tool used in combination with tandem mass spectrometry. *Anal. Chem.* 72:2331–36
77. Lapadula AJ, Hatcher PJ, Hanneman AJ, Ashline DJ, Zhang H, et al. 2005. Congruent strategies for carbohydrate sequencing. 3. OSCAR: an algorithm for assigning oligosaccharide topology from MS^n data. *Anal. Chem.* 77:6271–79
78. Ethier M, Saba JA, Spearman M, Krokhin O, Butler M, et al. 2003. Application of the StrOligo algorithm for the automated structure assignment of complex *N*-linked glycans from glycoproteins using tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 17:2713–20
79. Song EH, Pohl NL. 2009. Carbohydrate arrays: recent developments in fabrication and detection methods with applications. *Curr. Opin. Chem. Biol.* 13:626–32
80. Gornik O, Lauc G. 2008. Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* 25:267–78
81. Tateno H, Nakamura-Tsuruta S, Hirabayashi J. 2007. Frontal affinity chromatography: sugar-protein interactions. *Nat. Protoc.* 2:2529–37
82. Westerlund-Wikström B, Korhonen TK. 2005. Molecular structure of adhesin domains in *Escherichia coli* fimbriae. *Int. J. Med. Microbiol.* 295:479–86
83. Rüdiger H, Gabius HJ. 2001. Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconj. J.* 18:589–613
84. Pilobello KT, Krishnamoorthy L, Slawek D, Mahal LK. 2005. Development of a lectin microarray for the rapid analysis of protein glycopatterns. *ChemBioChem* 6:985–89
85. Chen P, Liu Y, Kang X, Sun L, Yang P, et al. 2008. Identification of *N*-glycan of α-fetoprotein by lectin affinity microarray. *J. Cancer Res. Clin. Oncol.* 134:851–60
86. Hsu KL, Mahal LK. 2006. A lectin microarray approach for the rapid analysis of bacterial glycans. *Nat. Protoc.* 1:543–49
87. Propheter DC, Hsu KL, Mahal LK. 2010. Fabrication of an oriented lectin microarray. *Chembiochem* 11:1203–7
88. Krishnamoorthy L, Bess J Jr, Preston AB, Nagashima K, Mahal LK. 2009. HIV-1 and microvesicles from T cells share a common glycome, arguing for a common origin. *Nat. Chem. Biol.* 5:244–50
89. Uchiyama N, Kuno A, Tateno H, Kubo Y, Mizuno M, et al. 2008. Optimization of evanescent-field fluorescence-assisted lectin microarray for high-sensitivity detection of monovalent oligosaccharides and glycoproteins. *Proteomics* 8:3042–50
90. Tateno H, Uchiyama N, Kuno A, Togayachi A, Sato T, et al. 2007. A novel strategy for mammalian cell surface glycome profiling using lectin microarray. *Glycobiology* 17:1138–46
91. Pilobello KT, Slawek D, Mahal LK. 2007. A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome. *Proc. Natl. Acad. Sci. USA* 104:11534–39
92. Kuno A, Uchiyama N, Koseki-Kuno S, Ebe Y, Takashima S, et al. 2005. Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat. Methods* 2:851–56
93. Chen S, Zheng T, Shortreed MR, Alexander C, Smith LM. 2007. Analysis of cell surface carbohydrate expression patterns in normal and tumorigenic human breast cell lines using lectin arrays. *Anal. Chem.* 79:5698–702
94. Zheng T, Peelen D, Smith LM. 2005. Lectin arrays for profiling cell surface carbohydrate expression. *J. Am. Chem. Soc.* 127:9982–83
95. Tao SC, Li Y, Zhou J, Qian J, Schnaar RL, et al. 2008. Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology* 18:761–69
96. Hsu KL, Pilobello KT, Mahal LK. 2006. Analyzing the dynamic bacterial glycome with a lectin microarray approach. *Nat. Chem. Biol.* 2:153–57
97. Gupta G, Surolia A, Sampathkumar SG. 2010. Lectin microarrays for glycomic analysis. *OMICS* 14:419–36
98. Jin S, Cheng Y, Reid S, Li M, Wang B. 2010. Carbohydrate recognition by boronolectins, small molecules, and lectins. *Med. Res. Rev.* 30:171–257
99. Manimala JC, Roach TA, Li Z, Gildersleeve JC. 2007. High-throughput carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity problems. *Glycobiology* 17:17–23C

100. Linman MJ, Taylor JD, Yu H, Chen X, Cheng Q. 2008. Surface plasmon resonance study of protein-carbohydrate interactions using biotinylated sialosides. *Anal. Chem.* 80:4007–13
101. Zhao J, Patwa TH, Pal M, Qiu W, Lubman DM. 2009. Analysis of protein glycosylation and phosphorylation using liquid phase separation, protein microarray technology, and mass spectrometry. *Methods Mol. Biol.* 492:321–51
102. Chen S, LaRoche T, Hamelinck D, Bergsma D, Brenner D, et al. 2007. Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nat. Methods* 4:437–44
103. Qiu Y, Patwa TH, Xu L, Shedden K, Misek DE, et al. 2008. Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. *J. Proteome Res.* 7:1693–703
104. Yue T, Goldstein IJ, Hollingsworth MA, Kaul K, Brand RE, et al. 2009. The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich array. *Mol. Cell Proteomics* 8:1697–707
105. Wu YM, Nowack DD, Omenn GS, Haab BB. 2009. Mucin glycosylation is altered by pro-inflammatory signaling in pancreatic-cancer cells. *J. Proteome Res.* 8:1876–86
106. Forrester S, Hung KE, Kuick R, Kucherlapati R, Haab BB. 2007. Low-volume, high-throughput sandwich immunoassays for profiling plasma proteins in mice: identification of early-stage systemic inflammation in a mouse model of intestinal cancer. *Mol. Oncol.* 1:216–25
107. Nilsson P, Paavilainen L, Larsson K, Odling J, Sundberg M, et al. 2005. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* 5:4327–37
108. Aoki-Kinoshita K. 2008. An introduction to bioinformatics for glycomics research. *PLoS Comput. Biol.* 4:e1000075
109. Frank M, Schloissnig S. 2010. Bioinformatics and molecular modeling in glycobiology. *Cell Mol. Life Sci.* 67:2749–72
110. Doubet SAP. 1992. CarbBank. *Glycobiology* 2:505
111. Porter A, Yue T, Heeringa L, Day S, Suh E, et al. 2010. A motif-based analysis of glycan array data to determine the specificities of glycan-binding proteins. *Glycobiology* 20:269–80
112. Gupta R, Brunak S. 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* 2002:310–22
113. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, et al. 1998. NetOglyc: prediction of mucin type *O*-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* 15:115–30
114. Julenius K. 2007. NetCGlyc 1.0: prediction of mammalian *C*-mannosylation sites. *Glycobiology* 17:868–76
115. York WS, Kochut, KJ, Miller JA. 2009. Integration of glycomics knowledge and data. In *Handbook of Glycomics*, ed. RD Cummings, JM Pierce, pp. 179–95. New York: Academic
116. Ito H, Kuno A, Sawaki H, Sogabe M, Ozaki H, et al. 2009. Strategy for glycoproteomics: identification of glyco-alteration using multiple glycan profiling tools. *J. Proteome Res.* 8:1358–67
117. Narimatsu H, Sawaki H, Kuno A, Kaji H, Ito H, et al. 2010. A strategy for discovery of cancer glycobiomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J.* 277:95–105

Annual Review of
Analytical Chemistry

Volume 4, 2011

Contents

Indexes

Errata

An online log of corrections to the *Annual Review of Analytical Chemistry* articles may be found at http://arjournals.annualreviews.org/errata/anchem