View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DSpace@MIT

NIH Public Access
Author Manuscript
*Electrophoresis*. Author manuscript; available in PMC 2013 August 14.

Published in final edited form as:
*Electrophoresis*. 2012 March ; 33(5): 797–814. doi:10.1002/elps.201100231.

# Harnessing glycomics technologies: integrating structure with function for glycan characterization

**Luke N. Robinson**[1], **Charlermchai Artpradit**[2], **Rahul Raman**[1], **Zachary H. Shriver**[1], **Mathuros Ruchirawat**[2,3], and **Ram Sasisekharan**[1,*]

[1]Department of Biological Engineering, Harvard-MIT Division of Health Sciences & Technology and Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA [2]Program in Applied Biological Sciences: Environmental Health, Chulabhorn Graduate Institute, Bangkok, Thailand [3]Laboratory of Environmental Toxicology, Chulabhorn Research Institute, Bangkok, Thailand

## Abstract

Glycans, or complex carbohydrates, are a ubiquitous class of biological molecules which impinge on a variety of physiological processes ranging from signal transduction to tissue development and microbial pathogenesis. In comparison to DNA and proteins, glycans present unique challenges to the study of their structure and function owing to their complex and heterogeneous structures and the dominant role played by multivalency in their sequence-specific biological interactions. Arising from these challenges, there is a need to integrate information from multiple complementary methods to decode structure-function relationships. Focusing on acidic glycans, we describe here key glycomics technologies for characterizing their structural attributes, including linkage, modifications, and topology, as well as for elucidating their role in biological processes. Two cases studies, one involving sialylated branched glycans and the other sulfated glycosaminoglycans, are used to highlight how integration of orthogonal information from diverse datasets enables rapid convergence of glycan characterization for development of robust structure-function relationships.

### Keywords

function; glycans; glycomics; structure

## 1. Introduction

Glycans, largely located at the interface between the cell surface and extracellular matrix, modulate cellular functions through specific, non-covalent interactions with growth factors, cytokines, enzymes, receptors, and other proteins present in this environment [1]. Glycans, depending on their structure and location, positively or negatively modulate cell growth and development [1–5], immunity [6–8], cell-cell interactions [9–11], tumor growth and metastasis [12–17], and microbial pathogenesis [18–20], among other processes. In addition to affecting physiology by binding proteins, glycans also affect the structural properties of the molecules to which they are conjugated (either covalently or non-covalently), such as by aiding folding, increasing solubility and stability, and influencing the clearance rate of

[*]Address correspondence to: Professor Ram Sasisekharan, Ph.D., 77 Massachusetts Avenue E25-519, Cambridge, MA 02139. Fax: 617-258-9409; rams@mit.edu.

**Conflict of interest:** None.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

proteins [21]. To highlight but one example, sialylation of erythropoietin is known to profoundly affect its *in vivo* half-life. Hypersialylation by, for instance, adding additional glycosylation sites results in creation of a pharmaceutical agent that can be given on a less frequent (more convenient) dosing schedule due to its extended residence in the circulation [22]. Conversely, desialylation of erythropoietin results in a molecule that has a very rapid disappearance from the bloodstream, but one that crosses into the central nervous system and has marked neuroprotective effects [23].

As with the erythropoietin example, glycans are typically found as covalent conjugates either to lipids (glycolipids) or to proteins (glycoproteins), although unconjugated forms are known to exist (e.g., hyaluronan). Broadly, glycans can be classified as either branched or linear (Figure 1). Branched glycans are present as *N*- and *O*-linked glycosylation on glycoproteins and on glycolipids, whereas linear glycans attached to proteins are mostly glycosaminoglycans (GAGs), which are polysaccharide chains made up of disaccharide units with differential sulfation and are typically *O*-linked to a protein core [1, 24]. In mammals, glycolipids are predominantly present as glycosphingolipids, which are characterized by the presence of branched or linear glycans that can be negatively charged (sialylated or sulfated) or neutral (no charged groups) [25].

Advancements in tools to decipher glycan fine structure and to define their physiological function have helped mature the field of glycomics, in part by addressing many of the challenges of glycan structure-function characterization (Table 1). Glycans have multiple potential structural elements, such as the branching pattern, linkage between monosaccharides, and site of attachment, which result in higher complexity and diversity than is typically found in proteins and nucleic acids. This structural diversity is further complicated by glycan biosynthesis, a non-proofreading, non-template-driven process which involves the concerted action of a variety of glycosyltransferases and glycosidases with tissue-, developmental- and metabolic-dependent expression and activity [2, 24, 26]. In terms of function, glycans are presented in physiological systems as an ‘ensemble’ of similar structures in which glycan-protein interactions achieve high specificity and high affinity through multivalent interactions (avidity) [27, 28]. Compared to proteins, capturing these key properties of glycan presentation complicates the ability to accurately and specifically measure glycan binding events. In fact, one of the key challenges in glycomics is to be able to address this notion that an ensemble of structures together modulate a biological function, and that all information is not necessarily encoded by sequence alone, with there being both finer and coarser determinants to activity.

Substantial progress in addressing many of the challenges of glycomics has been made in key areas, including analytical technologies [29–38], synthesis methods [39–43], glycan functional tools (including knockout mice) [44–46], specialized databases that store glycomics-specific information and computational tools [47–50]. Advances in many of these areas have enabled the scientific community to probe glycan structure-function relationships in unprecedented ways. Indeed, a variety of highly sensitive analytical methods, including mass spectrometry (MS), high performance liquid chromatography (HPLC), LC/MS, capillary electrophoresis (CE), nuclear magnetic resonance (NMR) spectroscopy, and lectin arrays, provide means to probe many fine structural attributes of carbohydrates (Table 2). However, each particular methodology has its strengths and weaknesses, especially in terms of the information content (i.e., structural attributes) which can be readily measured from the approach. The use of multiple methodologies for structural and functional glycomics overcomes these constraints but often leads to large, complex and diverse datasets, necessitating the integration of results from multiple complementary tools. Moreover, as new technologies have opened additional avenues for glycobiology-driven discovery, the need for integrating diverse datasets has expanded, especially given the non-binary

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

relationship between sequence and function. Thus, this review will focus on how integration of information content across multiple lines of inquiry has enabled - (1) convergence in glycan characterization in a more rapid manner than is possible using any one individual method as well as (2) development of robust structure-function relationships.

Numerous reviews have expertly outlined the use of particular analytical techniques for glycan structural determination, including MS [31, 51–58], NMR [32, 59–63], and lectins [33, 64–73]. To complement these reviews, we focus here on the role of integrated efforts towards determination of structure and structure-function relationships. This is best exemplified, in our opinion, by focusing on the class of acidic glycans where multiple recent advances have provided key insights into structure and biology and highlight the necessity of integrating information from multiple datasets to elucidate structure-function relationships. Within this review, we will focus on three major groups of acidic glycans: (1) sialylated glycans that often are present as lipid or protein conjugates (*N*- and *O*-linked), (2) sulfated/phosphorylated glycoconjugates, and (3) extended sulfated glycans *O*-linked to a protein core within the context of proteoglycans.

## 2. Structural elucidation of acidic glycans

Biological functions of glycans are mediated largely by their interactions with glycan-binding proteins (GBPs), and the specificity of these interactions is in turn regulated by the structural recognition of glycan motifs in the context of multivalent glycan-GBP interactions. Much of the structural variation on cells exists in the sugar sequence of the terminal residues and with modifications such as acetylation, sulfation, and phosphorylation [74–76]. Acidic motifs, represented by carboxylated (sialic acids and uronic acids), sulfated (GAGs and branched glycans with sulfation), and phosphorylated substituents (Figure 2), are often, but not always, located on the terminal sugars of glycan chains and play a dominant role in the regulation of many diverse biological activities. For example, as a result of their typical terminal location and negative charge, sialylated motifs extensively participate in varied physiological and pathologies processes, including immune surveillance, cell adhesion, proliferation, apoptosis, development, and as attachment ligands for a range of pathogenic bacteria, protozoa, fungi, and viruses (reviewed in [10, 77–79]). There are more than 50 structural variations of the sialic acid monosaccharide in nature, and further layers of complexity can be considered in terms of varied linkages, neighboring residues, presentation (e.g., branching diversity), and spatial organization along cell surfaces [80]. Other anionic modifications, including sulfation and phosphorylation, also play a pivotal role in the regulation of many biological processes. This can perhaps be best illustrated with GAGs, in which subtle changes in the amount and position of sulfates provide a “sulfation code” that largely governs specificity of protein-GAG interactions [81–84].

The diverse functions mediated by acidic glycans underscore the need to elucidate the structure-function relationships of this class of carbohydrates, and in particular, understand the fine structural features which govern their specificity of interactions. Measurement of acidic glycans, however, by “typical” analytical methods, even those that have proven useful for global glycomics analyses, is often complicated by the fact that acidic groups are labile in many sample workup and analytical methods, such as depolymerization (e.g., for CE- and MS-based compositional analysis [85]), derivatization (e.g., acidic reductive amination for fluorescence detection [86]) and analysis (e.g., mass spectrometry) [87, 88]. A second, related issue is that acidic groups on glycans strongly bind to monovalent and divalent cations, such as calcium and sodium, which further complicates their analysis by MS and NMR-based techniques through the creation of multiple adducts with different charge states. Moreover, their highly charged character limits the use of traditional chromatography, such

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



as reversed phase HPLC, which is the mode of choice for many advanced hyphenated techniques such as LC/MS.

## 2.1 Mass spectrometry-based approaches

Paralleling proteomics and lipidomics, MS has emerged as a corner stone for high throughput and sensitive structural characterization of glycans. As in the case in other –omics fields, combining electrospray ionization (ESI) with matrix-assisted laser desorption ionization (MALDI) often provides complementary information and enables a broader coverage of complex samples than either method alone. Acidic groups on glycans, however, can undergo significant fragmentation and loss of their acidic moieties during typical MALDI-MS conditions [87–89]. The acidic proton of anionic glycans also substantially exchanges with alkali metal cations, leading to the formation of metal adducts which increase the complexity of MS spectra as well as reduce signal intensity [90]. Moreover, acidic glycans analyzed in the positive-ion mode are usually suppressed due to their low ionization efficiency compared to their neutral counterparts. As a result, negative-ion mode is typically used, however, a loss of sensitivity occurs if negative-ion mode is exclusively used for glycan mixtures containing neutral and acidic carbohydrates, making quantification difficult [91].

Many of the challenges associated with MALDI-MS analysis of acidic glycans have been addressed with different strategies, including stabilizing acidic groups and improving ionization/sensitivity. The labile nature of acidic glycans can be partially overcome using matrices for softer ionization [92–95], atmospheric-pressure conditions [96], and additives such as basic peptides [97] or spermine [98]. The use of nafion [99] or cation exchange resins can also be effectively used to remove complicating counter ions, thus improving sensitivity and easing analysis of spectra.

Derivatization of native glycans also represents an effective strategy to solve issues with acidic carbohydrates. However, due to sample loss that frequently accompanies workup steps, derivatization may not be appropriate for certain samples containing acidic glycans due to their low abundance. Notably, derivatization methods such as permethylation improve sensitivity, enable the simultaneous analysis of both neutral and acidic glycans in positive-ion mode, increase the useful information that can be obtained from fragmentation experiments, and improve the accuracy of quantification [100, 101]. While permethylation is effective in stabilizing sialylated glycans, the method (in its most routinely used form developed by Ciucanu and Kerek [102]) causes the undesired effect of removing some acidic substituents, including sulfates and phosphates [103]. Thus, for many complex glycan mixtures that contain diverse acidic glycan motifs, permethylation will cause the loss of structural information, potentially destroying critical motifs. Other derivatization techniques, such as modified permethylation [104], methyl ester formation [90], amidation [105] and newer methods such as double permethylation [106], have helped address many of the limits of “typical” permethylation techniques, though these methods may require more involved processing steps and may lead to greater sample losses and accordingly, sample bias.

ESI-MS provides an alternative approach with the notable advantages of less fragmentation (softer ionization) and improved ionization efficiency for acidic glycans, thus making it a preferred analytical platform for analysis of many acidic glycans, particularly sulfated oligosaccharides [89, 107, 108]. However, glycan sequencing by ESI-MS/MS of native samples run in negative mode typically provides incomplete linkage-specific cleavage ions [109, 110]. Introduction of newer fragmentation techniques, including electron capture dissociation, have proven useful for the analysis of glycoproteins/peptides, particularly ones containing acidic modifications. Permethylation of samples, in turn, may be used to improve information content from fragmentation experiments, though limitations such as removal of

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

sulfates and phosphates remain. ESI-MS also leads to the formation of multiply charged ions, causing increased complexity in spectra analysis and reduced sensitivity. Due to the higher sensitivity of ESI-MS to the presence of sample contaminants such as salts and detergents (which are often used to extract glycoconjugates at the cell surface), experiments and workup strategies are limited to those which provide sufficient sample cleanup.

In terms of the wealth of information content associated with the technique, its sensitivity, and its ability to link both structural determination and quantification, MS has been and will likely continue to be one of the primary methods employed to analyze glycan structures. However, in our opinion, despite the advances that have been made, MS-related technologies require supplementation with other techniques in the context of a larger, integrated approach. This is, at least in part, due to the fact that assignment of glycosidic linkage and isomeric position of acidic groups, such as sulfates, is very difficult using MS strategies by themselves. Fortunately, the field has advanced a series of techniques that enable this analysis very precisely. Use of orthogonal analytical techniques enables one to arrive at detailed structural information in a more rapid, unbiased, and accurate manner than is possible through the use of any single analytical methodology.

## 2.2 NMR

NMR is a powerful technology for the characterization of many structural attributes of glycans. $^{1}$H and $^{13}$C chemical shifts of carbohydrate samples can be used to define features such as composition, linkage, anomeric configuration, and presence of chemical modifications [32, 60, 62]. Notably, NMR is non-destructive in nature, and thus sample material can be recovered for other subsequent structural or functional studies. While NMR can be effectively used for detailed and complete structural characterization of smaller oligosaccharides, such an analysis requires a high level of purity (>95%), which may be challenging or cause the loss of minor components during separation schemes. In such cases, NMR is an efficient tool to obtain quantitative properties and constraints on glycan mixtures with relatively small amounts of material. For example, $^{1}$H-NMR with as little as 100 micrograms of sample material has been used to characterize the ratio of specific linkages and branching patterns as well as the presence and absence of specific sugars (with their linkages) [111], information that, by MS methods alone, would require involved MS/MS techniques with complex peak analysis. To assist researchers in the assignment of glycan structures and features based on NMR data, the characteristic NMR chemical shifts and coupling constants of various glycans in literature have been compiled in accessible databases such as at the Glycosciences.de portal (http://www.glycosciences.de/sweetdb/) [112] and CASPER (http://www.casper.organ.su.se/casper/) [113–115], thus improving the accessibility of NMR as a tool for glycoscientists.

## 2.3 Lectins and glycan-modifying enzymes

The specificity of glycan-modifying enzymes and lectins make them useful tools for structural characterization studies as well as to directly aid other analytical techniques by providing an additional layer of information content. For example, one approach to decipher linkage information has been to employ exoglycosidases that specifically cleave monosaccharides with defined linkages from the nonreducing end of a glycan. The HPLC chromatographic profile of the parent glycan sample and the shifts in peaks which result from treatment with the exoglycosidases are matched with reference chromatographic profiles of known glycan structures using software tools to derive the exact structure of glycans in the sample [116–118]. Utilizing lectins for affinity-based chromatographic and electrophoretic methods has proved useful due to their unique selectivity and resolution in separation/fractionation studies as well as for enrichment of glycans with specific motifs which may be otherwise difficult to characterize, such as specific sialylated glycan and

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

glycoprotein populations [119–121]. In the case of GAGs, the selectivity of GAG-degrading enzymes, such as heparinases, chondroitinases, etc., has proven to be a key tool for the controlled and precise depolymerization or modification of GAG chains, thus increasing the information content from analytical assays. Indeed, informatics-based methods using GAG-degrading enzymes coupled with analytical techniques have been developed to capture the information density of GAGs and to enable the application of data from a combination of tools as constraints to sequence GAGs [122, 123].

Technologies incorporating lectins have been developed for directly probing structural attributes of glycan samples based on the glycan determinants (motifs) specifically recognized by lectins. Recent increased availability of recombinant lectins as well as knowledge of their binding specificity through glycan array studies have helped facilitate development of lectin microarrays, which contain a panel of lectins immobilized at high density to probe binding of carbohydrate analytes in high throughput (reviewed in [67–69, 124]). In addition to array platforms, lectins have been effectively used for probing biological samples to define the presence of specific glycan moieties. Importantly, these techniques enable analysis of glycans within their biological contexts. The functions of glycans depend in part on their physiological localization and context for access to proximal molecules, and thus lectin staining of cells and tissues helps decode glycan structure-function relationships by defining differences in localization as opposed to strictly structural composition after isolation [125, 126]. Indeed, this fact has been exploited in pathology, where reagents used to identify pathologic conditions include some that recognize specific glycan or glycoconjugate structures [127–130].

### 2.4 Databases and computational tools

In addition to experimental advances, several databases and computational tools have been developed to facilitate assignment of the most likely set of structures or structural attributes that satisfy one or more analytical datasets (as summarized in recent reviews and book chapters [47, 112, 131–137]). In the case of a typical high throughput dataset such as a MALDI-MS profile of glycans isolated from cells/tissues, tools such as GlycoMod [138] and Cartoonist [139] have been developed to assign a set of compositions and/or glycan topologies for a given mass peak based on biosynthetic constraints.

The next level of characterization involves matching the fragmentation patterns of the parent ion from the $MS^2$ (and possibly higher order $MS^3$ to $MS^n$) data to reference datasets and deducing the most likely glycan structures based on these data. The common reference datasets are those derived from the theoretical fragmentation of known glycan structures stored in different glycan structure databases [50, 135, 140]. Examples of tools developed for this purpose include Glyco-Search-MS (http://www.glycosciences.de/sweetdb/start.php?action=form_ms_search) [141] and GlycoWorkbench (http://www.glycoworkbench.org/) [142, 143]. Other approaches that have been utilized to create reference datasets include generation of possible theoretical fragments or compositions from a parent mass ion. The set of compositions is then used to generate possible glycan structures (using biosynthesis rules in some cases) whose theoretical fragmentation pathways are matched with the $MS^n$ data tree. Reference datasets have also been generated from experimental fragmentation profiles of well-characterized glycan structures or oligosaccharide fragments [144, 145]. The STAT [146], StrOligo [147] and OSCAR [148] programs use these approaches.

Enzymes involved in glycan modification have been annotated and classified in many databases including GlycoGene DataBase (GGDB) (http://riodb.ibase.aist.go.jp/rcmg/ggdb/), Carbohydrate-Active enZYme Database (CAZy) (http://www.cazy.org/), Kyoto Encyclopedia for Genes and Genomes (KEGG) (http://www.genome.jp/kegg/glycan/GT.html) and Consortium for Functional Glycomics (CFG) (http://

NIH-PA Author Manuscript

www.functionalglycomics.org/glycomics/molecule/jsp/glycoEnzyme/geMolecule.jsp). Detailed information resources on plant, animal and microbial lectins have been captured and annotated in many databases including Lectines 3D (http://www.cermav.cnrs.fr/lectines/), genomics resources for animal lectins (http://www.imperial.ac.uk/research/animallectins/), CFG GBP molecule pages (http://www.functionalglycomics.org/glycomics/molecule/jsp/gbpMolecule-home.jsp), and SugarBindDB (http://sugarbind.expasy.org/sugarbind/). The availability of these resources facilitate assignment of glycan structures based on their biosynthetic pathways and also screening or isolating glycans based on the binding specificity of specific motifs contained in these glycans to different lectins.

Taken together, it is clear that there have been significant advances in both experimental and computational approaches to address glycan structural complexity and provide insights into important structure-function relationships. Ultimately, given the set of analytical tools that have been developed, it is likely the case that the nuances of a specific biology will determine our ability to derive important insights, rather than the lack of effective tools to interrogate them.

## 3. Integrating structure with functional glycomics datasets

Characterization of glycan fine structural features by analytical methods is but the first (necessary) step towards decoding structure-function relationships for glycans. Several additional technologies described in the following section have recently come to the forefront in terms of critically informing glycan structure-function determination.

### 3.1 Glycogene analysis tools

The regulation of genes whose protein products are involved in glycan synthesis and glycan-protein interaction contributes to understanding of glycan structural diversity and function in biological systems. Efforts to understand expression of glycogenes are thus an important facet of glycomics research [45, 149]. Commercially available gene microarrays have been observed to have incomplete representation, poor annotation and restricted sensitivity for low abundance transcripts that are commonly involved in glycan biosynthesis [149–152]. To address these limitations, microarrays specific for understanding the transcriptional regulation of the glycome have been developed [150, 151]. Such tools enable clear and accurate signal detection of genes representing GBPs and glycoenzymes, including enzymes that encode for transfer and removal of acidic glycan motifs, such as sulfotransferases, sialyltransferases, and sulfatases. Gene manipulation tools, such as transgenic and knockout animal models, have aided discernment of the role of glycan motifs in complex biology [46, 153–155]. Indeed, conditional disruption of a heparan sulfate (HS)--polymerizing gene demonstrates the essential role of HS in mammalian brain morphogenesis [156], and transgenic models disrupting activity of *GNE*, whose protein product catalyzes the first two steps of sialic acid biosynthesis, emphasized the fundamental role of sialylation for early development in mice and normal renal function [157, 158].

### 3.2 Synthetic glycans

Synthetic glycans are also a critical component towards precisely defining glycan-protein interactions through provision of chemically defined oligosaccharides for binding assays and other biological readouts [41–45, 159, 160]. Strategies for chemical synthesis of oligosaccharides, particularly those representing terminal motifs of *N*- and *O*-linked glycans and glycolipids, have seen significant advancements during the previous decade [159, 161], however the labile nature and diversity of acidic substituents substantially challenges purely chemical approaches to generate synthetic sialosides and sulfated oligosaccharides, including GAGs. Discovery of the genes involved in formation of acidic glycan structures,

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

including sulfotransferases and sialyltransferases, has enabled development of chemoenzymatic synthesis tools for defined glycan structures and offers a promising synthetic strategy [77, 162–168]. To facilitate appropriate presentation of synthetic oligosaccharides for investigation of glycan interactions with their protein partners, scaffolds have been engineered for polyvalent glycan presentation with diverse multivalent architectures: low-molecular weight molecules, polymers (e.g., polyglutamic acid and polyacrylamide), dendrimers, liposomes, proteins (i.e., neoglycoproteins), and more recently on glycan arrays [27, 169–172].

### 3.3 Glycan arrays

Glycan arrays, composed of a panel of oligosaccharides immobilized to a solid support to which dilutions of GBPs are applied and their interaction subsequently quantified, provide a powerful technology to characterize GBP specificity (reviewed in [173–180]). A major asset of the array platform is the capability to rapidly interrogate GBP recognition with potentially hundreds of structurally defined glycans simultaneously, offering rapid insight into functional roles of fine structural features of glycans. Indeed, binding studies of members of the galectin family assessed by glycan array demonstrated that galectins 1–3 have differential specificity for sialylated glycans governed by fine structural features including linkage, and these binding differences were reflected in cellular functional readouts of galectin activity [181]. Furthermore, arrays containing structurally defined GAGs have been developed, and studies with them have further supported the hypothesis that specific sulfation and acetylation patterns on GAG sequences encode information for governing interactions and, consequently, physiological function [82, 182–187].

Many GBPs achieve their specificity and affinity through multivalent interaction with glycans. Capturing the physiological avidity of such interactions necessitates the appropriate presentation of multiple GBPs (or multiple glycan-binding domains within a GBP) and multiple glycan motifs. Oligosaccharide density, spacing, and orientation as well as linker flexibility and length become key parameters towards optimizing array strategies to capture biologically meaningful binding events [176]. Indeed, multiple studies have suggested that particular formats of glycan presentation prevent binding events that have previously been validated in other assays [175, 176, 188]. In one study, concanavalin A (a plant lectin) showed equivalent binding to high- and low-affinity mannose oligosaccharides immobilized at high density but demonstrated binding only to the high-affinity receptor at lower ligand density [189]. As these studies suggest, it can be expected that there is no clear single set of conditions that provides a universal optimum for all GBPs. Validation studies, such as binding assays to obtain dissociation constants (e.g., surface plasmon resonance) and those using physiologically relevant biological readouts (e.g., interaction with cells), are necessary to confirm predictions made from glycan array analyses.

### 3.4 Integration across techniques

Integration of these techniques (functional genetics, glycan synthesis, and glycan arrays) with analytical approaches is critical for a number of reasons (as summarized in recent reviews [29, 30, 38, 44, 45, 133, 190–198]). In general, as described above, multiple glycan sequences can bind to a given protein. Additional aspects, beyond sequence alone, govern the specificity and biology of glycans, including multivalency, glycan topology, presence of substituent groups, and other structural constraints. Moreover, while often genetic control of a signaling pathway can often be understood in a binary way, as evidenced in the utility of various gene manipulation techniques, glycan modulation of function is often more analog in nature.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Two case studies, one involving the binding of influenza A virus hemagglutinin (HA) to sialylated receptors on the surface of airway epithelial cells and the other describing the structural determination of a contaminant in heparin, highlight many of these concepts, and provide specific examples where integration of approaches provided robust elucidation of structure-function relationships.

## 4. Case studies

### 4.1 Receptor binding specificity of influenza virus

Influenza A virus interaction with sialylated glycan receptors has been studied for over five decades and thus serves as a rich example to highlight an integrated strategy for defining how fine structural features of sialylated glycans dictates interaction specificity, which in turn governs a biological function, namely infection and pathogenesis. Influenza A viruses exist commensally in wild aquatic bird populations without causing disease, thus making birds the natural reservoir for influenza strains, though viruses may intermittently spread through other species such as swine. As observed in the 1957, 1968, and the 2009 influenza pandemics, genetic reassortments of avian-derived viruses can lead to adaptation of virus such that they efficiently transmit between humans. Strains which demonstrate high human-to-human transmissibility in addition to high virulence have increased risk to cause severe disease outbreaks in people. Thus, one critical area with important scientific as well as medical implications to address in this field is to define the molecular changes that govern the shift from avian to human infectivity and transmissibility of influenza A viruses.

The infection process begins with attachment of viral HA to sialylated glycan receptors located on host cells, and as such, the HA-glycan interaction is a critical step regulating virus infectivity and selectivity. Biochemical studies of the HA-glycan interaction have identified the presence of glycans with terminal sialic acid (Neu5Ac) linked α2→6 to galactose (Neu5Ac α2→6 D-galactose [Gal], hereafter referred to as α2→6) to be a key feature for human-adapted HA, whereas avian-adapted viruses have preference for terminal sialic acid linked α2→3 to galactose (Neu5Acα2→3Gal, hereafter referred to as α2→3) [199–203]. This biochemical characterization along with studies describing co-crystal structures of HAs in complex with various glycan receptors [204–209] support the hypothesis that for an influenza virus to cross over from avian to human-adapted, the recognition specificity of its HA must switch preference from α2→3 glycans found on avian tissues to α2→6 glycans. Although this hypothesis explained transmission specificity for multiple viruses, cases were identified in which HAs representing different strains exhibited mixed α2→3/α2→6 binding specificity but showed differential transmission in a ferret animal model of human transmission, thereby confounding the correlation between α2→6 glycan binding specificity and efficient human transmission [210]. These results suggested HA binding preference defined by glycosidic linkage alone, namely α2→3 or α2→6, is not a sufficient determinant to differentiate human- from avian-adapted HA.

To more fully explain the classification of avian- versus human-adapted HA in the context of HA glycan binding specificity, application of additional tools was required to elucidate finer structural properties of sialylated glycan receptors beyond the terminal Neu5Ac-galactose glycosidic linkage. In a study by Chandrasekaran et al., multiple complementary methodologies which both affirm and inform each other were employed to define fine structural properties of sialylated glycan receptors and bridge these structures with biophysical and biochemical binding features of HA, the results of which demonstrated a specific glycan topology recognized by human- versus avian-adapted HA [211].

The upper respiratory tract serves as the primary target area for human-to-human transmission of influenza A virus. In order to characterize potential target glycan receptors

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

for influenza A virus, an analysis of the distribution and diversity of sialylated glycans found on cell surfaces of upper airway tissues and cells was performed. Structures of physiologically relevant sialylated glycan receptors were probed using the complementary techniques of lectin staining and MALDI-MS [211]. A matrix of lectins was applied to stain different tissue sections of upper respiratory tract, and glycan fine structure was analyzed and quantified using MALDI-MS /MS coupled with pretreatments with sialidases having different specificities. Incorporating complementary tools (i.e., lectins and glycosidases) to MS aided the identification of a diversity of multiantennary sialylated glycans with the presence of long oligosaccharide branches composed of multiple lactosamine repeats.

Structural information gathered from an MS-coupled approach informs the selection of glycan targets for use in functional assays as well as guides interpretation of functional readouts. Utilizing this strategy, a two-pronged methodology was employed to correlate physiologically relevant glycan structures with HA binding preference [211]. First, data-mining tools were applied to existing glycan array data of binding specificity of influenza A virus and recombinant HA from various avian- and human-adapted strains. This informatics-based approach identified patterns of structural features present (or absent) in glycan motifs which demonstrated binding or non-binding in array experiments. Abstraction of binding patterns led to rules indicating the importance of extension length on the non-reducing end of sialic acid to classify α2→6 binders. Importantly, much of the existing glycan array data was obtained from experiments utilizing only a single concentration of HA or titer of virus in the binding assay. Interpretation of binding results using a single concentration is limited and is generally framed in a binary "on" or "off" fashion [203, 212–214]. Such an analysis necessarily obscures finer differences in specificity and affinity, particularly when saturating concentrations are used.

To overcome these limitations as well as validate predictions generated by the data-mining approach, a robust and quantitative binding assay was developed to measure HA and virus specificity [215]. A panel of synthetic oligosaccharides was selected based on sialylated glycan motifs identified in the coupled-MS approach as well as from predictions generated by glycan array data mining. To accurately capture and compare specificity and affinity of multivalent HA-glycan interactions across different HAs, two key variables were addressed. First, to capture the avidity of HA-glycan interactions, recombinant HA was expressed as a trimer with further multivalent presentation facilitated by precomplexation of trimeric HA with primary and secondary antibodies, resulting in a locked spatial arrangement [215]. Such an approach fixed avidity effects across HA, thereby allowing for comparison of binding affinity across HAs of different subtypes. The second variable was quantifying relative binding affinity of HA to α2→3 and α2→6 linked glycan motifs. This was accomplished by generating a binding isotherm across serial dilution of precomplexed HA and calculation of an apparent dissociation constant ($K'_d$) to capture the relative binding affinities.

Designing binding assays that address these variables provided a framework to quantitatively assess fine differences of binding specificity of HA from diverse strains. Previous hypotheses were corroborated by demonstration that human-adapted HAs share binding preference for α2→6 glycans containing multiple lactosamine (or lactose) repeats, whereas avian-adapted HAs had high affinity for α2→3 glycans and α2→6 glycans containing short extensions.

The above analyses provided a direct connection between structural analyses of physiologically relevant sialylated glycan receptors from upper airways and the binding specificity of avian- and human-adapted HA in the context of relevant glycan motifs. The results suggested that, in addition to the glycosidic linkage between terminal sialic acid and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

penultimate galactose, the sugars extending beyond the penultimate galactose, and particularly the length of the extension, play a role in binding differentiation of avian- versus human-adapted HA. Molecular modeling analyses, utilizing existing co-crystal structures of HA complexed with different glycan receptors, corroborated previous results and revealed that a specific topology (termed *umbrella-like* topology) was adopted by α2→6 glycans containing multiple lactosamine repeats [211]. In contrast, a distinct topology with a cone-like shape was adopted by α2→3 and short α2→6 glycans. Cumulatively, these studies illustrate an integrated approach for generating a direct and specific structural rationale for binding specificity of avian-versus human-adapted HA (and viruses) to sialylated glycans, which in turn correlates with airborne transmissibility of viruses (Figure 3).

The set of studies described here provides tangible insight into an integrated framework for the structure-function study of acidic glycans. First, orthogonal techniques are typically necessary to probe finer structural details beyond sequence, such as glycosidic linkage, modifications, and conformation. This case study highlights the utility of coupling MS analysis with lectin staining of tissue sections (distribution of physiologically relevant glycans), glycosidase treatments (linkage), and molecular modeling (glycan conformation and topology). Furthermore, this case study exemplifies how glycan sequence alone can be insufficient to explain a glycan-mediated function, and that probing deeper simply by additional layers of MS and MS/MS analysis may not necessarily reveal the key structural determinants for a given activity. Indeed, the linking of structural results (MS-coupled experiments) with functional results (glycan array studies) was critical to inform and guide subsequent molecular analyses focused on structural and conformational aspects of extended versus short sialylated terminal motifs.

The importance of informatics tools to decipher complex datasets is highlighted in this case study. Several strains of HA (and virus) have been screened on glycan array platforms comprising of hundreds of distinct glycan motifs. The size of this dataset overwhelms manual interpretation methods, and thus informatics became a key tool to systematically analyze the large dataset in an integrated manner so as to make all data comparable for extracting useful information. Moreover, given that a multitude of related (or sometimes unrelated) glycan motifs show binding signals on the array, it was necessary to develop informatics tools to extract key glycan structural features shared by binders versus nonbinders for a given HA.

### 4.2 Heparin contamination

Heparin is a mixture of highly sulfated GAG chains isolated from a biological source, typically porcine intestinal mucosa, and is used clinically primarily as a prophylactic agent to prevent thrombosis as well as an initial treatment of established venous thrombosis. These heterogenous acidic polysaccharides are composed of 10–100 disaccharide repeat units of *N*-acetyl/sulfo glucosamine (1→4) linked to hexuronic acid. In its simplest form, the glucosamine (H) is *N*-acetylated and the uronic acid is β-D-glucuronic acid (G). Modifications to this basic unit include *N*- and *O*-sulfonation (6-*O* and 3-*O* sulfo groups are present in the glucosamine and 2-*O* sulfo groups in the uronic acid) as well as epimerization of β-D-glucuronic acid (G) to α-L-iduronic acid (I) [216]. Heparin has enjoyed widespread use as a medicinal agent; it and insulin are, on a unit basis, the most widely used medications in the clinic. Heparin is standard therapy in a number of situations, such as kidney dialysis and medical intervention of acute coronary events. In terms of worldwide production, most of the crude, partially purified, heparin is produced in China; in addition, China is a major locale for purification of heparin active pharmaceutical ingredient (API) from crude heparin.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Commercial heparins are derived from a number of sources, including porcine and bovine intestine and lung mucosa, though heparin sold in the United States is almost exclusively from porcine intestinal mucosa. Due to their biological source, as well as the employed isolation procedure, heparin primarily contains highly sulfated chains, with ~60–90% of the disaccharide units consisting of 2-*O* sulfo iduronic acid and 6-*O* sulfo, *N*-sulfo glucosamine (abbreviated $I_{2S}H_{NS,6S}$). In a typical polysaccharide chain of heparin, repeats of this trisulfated disaccharide unit are interrupted on occasion by other minor sequences, including the antithrombin III binding pentasaccharide sequence, a key sequence for heparin's anticoagulant activity. On average, heparin consists of chains which contain 2.4–2.8 sulfates per disaccharide; however, a combination of *polydispersity*, resulting in variable molecular weight, as well as *microheterogeneity*, resulting in structural (sequence) variability between chains, make heparin a complex pharmaceutical agent, one that orchestrates a range of biological activities.

In late 2007 and early 2008, clusters of serious allergic-type events were reported in patients undergoing hemodialysis who were receiving heparin. Efforts were made to identify the source of these allergic-type reactions, and these initial investigations ruled out many obvious causes, such as the presence of adventitious viral agents or the presence of greater levels of protein impurities in suspect as compared to non-suspect lots of the drug [217]. Analysis of intact heparin by analytical techniques, such as CE and one-dimensional NMR, demonstrated profile differences in suspect lots *versus* clean lots. However, while these techniques could be used to screen the heparin supply and provide some indication of the nature of the contamination, detailed structural work was complicated by the heparin's polydispersity and microheterogeneity to the point that structural elucidation of the contaminant was not possible using strategies that had worked in the past to identify structural signatures with heparin. Further complicating structural work was the fact that the contaminant was refractory to digestion with heparinases, which, as described in section 3, are often used to break heparin chains down into constituent parts for analysis [123, 218, 219].

Only through a combination of approaches was the contaminant ultimately identified as oversulfated chondroitin sulfate (OSCS) [220, 221], a complex polysaccharide mixture which had never before been observed in heparin (Figure 4A). In this case, to definitively identify OSCS required the use of multiple analytical techniques including multidimensional NMR, HPLC analysis, and MS. OSCS was found to have four sulfate groups per disaccharide unit, which is rarely, if ever, found in nature and is structurally distinct from other GAG impurities, such as dermatan sulfate. Consequently, it is unlikely that OSCS arose from a natural source and was probably added, either accidentally or deliberately, to heparin. Indeed, some suspect lots of heparin were found to contain approximately 30% of contaminant [221]. The advent of OSCS contamination in heparin has led to the introduction of new quality control tests, based on CE and NMR, which measure molecular level attributes of heparin and that can readily detect OSCS contamination if present at levels greater than 0.1–0.2%. However, OSCS is but one possible contaminant that could enter the heparin supply, given that numerous persulfonated polysaccharides have global anticoagulant activity.

Given heparin's structural complexity as well as its use as a critical drug in the hospital, a rigorous testing regimen will necessarily involve multiple tests. Such a testing strategy can be thought of as a series of filters, where suspect heparin is captured by at least one of the filters (i.e., fails testing specifications) and, accordingly, is removed from the supply chain. Thus, the analytical testing regimen, or protocol, should be designed such that will be some overlap in information content obtained from each test. Because of this overlap, through

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cross-correlation of results, it is possible to ensure the accuracy of testing results and increase the overall confidence that a sample which passes is truly free of contamination.

Without this type of approach, there is a real risk of missing potential impurities/ contaminants within heparin. For example, with regards to a $^1$H NMR-based test, its limitations should be expressly recognized. First, while this method can detect OSCS containing approximately four sulfates per disaccharide, if the degree of sulfation for OSCS is reduced, the prominent *N*-acetyl signal is lost behind that of heparin (Figure 4B). Another issue at present is that the current $^1$H NMR test as practiced may not capture the presence of other contaminants. For example, alginate sulfate, even at a level of 4%, is not detectable in the proton NMR spectrum (Figure 4C).

The above situation is even more acute for CE/HPLC-based methods which analyze intact heparin. CE/HPLC, which relies on absorbance at 210 nm, also cannot detect several potential contaminants. Again, referring the above examples, any polysaccharide, including alginate sulfate, which does not have appreciable absorbance at this wavelength would not be detected. Furthermore, any contaminant with a sulfate (charge) density close to that of heparin would migrate similarly to heparin and remain undetected. Clearly, there is the need to develop alternative testing strategies that incorporate molecular-level information and hence measure and account for heparin's heterogeneity, thus resolving potential contaminants.

Utilization and integration of the information provided by orthogonal methodologies avoids these pitfalls encountered with a single technique. In this context, a method like $^1$H NMR is not intended to capture *every* possible contaminant in heparin, and it is supplemented with additional analytical technologies, such as enzymatic digestion and HPLC analysis, that would detect contaminants invisible to NMR, including the aforementioned alginate sulfate.

## 5. Future of glycomics

Over the past several years, there has been a dramatic transformation in understanding the structure and biology of complex glycans owing to development of synthesis, analytical and informatics tools and technologies to study glycan structure-function relationships. These developments have changed the notion of glycomics from a complex research field to an area that is now readily accessible to a wide range of scientific disciplines including analytical chemistry, biochemistry, structural biology, immunology, microbiology, cancer research, computer science and informatics.

An important factor in opening up glycomics research to the broader scientific community is the accelerated development of databases and computational and informatics tools to acquire, integrate, annotate, mine and disseminate glycomics datasets such as analytical data, glycan array data and glycogene expression data, etc. Much of the earlier efforts in glycomics focused on structural characterization of glycans, development of glycan structure databases and computational tools to assist assignment of glycan structures from high-throughput analytical datasets. The development of these tools has advanced to a point where it is possible to obtain robust and detailed profiling of a majority of glycans isolated from cells, tissues and individual glycoproteins. As covered in this review, the acidic glycans pose unique challenges in terms of structural characterization and efforts are ongoing to improve the analytical tools for sensitive measurement of acidic glycans and integrate information across multiple tools (facilitated by informatics) to address these challenges.

An important future direction building on technologies to characterize glycan structure is to bridge the space of the glycome with functional glycome. Advances in chemical and

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

chemoenzymatic synthesis of glycans, neoglycoproteins and neoglycolipids that have led to development of glycan array platforms, and functional readouts in terms of glycogene expression and animal models have positioned the field to begin making this bridge. However, unlike DNA and proteins where structural specificity can be readily translated into sequence space, it is challenging to bridge the biophysical and biochemical specificity of glycan-protein interactions with the biology.

A typical screening of a GBP on glycan array platforms gives several glycan sequence motifs that show comparable high binding to that GBP. However, at the level of the three-dimensional structural interactions, these diverse glycan sequence motifs adopt a distinct shape or topology that optimally interacts with the GBP (Figure 5). For example, it has been demonstrated that binding of a GAG chain to a protein induces a distinct kink in the three-dimensional helical structure of GAG and the topological arrangement of sulfate groups to make optimal contacts with the protein is governed by the nature and extent of the kink [83]. The importance of glycan shape and conformational constraints induced upon protein binding has also been demonstrated for C-type lectins [188] and influenza A virus HA [211]. Therefore it is important to define specificity of glycan-protein interactions in context of structural motif and possible set of sequence motifs that would satisfy the structural constraints for optimal interaction with a GBP.

The aforementioned issues have motivated the development of computational tools for determining binding motifs from glycan array data [211, 222] and for three-dimensional modeling of glycan-protein interactions [223]. These computational tools will also facilitate in guiding the expansion of functional glycome space by informing glycan synthesis strategies to generate novel binding motifs that need to be explored in the context of understanding biology governed by glycan-protein interactions.

## Acknowledgments

This work was supported by National Institutes of Health GM R37 GM057073-13 and in part by the Singapore–MIT Alliance for Research and Technology (SMART).

## Abbreviations used

| | |
|---|---|
| **GAG** | glycosaminoglycan |
| **HS** | heparan sulfate |
| **CS** | chondroitin sulfate |
| **DS** | dermatan sulfate |
| **HPLC** | high performance liquid chromatography |
| **MS** | mass spectrometry |
| **CE** | capillary electrophoresis |
| **NMR** | nuclear magnetic resonance |
| **GBP** | glycan-binding protein |
| **MALDI** | matrix-assisted laser desorption ionization |
| **ESI** | electrospray ionization |
| **GGDB** | GlycoGene DataBase |
| **CAZy** | Carbohydrate-Active enZYme Database |

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

| | |
|---|---|
| **KEGG** | Kyoto Encyclopedia for Genes and Genomes |
| **CFG** | Consortium for Functional Glycomics |
| **HA** | hemagglutinin |
| **Neu5Ac** | N-Acetylneuraminic acid (sialic acid) |
| **Gal** | galactose |
| **SNA** | Sambucus nigra agglutinin |
| **H** | glucosamine |
| **G** | β-D-glucuronic acid |
| **I** | α-L-iduronic acid |
| **API** | active pharmaceutical ingredient |
| **OSCS** | oversulfated chondroitin sulfate |
| **ds** | degree of sulfation |
| **FGF** | fibroblast growth factor |

## References

1. Varki, A. Essentials of glycobiology. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2009.
2. Lowe JB, Marth JD. Annu Rev Biochem. 2003; 72:643–691. [PubMed: 12676797]
3. Tabak LA. Semin Cell Dev Biol. 2010; 21:616–621. [PubMed: 20144722]
4. Haltiwanger RS, Lowe JB. Annu Rev Biochem. 2004; 73:491–537. [PubMed: 15189151]
5. Hwang HY, Olson SK, Esko JD, Horvitz HR. Nature. 2003; 423:439–443. [PubMed: 12761549]
6. Marth JD, Grewal PK. Nat Rev Immunol. 2008; 8:874–887. [PubMed: 18846099]
7. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA. Science. 2001; 291:2370–2376. [PubMed: 11269318]
8. van Kooyk Y, Rabinovich GA. Nat Immunol. 2008; 9:593–601. [PubMed: 18490910]
9. Haines N, Irvine KD. Nat Rev Mol Cell Biol. 2003; 4:786–797. [PubMed: 14570055]
10. Varki A. Nature. 2007; 446:1023–1029. [PubMed: 17460663]
11. Ohtsubo K, Marth JD. Cell. 2006; 126:855–867. [PubMed: 16959566]
12. Fuster MM, Esko JD. Nat Rev Cancer. 2005; 5:526–542. [PubMed: 16069816]
13. Lau KS, Dennis JW. Glycobiology. 2008; 18:750–760. [PubMed: 18701722]
14. Wagner KW, Punnoose EA, Januario T, Lawrence DA, Pitti RM, Lancaster K, Lee D, von Goetz M, Yee SF, Totpal K, Huw L, Katta V, Cavet G, Hymowitz SG, Amler L, Ashkenazi A. Nat Med. 2007; 13:1070–1077. [PubMed: 17767167]
15. Liu FT, Rabinovich GA. Nat Rev Cancer. 2005; 5:29–41. [PubMed: 15630413]
16. Sasisekharan R, Shriver Z, Venkataraman G, Narayanasami U. Nat Rev Cancer. 2002; 2:521–528. [PubMed: 12094238]
17. Dube DH, Bertozzi CR. Nat Rev Drug Discov. 2005; 4:477–488. [PubMed: 15931257]
18. Nilsson EC, Storm RJ, Bauer J, Johansson SM, Lookene A, Angstrom J, Hedenstrom M, Eriksson TL, Frangsmyr L, Rinaldi S, Willison HJ, Domellof FP, Stehle T, Arnberg N. Nat Med. 2011; 17:105–109. [PubMed: 21151139]
19. Scanlan CN, Offer J, Zitzmann N, Dwek RA. Nature. 2007; 446:1038–1045. [PubMed: 17460665]
20. Chen Y, Maguire T, Hileman RE, Fromm JR, Esko JD, Linhardt RJ, Marks RM. Nat Med. 1997; 3:866–871. [PubMed: 9256277]
21. Helenius A, Aebi M. Science. 2001; 291:2364–2369. [PubMed: 11269317]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript


22. Macdougall IC, Gray SJ, Elston O, Breen C, Jenkins B, Browne J, Egrie J. Journal of the American Society of Nephrology. 1999; 10:2392–2395. [PubMed: 10541299]
23. Mattio M, Ceaglio N, Oggero M, Perotti N, Amadeo I, Orozco G, Forno G, Kratje R, Etcheverrigaray M. Biotechnology Progress. 2011; 27:1018–1028. [PubMed: 21608141]
24. Sasisekharan R, Raman R, Prabhakar V. Annu Rev Biomed Eng. 2006; 8:181–231. [PubMed: 16834555]
25. Schnaar, RL.; Suzuki, A.; Stanley, P. Essentials of Glycobiology. Varki, A.; Cummings, R.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME., editors. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009.
26. Taylor, ME.; Drickamer, K. Introduction to glycobiology. New York: Oxford University Press, Oxford; 2003.
27. Collins BE, Paulson JC. Curr Opin Chem Biol. 2004; 8:617–625. [PubMed: 15556405]
28. Kiessling, LL.; Young, T.; Gruber, TD.; Mortell, KH. Fraser-Reid, BO.; Tatsuta, K.; Thiem, J., editors. Springer Berlin Heidelberg; 2008. p. 2483-2523.
29. Marino K, Bones J, Kattla JJ, Rudd PM. Nat Chem Biol. 2010; 6:713–723. [PubMed: 20852609]
30. Vanderschaeghe D, Festjens N, Delanghe J, Callewaert N. Biological Chemistry. 2010; 391:149–161. [PubMed: 20128687]
31. Zaia J. OMICS. 2010; 14:401–418. [PubMed: 20443730]
32. Kato, K.; Yamaguchi, Y. Experimental Glycoscience. Taniguchi, N.; Suzuki, A.; Ito, Y.; Narimatsu, H.; Kawasaki, T.; Hase, S., editors. Springer Japan; 2008. p. 45-50.
33. Cummings, RD. Lectins as Tools for Glycoconjugate Purification and Characterization. Wiley-VCH Verlag GmbH; 2008.
34. El Rassi, Z., editor. Carbohydrate analysis by modern chromatography and electrophoresis. Elsevier; 2002.
35. Nakano, M.; Kakehi, K.; Taniguchi, N.; Kondo, A. Volpi, N., editor. Humana Press; 2011. p. 205-235.
36. Mechref Y, Novotny MV. Mass Spectrom Rev. 2009; 28:207–222. [PubMed: 18973241]
37. Pilobello KT, Mahal LK. Curr Opin Chem Biol. 2007; 11:300–305. [PubMed: 17500024]
38. Mechref Y, Novotny MV. Chem Rev. 2002; 102:321–369. [PubMed: 11841246]
39. Seeberger, PH.; Finney, N.; Rabuka, D.; Bertozzi, CR. Essentials of glycobiology. Varki, A.; Cummings, RD.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME., editors. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009.
40. Blixt O, Razi N. Methods Enzymol. 2006; 415:137–153. [PubMed: 17116472]
41. Bertozzi CR, Kiessling LL. Science. 2001; 291:2357–2364. [PubMed: 11269316]
42. Bernardes, Ga. J. L; Castagner, B.; Seeberger, PH. ACS Chemical Biology. 2009; 4:703–713. [PubMed: 19271728]
43. Seeberger PH, Werz DB. Nature. 2007; 446:1046–1051. [PubMed: 17460666]
44. Paulson JC, Blixt O, Collins BE. Nat Chem Biol. 2006; 2:238–248. [PubMed: 16619023]
45. Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. Nat Methods. 2005; 2:817–824. [PubMed: 16278650]
46. Honke, K.; Taniguchi, N. The sugar code. Fundamentals of glycosciences. Gabius, HJ., editor. Weinheim: Wiley-VCH; 2009. p. 385-401.
47. von der Lieth, C-W. Bioinformatics for Glycobiology and Glycomics. John Wiley & Sons, Ltd; 2009. p. 1-20.
48. Ceroni, A.; Joshi, HJ.; Maaß, K.; Ranzinger, R.; Lieth, C-W. von d Glycoscience. Fraser-Reid, BO.; Tatsuta, K.; Thiem, J., editors. Springer Berlin Heidelberg; 2008. p. 2219-2240.
49. Perez S, Mulloy B. Curr Opin Struct Biol. 2005; 15:517–524. [PubMed: 16143513]
50. Raman, R.; Sasisekharan, R. Wiley Encyclopedia of Chemical Biology. Begley, TP., editor. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2007.
51. Bocker S, Kehr B, Rasche F. IEEE/ACM Trans Comput Biol Bioinform. 2010
52. Harvey DJ. International Journal of Mass Spectrometry. 2003; 226:1–35.
53. Harvey DJ. Expert Rev Proteomics. 2005; 2:87–101. [PubMed: 15966855]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript


54. Liu X, McNally DJ, Nothaft H, Szymanski CM, Brisson JR, Li J. Anal Chem. 2006; 78:6081–6087. [PubMed: 16944887]
55. Mechref Y, Novotny MV, Krishnan C. Anal Chem. 2003; 75:4895–4903. [PubMed: 14674469]
56. Morelle W, Michalski JC. Curr Pharm Des. 2005; 11:2615–2645. [PubMed: 16101462]
57. Zaia J. Chem Biol. 2008; 15:881–892. [PubMed: 18804025]
58. Haslam SM, North SJ, Dell A. Curr Opin Struct Biol. 2006; 16:584–591. [PubMed: 16938453]
59. Brisson JR, Vinogradov E, McNally DJ, Khieu NH, Schoenhofen IC, Logan SM, Jarrell H. Methods Mol Biol. 2010; 600:155–173. [PubMed: 19882127]
60. Duus J, Gotfredsen CH, Bock K. Chem Rev. 2000; 100:4589–4614. [PubMed: 11749359]
61. Leeflang BR, Faber EJ, Erbel P, Vliegenthart JF. J Biotechnol. 2000; 77:115–122. [PubMed: 10674218]
62. Vliegenthart Johannes, FG. NMR Spectroscopy and Computer Modeling of Carbohydrates. American Chemical Society; 2006. p. 1-19.
63. von der Lieth, C-W. NMR Databases and Tools for Automatic Interpretation of Spectra of Carbohydrates. John Wiley & Sons, Ltd; 2009.
64. Cummings, RD. Methods in Enzymology. William, J.; Lennarz, GWH., editors. Academic Press; 1994. p. 66-86.
65. Endo, T. Journal of Chromatography Library. El Ziad, R., editor. Elsevier; 2002. p. 251-265.
66. Hirabayashi J. Glycoconj J. 2004; 21:35–40. [PubMed: 15467396]
67. Gupta G, Surolia A, Sampathkumar SG. OMICS. 2010; 14:419–436. [PubMed: 20726799]
68. Hsu KL, Mahal LK. Curr Opin Chem Biol. 2009; 13:427–432. [PubMed: 19716334]
69. Krishnamoorthy, L.; Mahal, LK. Functional and Structural Proteomics of Glycoproteins. Owens, R.; Nettleship, J., editors. Springer Netherlands; 2011. p. 91-102.
70. Rüdiger H, Gabius H-J. Glycoconjugate Journal. 2001; 18:589–613. [PubMed: 12376725]
71. Tao SC, Li Y, Zhou J, Qian J, Schnaar RL, Zhang Y, Goldstein IJ, Zhu H, Schneck JP. Glycobiology. 2008; 18:761–769. [PubMed: 18625848]
72. Wu A, Lisowska E, Duk M, Yang Z. Glycoconjugate Journal. 2009; 26:899–913. [PubMed: 18368479]
73. Hirabayashi J. Journal of Biochemistry. 2008; 144:139–147. [PubMed: 18390573]
74. Lowe, JB.; Marth, JD. Essentials of Glycobiology. Varki, A., editor. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1999. p. 211-252.
75. Gagneux P, Varki A. Glycobiology. 1999; 9:747–755. [PubMed: 10406840]
76. Stick, RV.; Williams, SJ. Carbohydrates: the essential molecules of life. Stick, RV.; Williams, SJ., editors. Amsterdam: Elsevier Science; 2008. p. 364-367.
77. Chen X, Varki A. ACS Chem Biol. 2010; 5:163–176. [PubMed: 20020717]
78. Schauer R. Curr Opin Struct Biol. 2009; 19:507–514. [PubMed: 19699080]
79. Varki A. Trends Mol Med. 2008; 14:351–360. [PubMed: 18606570]
80. Cohen M, Varki A. OMICS. 2010; 14:455–464. [PubMed: 20726801]
81. Gama CI, Hsieh-Wilson LC. Current Opinion in Chemical Biology. 2005; 9:609–619. [PubMed: 16242378]
82. Gama CI, Tully SE, Sotogaku N, Clark PM, Rawat M, Vaidehi N, Goddard WA, Nishi A, Hsieh-Wilson LC. Nat Chem Biol. 2006; 2:467–473. [PubMed: 16878128]
83. Raman R, Sasisekharan V, Sasisekharan R. Chemistry & Biology. 2005; 12:267–277. [PubMed: 15797210]
84. Lamanna WC, Kalus I, Padva M, Baldwin RJ, Merry CLR, Dierks T. Journal of Biotechnology. 2007; 129:290–307. [PubMed: 17337080]
85. Lamari F, Militsopoulou M, Gioldassi X, Karamanos N. Fresenius' Journal of Analytical Chemistry. 2001; 371:157–167. [PubMed: 11678186]
86. Ruhaak L, Zauner G, Huhn C, Bruggink C, Deelder A, Wuhrer M. Analytical and Bioanalytical Chemistry. 2010; 397:3457–3481. [PubMed: 20225063]
87. Harvey DJ. Mass Spectrometry Reviews. 1999; 18:349–450. [PubMed: 10639030]

88. Talbo G, Mann M. Rapid Communications in Mass Spectrometry. 1996; 10:100–103. [PubMed: 8563011]
89. Zaia J. Mass Spectrometry Reviews. 2004; 23:161–227. [PubMed: 14966796]
90. Powell AK, Harvey DJ. Rapid Communications in Mass Spectrometry. 1996; 10:1027–1032. [PubMed: 8755235]
91. An HJ, Lebrilla CB. Israel Journal of Chemistry. 2001; 41:117–128.
92. Fukuyama Y, Nakaya S, Yamazaki Y, Tanaka K. Analytical Chemistry. 2008; 80:2171–2179. [PubMed: 18275166]
93. Laremore TN, Murugesan S, Park T-J, Avci FY, Zagorevski DV, Linhardt RJ. Analytical Chemistry. 2006; 78:1774–1779. [PubMed: 16536411]
94. Papac DI, Wong A, Jones AJS. Analytical Chemistry. 1996; 68:3215–3223. [PubMed: 8797382]
95. Pitt JJ, Gorman JJ. Rapid Communications in Mass Spectrometry. 1996; 10:1786–1788.
96. Zhang J, LaMotte L, Dodds ED, Lebrilla CB. Analytical Chemistry. 2005; 77:4429–4438. [PubMed: 16013856]
97. Juhasz P, Biemann K. Proceedings of the National Academy of Sciences. 1994; 91:4333–4337.
98. Mechref Y, Novotny MV. Journal of the American Society for Mass Spectrometry. 1998; 9:1293–1302. [PubMed: 9835074]
99. Jacobs A, Dahlman O. Analytical Chemistry. 2000; 73:405–410. [PubMed: 11217739]
100. Harvey DJ. J Chromatogr B Analyt Technol Biomed Life Sci. 2010
101. Wada Y, Azadi P, Costello CE, Dell A, Dwek RA, Geyer H, Geyer R, Kakehi K, Karlsson NG, Kato K, Kawasaki N, Khoo K-H, Kim S, Kondo A, Lattova E, Mechref Y, Miyoshi E, Nakamura K, Narimatsu H, Novotny MV, Packer NH, Perreault H, Peter-Katalini J, Pohlentz G, Reinhold VN, Rudd PM, Suzuki A, Taniguchi N. Glycobiology. 2007; 17:411–422. [PubMed: 17223647]
102. Ciucanu I, Kerek F. Carbohydrate Research. 1984; 131:209–217.
103. Ciucanu I. Analytica Chimica Acta. 2006; 576:147–155. [PubMed: 17723627]
104. Dell, A.; Reason, AJ.; Khoo, K-H.; Panico, M.; McDowell, RA.; Morris, HR. Methods in Enzymology. William, J.; Lennarz, GWH., editors. Academic Press; 1994. p. 108-132.
105. Sekiya S, Wada Y, Tanaka K. Analytical Chemistry. 2005; 77:4962–4968. [PubMed: 16053310]
106. Lei M, Mechref Y, Novotny MV. Journal of the American Society for Mass Spectrometry. 2009; 20:1660–1671. [PubMed: 19546010]
107. Thomsson KA, Karlsson NG, Hansson GC. Journal of Chromatography A. 1999; 854:131–139. [PubMed: 10497934]
108. Hitchcock AM, Costello CE, Zaia J. Biochemistry. 2006; 45:2350–2361. [PubMed: 16475824]
109. Thomsson KA, Schulz BL, Packer NH, Karlsson NG. Glycobiology. 2005; 15:791–804. [PubMed: 15814823]
110. Robbe C, Capon C, Coddeville B, Michalski J-C. Rapid Communications in Mass Spectrometry. 2004; 18:412–420. [PubMed: 14966848]
111. Manzi AE, Norgard-Sumnicht K, Argade S, Marth JD, van Halbeek H, Varki A. Glycobiology. 2000; 10:669–689. [PubMed: 10910972]
112. Lütteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth C-W. Glycobiology. 2006; 16:71R–81R.
113. Lundborg M, Widmalm G. Anal Chem. 2011
114. Jansson P-E, Stenutz R, Widmalm G. Carbohydrate Research. 2006; 341:1003–1010. [PubMed: 16564037]
115. Lütteke T. ChemBioChem. 2008; 9:2155–2160. [PubMed: 18693281]
116. Royle L, Mattu TS, Hart E, Langridge JI, Merry AH, Murphy N, Harvey DJ, Dwek RA, Rudd PM. Anal Biochem. 2002; 304:70–90. [PubMed: 11969191]
117. Rudd PM, Colominas C, Royle L, Murphy N, Hart E, Merry AH, Hebestreit HF, Dwek RA. Proteomics. 2001; 1:285–294. [PubMed: 11680875]
118. Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. Bioinformatics. 2008; 24:1214–1216. [PubMed: 18344517]


NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

119. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM. Journal of Proteome Research. 2006; 5:1792–1802. [PubMed: 16823988]
120. Stadlmann J, Weber A, Pabst M, Anderle H, Kunert RJ, Ehrlich H, Peter Schwarz H, Altmann F. Proteomics. 2009; 9:4143–4153. [PubMed: 19688751]
121. McDonald CA, Yang JY, Marathe V, Yen T-Y, Macher BA. Molecular & Cellular Proteomics. 2009; 8:287–301. [PubMed: 18923192]
122. Guerrini M, Raman R, Venkataraman G, Torri G, Sasisekharan R, Casu B. Glycobiology. 2002; 12:713–719. [PubMed: 12460939]
123. Venkataraman G, Shriver Z, Raman R, Sasisekharan R. Science. 1999; 286:537–542. [PubMed: 10521350]
124. Hirabayashi J, Kuno A, Tateno H. Electrophoresis. 2011; 32:1118–1128. [PubMed: 21544837]
125. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. Nature. 2006; 440:435–436. [PubMed: 16554799]
126. Nicholls J, Bourne A, Chen H, Guan Y, Peiris JM. Respiratory Research. 2007; 8:73. [PubMed: 17961210]
127. Brooks SA, Hall DMS, Buley I. Br J Cancer. 2001; 85:1014–1022. [PubMed: 11592774]
128. Danguy A, Decaestecker C, Genten F, Salmon I, Kiss R. Cells Tissues Organs. 1998; 161:206–218.
129. Caron, M.; Seve, A-P., editors. Lectins and Pathology. Amsterdam: Harwood Academic Publishers; 2000.
130. Varki NM, Varki A. Lab Invest. 2007; 87:851–857. [PubMed: 17632542]
131. Raman, R.; Sasisekharan, R.; Begley, TP. Wiley Encyclopedia of Chemical Biology. John Wiley & Sons, Inc; 2007.
132. von der Lieth CW, Lutteke T, Frank M. Biochim Biophys Acta. 2006; 1760:568–577. [PubMed: 16459020]
133. von der Lieth, CW. Bioinformatics for Glycobiology and Glycomics: An Introduction. Von der Lieth, CW.; Lutteke, T.; Frank, M., editors. Chichester: John Wiley & Sons Ltd; 2009. p. 195-202.
134. Von der Lieth, CW. Comprehensive Glycoscience. Johannis, PK., editor. Oxford: Elsevier; 2007. p. 329-346.Editor-in-Chief
135. Aoki-Kinoshita KF. PLoS Comput Biol. 2008; 4:e1000075. [PubMed: 18516240]
136. Ranzinger, R.; Maaß, K.; Lütteke, T. Owens, R.; Nettleship, J., editors. Springer Netherlands; 2011. p. 59-90.
137. von der Lieth C-W, Freire AA, Blank D, Campbell MP, Ceroni A, Damerell DR, Dell A, Dwek RA, Ernst B, Fogh R, Frank M, Geyer H, Geyer R, Harrison MJ, Henrick K, Herget S, Hull WE, Ionides J, Joshi HJ, Kamerling JP, Leeflang BR, Lütteke T, Lundborg M, Maass K, Merry A, Ranzinger R, Rosen J, Royle L, Rudd PM, Schloissnig S, Stenutz R, Vranken WF, Widmalm G, Haslam SM. Glycobiology. 2011; 21:493–502. [PubMed: 21106561]
138. Cooper CA, Gasteiger E, Packer NH. Proteomics. 2001; 1:340–349. [PubMed: 11680880]
139. Goldberg D, Sutton-Smith M, Paulson J, Dell A. Proteomics. 2005; 5:865–875. [PubMed: 15693066]
140. Packer NH, von der Lieth CW, Aoki-Kinoshita KF, Lebrilla CB, Paulson JC, Raman R, Rudd P, Sasisekharan R, Taniguchi N, York WS. Proteomics. 2008; 8:8–20. [PubMed: 18095367]
141. Lohmann KK, von der Lieth C-W. Nucleic Acids Research. 2004; 32:W261–W266. [PubMed: 15215392]
142. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. J Proteome Res. 2008; 7:1650–1659. [PubMed: 18311910]
143. Ceroni A, Dell A, Haslam S. Source Code for Biology and Medicine. 2007; 2:3. [PubMed: 17683623]
144. Tang H, Mechref Y, Novotny MV. Bioinformatics. 2005; 21(Suppl 1):431–i439.
145. Zhang H, Singh S, Reinhold VN. Anal Chem. 2005; 77:6263–6270. [PubMed: 16194087]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


146. Gaucher SP, Morrow J, Leary JA. Analytical Chemistry. 2000; 72:2331–2336. [PubMed: 10857602]
147. Ethier M, Saba JA, Ens W, Standing KG, Perreault H. Rapid Commun Mass Spectrom. 2002; 16:1743–1754. [PubMed: 12207362]
148. Lapadula AJ, Hatcher PJ, Hanneman AJ, Ashline DJ, Zhang H, Reinhold VN. Anal Chem. 2005; 77:6271–6279. [PubMed: 16194088]
149. Nairn, A.; Moremen, K. Handbook of glycomics. Cummings, R., editor. London: Elsevier; 2009. p. 95-136.
150. Comelli EM, Amado M, Head SR, Paulson JC. Biochem Soc Symp. 2002:135–142. [PubMed: 12655780]
151. Comelli EM, Head SR, Gilmartin T, Whisenant T, Haslam SM, North SJ, Wong NK, Kudo T, Narimatsu H, Esko JD, Drickamer K, Dell A, Paulson JC. Glycobiology. 2006; 16:117–131. [PubMed: 16237199]
152. Nairn AV, York WS, Harris K, Hall EM, Pierce JM, Moremen KW. J Biol Chem. 2008; 283:17298–17313. [PubMed: 18411279]
153. Bishop JR, Schuksz M, Esko JD. Nature. 2007; 446:1030–1037. [PubMed: 17460664]
154. Forsberg E, Kjellén L. The Journal of Clinical Investigation. 2001; 108:175–180. [PubMed: 11457868]
155. Perrimon N, Bernfield M. Nature. 2000; 404:725–728. [PubMed: 10783877]
156. Inatani M, Irie F, Plump AS, Tessier-Lavigne M, Yamaguchi Y. Science. 2003; 302:1044–1046. [PubMed: 14605369]
157. Schwarzkopf M, Knobeloch K-P, Rohde E, Hinderlich S, Wiechens N, Lucka L, Horak I, Reutter W, Horstkorte R. Proceedings of the National Academy of Sciences. 2002; 99:5267–5270.
158. Galeano B, Klootwijk R, Manoli I, Sun M, Ciccone C, Darvish D, Starost MF, Zerfas PM, Hoffmann VJ, Hoogstraten-Miller S, Krasnewich DM, Gahl WA, Huizing M. The Journal of Clinical Investigation. 2007; 117:1585–1594. [PubMed: 17549255]
159. Lepenies B, Yin J, Seeberger PH. Curr Opin Chem Biol. 2010; 14:404–411. [PubMed: 20227905]
160. Park S, Lee M-R, Shin I. Chemical Society Reviews. 2008; 37:1579–1591. [PubMed: 18648683]
161. Muthana S, Cao H, Chen X. Curr Opin Chem Biol. 2009; 13:573–581. [PubMed: 19833544]
162. Chen J, Jones CL, Liu J. Chemistry & Biology. 2007; 14:986–993. [PubMed: 17884631]
163. Linhardt RJ, Dordick JS, Deangelis PL, Liu J. Semin Thromb Hemost. 2007; 33453:465.
164. Avci Fikri, Y.; DeAngelis Paul, L.; Liu, J.; Linhardt Robert, J. Frontiers in Modern Carbohydrate Chemistry. American Chemical Society; 2007. p. 253-284.
165. Liu R, Xu Y, Chen M, Weïwer M, Zhou X, Bridges AS, DeAngelis PL, Zhang Q, Linhardt RJ, Liu J. Journal of Biological Chemistry. 2010; 285:34240–34249. [PubMed: 20729556]
166. Kiefel MJ, von Itzstein M. Chemical Reviews. 2002; 102:471–490. [PubMed: 11841251]
167. Blixt O, Allin K, Pereira L, Datta A, Paulson JC. Journal of the American Chemical Society. 2002; 124:5739–5746. [PubMed: 12010048]
168. Chokhawala HA, Huang S, Lau K, Yu H, Cheng J, Thon V, Hurtado-Ziola N, Guerrero JA, Varki A, Chen X. ACS Chemical Biology. 2008; 3:567–576. [PubMed: 18729452]
169. Kiessling LL, Splain RA. Annual Review of Biochemistry. 2010; 79:619–653.
170. Kiessling LL, Gestwicki JE, Strong LE. Angewandte Chemie International Edition. 2006; 45:2348–2368.
171. Pieters RJ. Organic & Biomolecular Chemistry. 2009; 7:2013–2025. [PubMed: 19421435]
172. Lundquist JJ, Toone EJ. Chemical Reviews. 2002; 102:555–578. [PubMed: 11841254]
173. Smith, DF.; Cummings, RD. Handbook of glycomics. Cummings, RD.; Pierce, JM., editors. Amsterdam: Elsevier; 2009. p. 139-160.
174. Smith, DF.; Song, X.; Cummings, RD. Methods in Enzymology. Minoru, F., editor. Academic Press; 2010. p. 417-444.
175. Rillahan, CD.; Paulson, JC. Annual Review of Biochemistry. Vol. 80. null; 2011.
176. Oyelaran O, Gildersleeve JC. Current Opinion in Chemical Biology. 2009; 13:406–413. [PubMed: 19625207]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

177. Horlacher T, Seeberger PH. Chemical Society Reviews. 2008; 37:1414–1422. [PubMed: 18568167]
178. Lonardi E, Balog CI, Deelder AM, Wuhrer M. Expert Rev Proteomics. 2010; 7:761–774. [PubMed: 20973647]
179. Liu Y, Palma AS, Feizi T. Biological Chemistry. 2009; 390:647–656. [PubMed: 19426131]
180. Liang P-H, Wu C-Y, Greenberg WA, Wong C-H. Current Opinion in Chemical Biology. 2008; 12:86–92. [PubMed: 18258211]
181. Stowell SR, Arthur CM, Mehta P, Slanina KA, Blixt O, Leffler H, Smith DF, Cummings RD. Journal of Biological Chemistry. 2008; 283:10109–10123. [PubMed: 18216021]
182. Tully SE, Rawat M, Hsieh-Wilson LC. Journal of the American Chemical Society. 2006; 128:7740–7741. [PubMed: 16771479]
183. Shipp EL, Hsieh-Wilson LC. Chemistry & Biology. 2007; 14:195–208. [PubMed: 17317573]
184. de Paz JL, Moseman EA, Noti C, Polito L, von Andrian UH, Seeberger PH. ACS Chemical Biology. 2007; 2:735–744. [PubMed: 18030990]
185. Rogers CJ, Clark PM, Tully SE, Abrol R, Garcia KC, Goddard WA, Hsieh-Wilson LC. Proceedings of the National Academy of Sciences. 2011; 108:9747–9752.
186. de Paz JL, Seeberger PH. Molecular BioSystems. 2008; 4:707–711. [PubMed: 18563243]
187. Noti C, de Paz JL, Polito L, Seeberger PH. Chemistry – A European Journal. 2006; 12:8664–8686.
188. Taylor ME, Drickamer K. Glycobiology. 2009; 19:1155–1162. [PubMed: 19528664]
189. Zhang Y, Li Q, Rodriguez LG, Gildersleeve JC. Journal of the American Chemical Society. 2010; 132:9653–9662. [PubMed: 20583754]
190. York, WS.; Kochut, KJ.; Miller, JA. Handbook of Glycomics. Richard, DC.; Pierce, JM., editors. San Diego: Academic Press; 2010. p. 177-195.
191. Turnbull JE, Linhardt RJ. Nat Chem Biol. 2006; 2:449–450. [PubMed: 16921352]
192. Visvanathan M, Siddam SR, Lee IH, Lushington GH, Bousfield GR. Open Med Inform J. 2011; 5:9–16. [PubMed: 21603090]
193. Bertozzi, CR.; Sasisekharan, R. Essentials of Glycobiology. Varki, A.; Cummings, R.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME., editors. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009.
194. Turnbull JE, Field RA. Nat Chem Biol. 2007; 3:74–77. [PubMed: 17235338]
195. Prescher JA, Bertozzi CR. Cell. 2006; 126:851–854. [PubMed: 16959565]
196. Turnbull JE, Sasisekharan R. OMICS: A Journal of Integrative Biology. 2010; 14:385–387. [PubMed: 20726798]
197. Hart GW, Copeland RJ. Cell. 2010; 143:672–676. [PubMed: 21111227]
198. Li J, Richards JC. 2009:1–8.
199. Ibricevic A, Pekosz A, Walter MJ, Newby C, Battaile JT, Brown EG, Holtzman MJ, Brody SL. J. Virol. 2006; 80:7469–7480. [PubMed: 16840327]
200. Russell R, Stevens D, Haire L, Gamblin S, Skehel J. Glycoconjugate Journal. 2006; 23:85–92. [PubMed: 16575525]
201. Skehel JJ, Wiley DC. Annual Review of Biochemistry. 2000; 69:531–569.
202. van Riel D, Munster VJ, de Wit E, Rimmelzwaan GF, Fouchier RAM, Osterhaus ADME, Kuiken T. The American Journal of Pathology. 2007; 171:1215–1223. [PubMed: 17717141]
203. Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, Wilson IA. Science. 2006; 312:404–410. [PubMed: 16543414]
204. Eisen MB, Sabesan S, Skehel JJ, Wiley DC. Virology. 1997; 232:19–31. [PubMed: 9185585]
205. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ. Science. 2004; 303:1838–1842. [PubMed: 14764886]
206. Ha Y, Stevens DJ, Skehel JJ, Wiley DC. Proceedings of the National Academy of Sciences. 2001; 98:11181–11186.
207. Ha Y, Stevens DJ, Skehel JJ, Wiley DC. Virology. 2003; 309:209–218. [PubMed: 12758169]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

208. Sauter NK, Hanson JE, Glick GD, Brown JH, Crowther RL, Park SJ, Skehel JJ, Wiley DC. Biochemistry. 1992; 31:9609–9621. [PubMed: 1327122]

209. Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC. Nature. 1988; 333:426–431. [PubMed: 3374584]

210. Tumpey TM, Maines TR, Van Hoeven N, Glaser L, Solórzano A, Pappas C, Cox NJ, Swayne DE, Palese P, Katz JM, García-Sastre A. Science. 2007; 315:655–659. [PubMed: 17272724]

211. Chandrasekaran A, Srinivasan A, Raman R, Viswanathan K, Raguram S, Tumpey TM, Sasisekharan V, Sasisekharan R. Nat Biotechnol. 2008; 26:107–113. [PubMed: 18176555]

212. Coombs PJ, Taylor ME, Drickamer K. Glycobiology. 2006; 16:1C–7C. [PubMed: 16118287]

213. Stevens J, Blixt O, Paulson JC, Wilson IA. Nat Rev Micro. 2006; 4:857–864.

214. Stevens J, Blixt O, Glaser L, Taubenberger JK, Palese P, Paulson JC, Wilson IA. Journal of Molecular Biology. 2006; 355:1143–1155. [PubMed: 16343533]

215. Srinivasan A, Viswanathan K, Raman R, Chandrasekaran A, Raguram S, Tumpey TM, Sasisekharan V, Sasisekharan R. Proceedings of the National Academy of Sciences. 2008; 105:2800–2805.

216. Sasisekharan R, Venkataraman G. Current Opinion in Chemical Biology. 2000; 4:626–631. [PubMed: 11102866]

217. Sasisekharan R, Shriver Z. Thromb Haemost. 2009; 102:854–858. [PubMed: 19888519]

218. Shriver Z, Raman R, Venkataraman G, Drummond K, Turnbull J, Toida T, Linhardt R, Biemann K, Sasisekharan R. Proceedings of the National Academy of Sciences. 2000; 97:10359–10364.

219. Xiao Z, Tappen BR, Ly M, Zhao W, Canova LP, Guan H, Linhardt RJ. Journal of Medicinal Chemistry. 2010; 54:603–610. [PubMed: 21166465]

220. Guerrini M, Beccati D, Shriver Z, Naggi A, Viswanathan K, Bisio A, Capila I, Lansing JC, Guglieri S, Fraser B, Al-Hakim A, Gunay NS, Zhang Z, Robinson L, Buhse L, Nasr M, Woodcock J, Langer R, Venkataraman G, Linhardt RJ, Casu B, Torri G, Sasisekharan R. Nat Biotech. 2008; 26:669–675.

221. Kishimoto TK, Viswanathan K, Ganguly T, Elankumaran S, Smith S, Pelzer K, Lansing JC, Sriranganathan N, Zhao G, Galcheva-Gargova Z, Al-Hakim A, Bailey GS, Fraser B, Roy S, Rogers-Cotrone T, Buhse L, Whary M, Fox J, Nasr M, Dal Pan GJ, Shriver Z, Langer RS, Venkataraman G, Austen KF, Woodcock J, Sasisekharan R. N Engl J Med. 2008

222. Porter A, Yue T, Heeringa L, Day S, Suh E, Haab BB. Glycobiology. 2010; 20:369–380. [PubMed: 19946132]

223. Woods RJ, Tessier MB. Current Opinion in Structural Biology. 2010; 20:575–583. [PubMed: 20708922]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

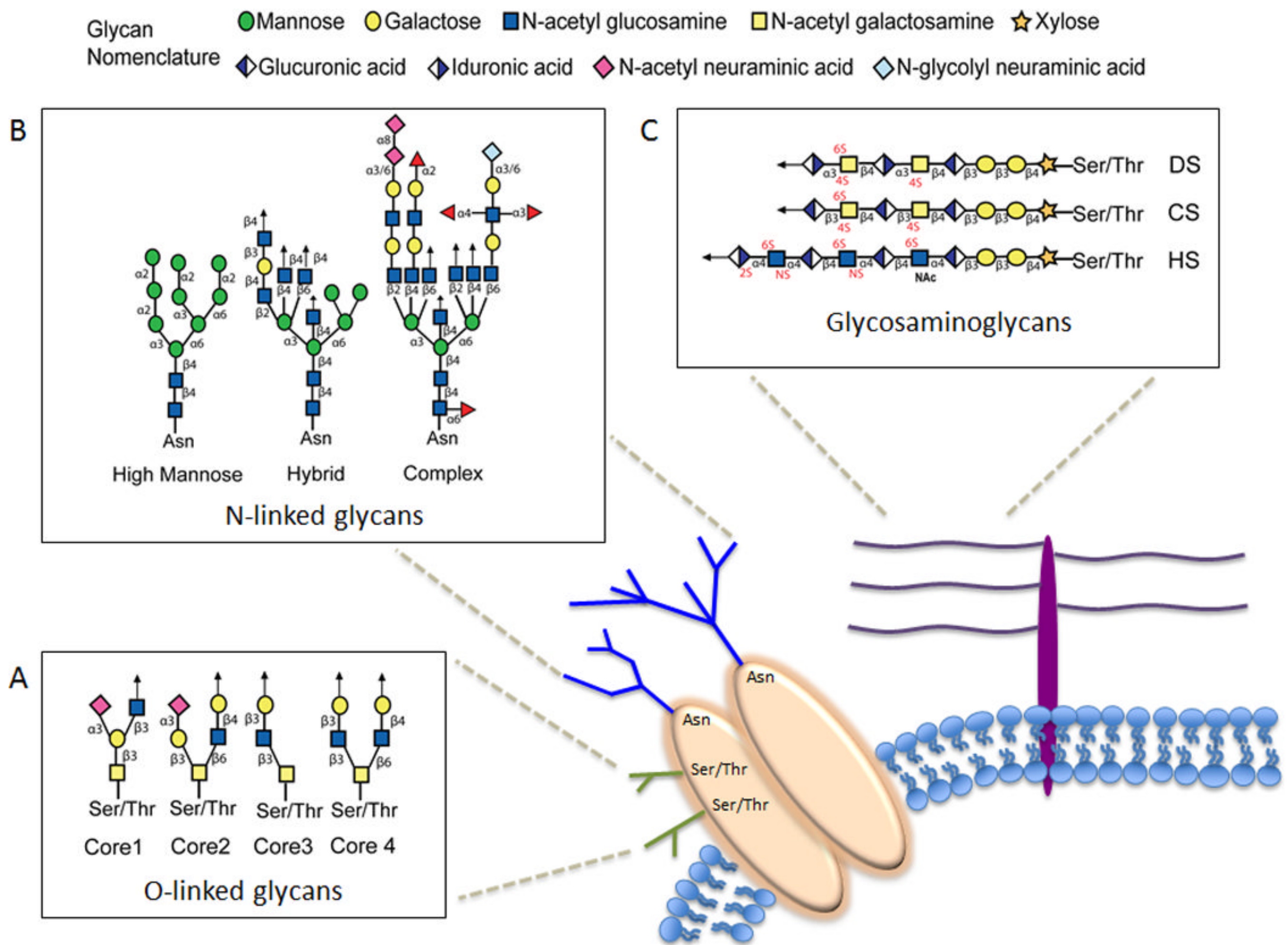
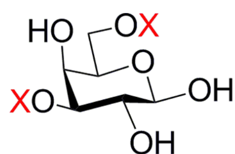



**Figure 1. Structural diversity of glycans**

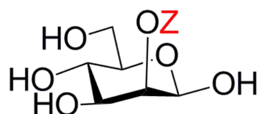
Based on their backbone structure, glycans can be broadly categorized as linear (GAGs) or branched (*N*-linked and *O*-linked glycans). (**A**) *O*-linked glycans, attached to proteins via covalent linkage typically to serine or threonine, can be subgrouped based on their core glycan structures. (**B**) In *N*-linked glycosylation, the glycan is typically covalently linked to asparagine on a protein containing the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except proline. *N*-linked glycans generally display larger and more complex structures than *O*-glycans but have less variability. All *N*-glycans share a common trimannosyl core sugar structure and can be categorized into three types based on the nature of monosaccharides extended from the core: high-mannose, complex, and hybrid. The antenna structures displayed represent example extensions; typically a maximum of four antennas with a bisecting *N*-acetylglucosamine exist for a given glycan. (**C**) GAGs, the predominant group of linear glycans, are polymeric chains of sulfated disaccharide repeat units consisting of an uronic acid linked to a hexosamine, which when linked to a protein are termed a proteoglycan (protein and GAG). GAGs are classified based on the monosaccharides making up their base disaccharide unit. Three examples are heparan sulfate glycosaminoglycans (HS), chondroitin sulfate glycosaminoglycans (CS), and dermatan sulfate glycosaminoglycans (DS).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

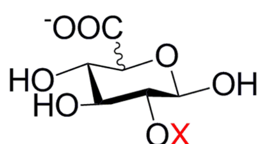




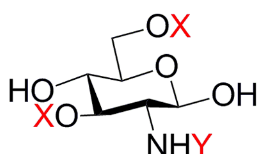

Sulfated Galactose



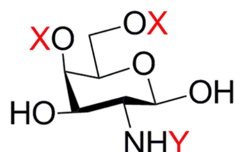

Phosphorylated Mannose



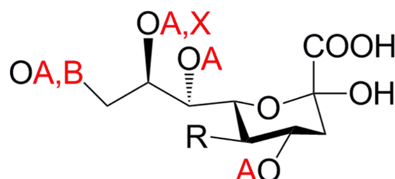

Sulfated Iduronic/Glucuronic Acid



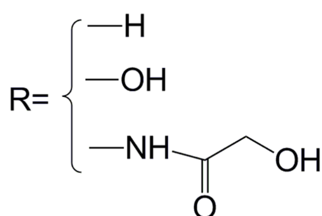

*O*-Sulfated and *N*-Sulfated (or *N*-Acetylated) Glucosamine




*O*-Sulfated and *N*-Sulfated (or *N*-Acetylated) Galactosamine




4,7,8,9-Acetylated and 8-Sulfated *N*-Acetylneuraminic Acid (where R= NHAc)




= Neuraminic Acid

= 2-Keto-3-deoxynononic acid

= *N*-Glycoylneuraminic acid

**Figure 2. Major classes of acidic glycans that appear on *N*- and *O*-linked glycans in mammals**

There are six monosaccharides that appear on the majority of acidic glycans. Mannose and *N*-acetylneuraminic acid (along with *N*-glycolylneuraminic acid and 2-keto-3-deoxynononic acid) primarily appear as glycoconjugates on proteins and lipids. Uronic acid and *N*-acetylgalactosamine appear primarily on proteoglycans. *N*-acetylglucosamine and galactose can appear in both glycoconjugates and proteoglycans. In the figure, a position labeled as X is either sulfated or unsulfated, Y is either acetylated or sulfated (or in rare circumstances unsubstituted), Z is either phosphorylated or not, A is either acetylated or not, and B indicates the potential presence of a lactate residue.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

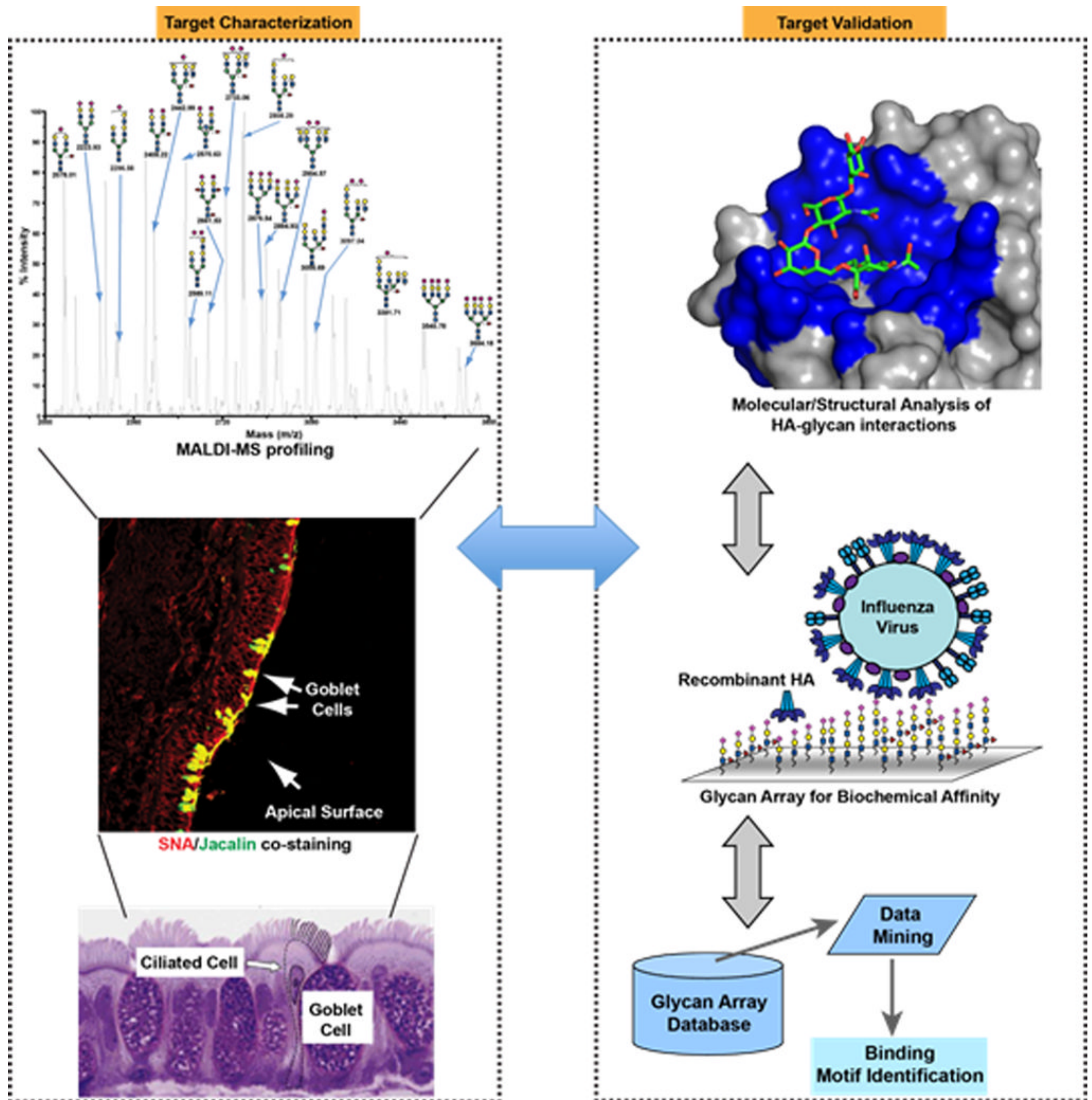



**Figure 3. Framework for integrated analyses of glycan-protein interactions modulating influenza A virus pathogenesis**

Shown on the left is detailed structural characterization of target glycan receptors in the human upper respiratory epithelia using a combination of lectin-staining (SNA, Sambucus nigra agglutinin lectin) and detailed characterization using MALDI-MS. Shown on the right is the validation of target glycan receptors by developing glycan array platforms comprising of glycan motifs representing target structures. These platforms are used to screen and determine quantitative relative binding affinities of different HAs and binding motifs are validated using data mining platform with the guidance of X-ray crystal structure of HA-glycan complexes.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

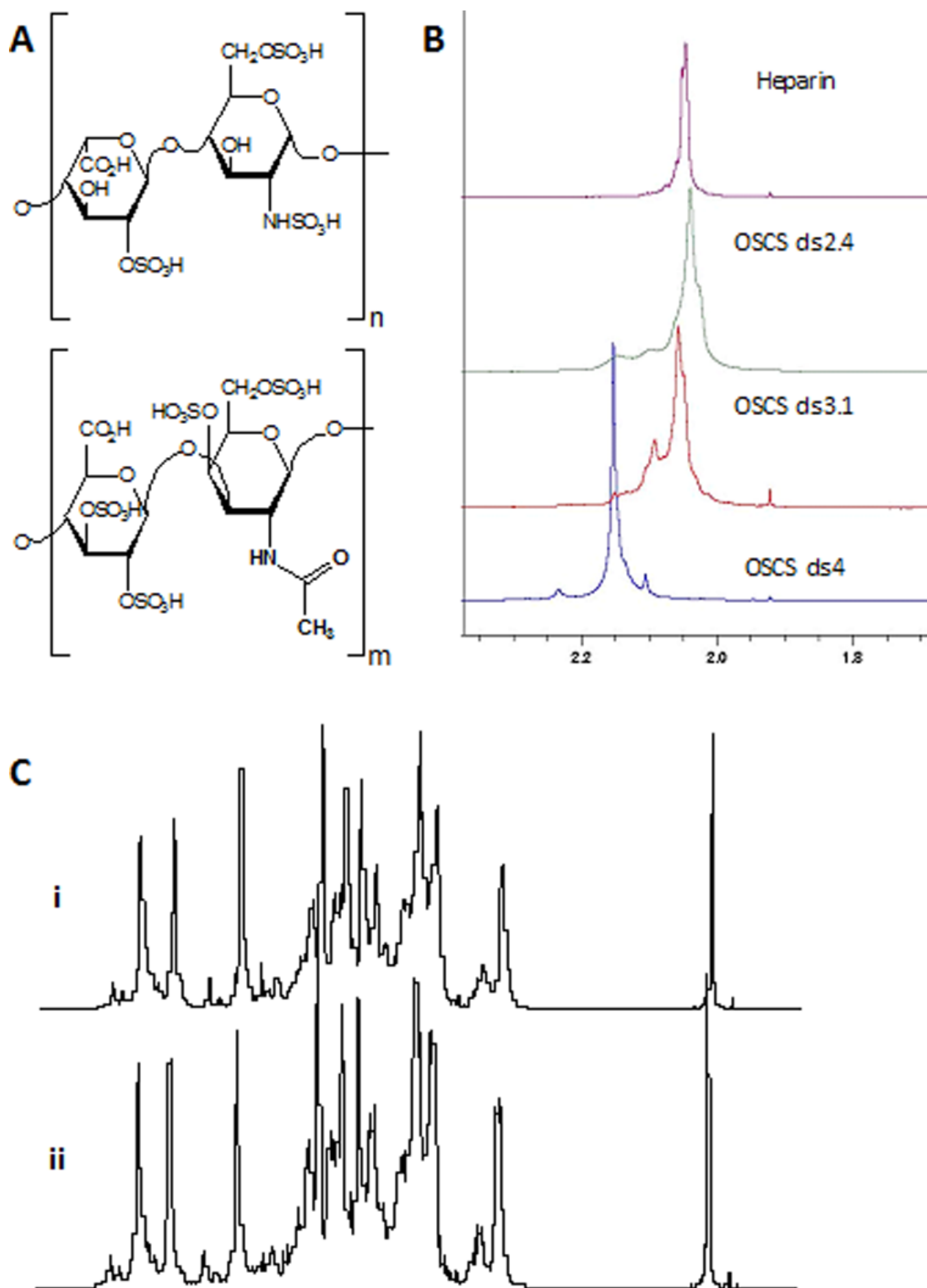



**Figure 4. Structural determination of OSCS demonstrates the need for analytical integration**

(**A**) Inspection of the disaccharide repeat units of heparin (top) and OSCS (bottom) indicate the close structural similarity between the two. Heparin, a polydisperse mixture of polysaccharide chains with an average length *n*, also has sequence diversity – the major disaccharide unit is trisulfated. OSCS is also polydisperse, with polysaccharide chains of average length *m*. However, all potential sites for sulfonation are "occupied" (degree of sulfation [ds] = 4), such that there is no real sequence variability in OSCS. (**B**) The acetyl signal for OSCS (ds=4) is readily distinct from that of heparin ($\delta_{methyl}$=2.12–2.16 compared to 2.07ppm, respectively), yielding a sensitive signal for detection of OSCS within heparin.

However, systematically changing the ds from 4 to 2.4 results in a shift in the methyl signal, resulting in substantial overlap with the signal arising from heparin. (**C**) 500 MHz proton spectra of (i) heparin alone or (ii) heparin spiked with 4% of alginate sulfate. The signals arising from alginate sulfate are not distinguishable from those arising from heparin, indicating that proton NMR is not a useful filter to detect alginate sulfate, if present within heparin.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

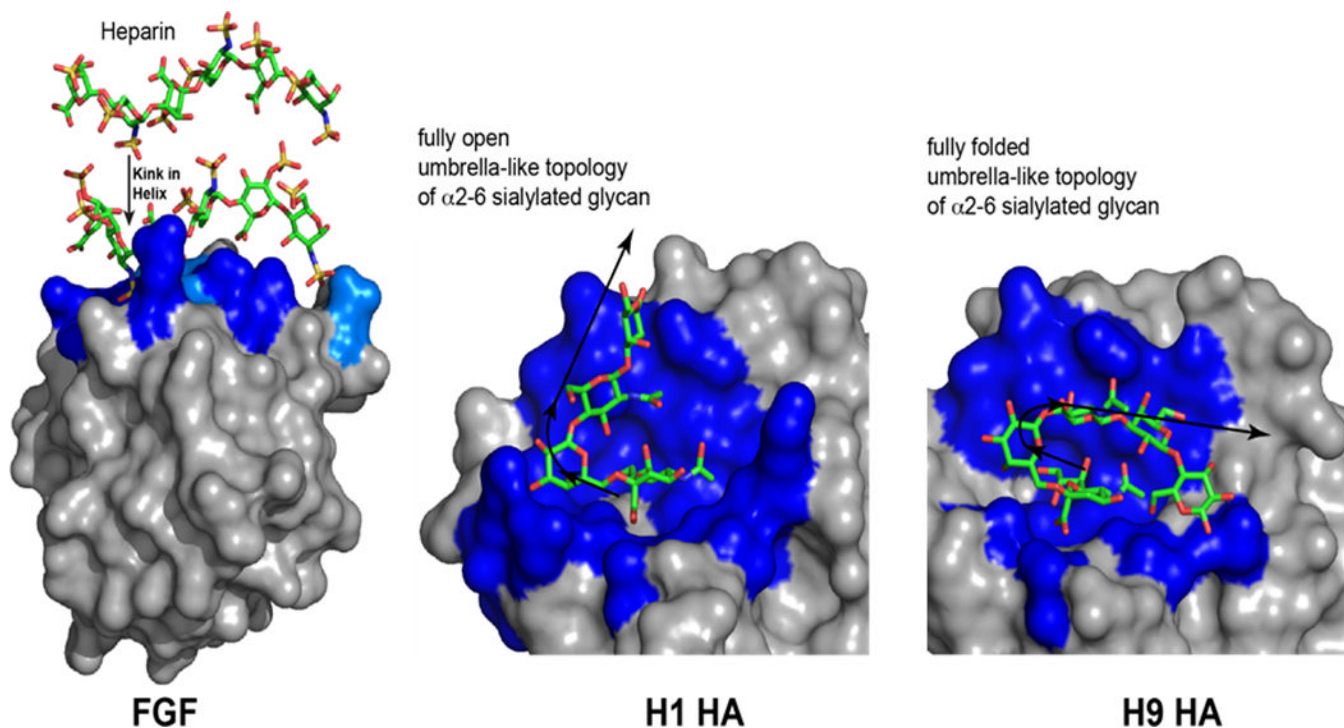



**Figure 5. Structural specificity in glycan-protein interactions**

In the case of linear glycans such as GAGs, a distinct *kink* in the helical axis of the unbound GAG chain is induced upon binding to protein such as fibroblast growth factor (FGF). The kink provides optimal *van der Waals* contact and also positions the key sulfate groups for optimal interactions with the protein. In the case of sialylated glycan receptors for influenza HA, depending on the nature of the binding pocket across different HA subtypes, the glycan receptor adopts a distinct umbrella-like topology spanning a fully open umbrella shape to fully closed umbrella shape. The protein is shown as a surface colored in gray with glycan-binding site colored in blue.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



**Table 1**

Challenges to the structure-function understanding of glycans.

| Key challenge | Features | Impact on study of glycans |
|---|---|---|
| Glycan biosynthesis | Nontemplate-driven process, unlike DNA/RNA and protein | Replication- or translation-like ‘rules’ cannot be easily applied; no direct methods to amplify glycans, unlike DNA (PCR) and protein (recombinant expression) |
| | Limited availability of glycans from natural sources (e.g., cells, tissues) | Without amplification tools, analytical and functional methods often require high sensitivity |
| | Tissue-, developmental-, and metabolic-dependent expression of glycan biosynthetic machinery (glycosyltransferases and glycosidases) | Glycan structure is sensitive to cellular conditions, tissue type, and developmental stages |
| | Lack of proofreading in glycan biosynthetic process | Increases structural diversity of glycans to be analyzed |
| Glycan structural complexity and heterogeneity | Presence of isomers and different anomeric configurations | Properties generally not present in DNA and proteins; challenges structural characterization by single method |
| | Microheterogeneity – a range of glycan structures (length, composition, branching) found at any given glycosylation site on a glycoprotein | Highly similar physicochemical properties of glycan microheterogeneities challenges their characterization |
| | Branching | Unambiguous designation of branches and their locations challenged by analytical approaches |
| | Presence of multiple modifications (sulfation, acetylation, methylation) and high diversity of linkages (location of linkages and anomericity) | Chemical synthesis is difficult and limited to small oligosaccharides due to the need of complex protecting and deprotecting strategies |
| | Site of attachment to protein/lipid | Requires glycan-protein and/or glycan-lipid characterization in addition to glycan structure |
| Glycan presentation and interactions | Presentation of an ensemble of different (often related) structures within a biological system or interaction | Studies must account for a population of glycans with similar structures, rather than an ‘average’ single structure |
| | Glycan-protein interactions often achieve high affinity and specificity by multivalency | Correct presentation of glycan and glycan-binding protein/ domain(s) is critical for experimental design |
| | Glycan-protein interactions modulate biology in an analog-like nature | Functional readouts must be characterized in terms of gradation of effects (not binary “on/off” effects) |
| | High torsional flexibility of glycans mediates presentation of a range of conformations for a particular glycan | Sequence of glycan is often not sufficient to characterize glycan-protein interactions; analysis of conformations and topologies should be considered |

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

**Table 2**

Overview of analytical methods for glycan structural characterization

| Method | Structural information | Strengths | Weaknesses |
|---|---|---|---|
| MS | Composition, profiling, chemical modifications, glycosylation site | High sensitivity and accuracy, potential for automation | Isomers indistinguishable, no separation of complex mixtures, quantification difficult |
| $MS^n$ | Sequence, linkages, chemical modifications | Detailed structural information | Significant expertise required, low throughput |
| CE | Profiling, separation | High separation efficiency, can resolve isomers, high throughput and automation, complements MS analysis, amenable to online connection with MS | Requires standards and/or complementary techniques (e.g., MS, enzymes) for Unambiguous characterization, limited analyte capacity |
| HPLC | Profiling, separation | Multiple separation modes/stationary phases, range of analyte capacity, complements MS analysis, amenable to online connection with MS | Requires standards and/or complementary techniques (e.g., MS, enzymes) for Unambiguous characterization |
| NMR | Complete structure, stereochemistry | Detailed structural information | Significant expertise required, low throughput, high amount of sample material required |
| Lectin binding/arrays | Structural motifs (terminal residues) | Potential for high throughput, probe glycans in their physiological context | Limited by availability and specificity of characterized lectins |