Published in final edited form as:

Adv Cancer Res. 2015 ; 126: 167–202. doi:10.1016/bs.acr.2014.11.003.

The detection and discovery of glycan motifs in biological samples using lectins and antibodies: new methods and opportunities

Huiyuan Tang, Peter Hsueh, Doron Kletter, Marshall Bern, and Brian Haab

Van Andel Research Institute, 333 Bostwick NE, Grand Rapids, MI, 49503 Protein Metrics, 1622 San Carlos Ave., Suite C, San Carlos, CA 94070

Abstract

Recent research is uncovering unexpected ways that glycans contribute to biology, as well as new strategies for combatting disease using approaches involving glycans. To make full use of glycans for clinical applications, we need more detailed information on the location, nature, and dynamics of glycan expression *in vivo*. Such studies require the use of specimens acquired directly from patients. Effective studies of clinical specimens require low-volume assays, high precision measurements, and the ability to process many samples. Assays using affinity reagents—lectins and glycan-binding antibodies—can meet these requirements, but further developments are needed to make the methods routine and effective. Recent advances in the use of glycan-binding proteins could meet that need. The advances involve improved determination of specificity using glycan arrays; the availability of databases for mining and analyzing glycan array data; lectin engineering methods; and the ability to quantitatively interpret lectin measurements. Here we describe many of the challenges and opportunities involved in the application of these new approaches to the study of biological samples. The new tools hold promise for developing methods to improve the outcomes of patients afflicted with diseases characterized by aberrant glycan expression.

Keywords

Glycans; glycan arrays; lectins; antibody-lectin sandwich arrays; glycan-binding proteins; GlycoSearch

Introduction

A growing body of research is showing the importance of glycobiology for human health and providing opportunities for improving health (Fuster & Esko, 2005). For example, recent studies showed that botulinum neurotoxin requires carbohydrate binding to effect toxicity (Lee et al.) and that a sialic acid ligand coupled to doxorubicin that could effectively kill B lymphoma cells through the CD22 B cell receptor (W. C. Chen et al.). The range of conditions involving glycans is expansive, including infection, immune disorders, developmental defects, cancer, and other diseases (Varki et al., 2009). Although the opportunities are many, progress in glycan-based strategies has been slow, hampered in many cases by a lack of detail about the contributions of glycans to the pathology.

To acquire this missing information, researchers will need to work with samples obtained directly from patients or animal models. Clinical specimens, as opposed to cell lines, provide an immediate look at the location, dynamics, and amount of expression of specific glycans in the relevant biological setting, as well as information about associations between clinical and molecular features. The methods used for complete structural analysis of glycans, including enzymatic digestions, chromatography, and mass spectrometry, have been foundational for glycobiology research, but they are not suitable for answering questions about associations with disease. Methods used for that purpose should be able to use small amounts of material, give precise measurements, and process many samples. These requirements stem from the limited volumes available for most clinical samples and the need to make statistical comparisons among many samples.

Affinity reagents can be effective for the study of glycans in biological samples. Two main types of affinity reagents are available for detecting glycans: lectins and glycan-binding antibodies. Lectins are proteins that bind specific glycan motifs without enzymatic activity, such as wheat germ agglutinin, a lectin found mainly in the kernel of wheat that binds to oligosaccharides commonly found on insects or other invaders, and E-selectin, which binds a glycan motif displayed on activated lymphocytes and some other cell types. Glycan-binding antibodies, on the other hand, are products of an immune response to a foreign glycan. In the remainder of the article, we use the term glycan-binding protein (GBP) to refer to both glycan-binding antibodies and lectins.

This article covers many of the current topics relating to the use of GBPs to probe glycans in biological samples. We begin with an overview of various formats that use GBPs as analytical reagents. We then describe several challenges in the use of GBPs, review the progress made in each area, and give viewpoints on promising areas of research and development. The challenges are: determining the specificity of a GBP, identifying the factors that alter GBP binding, interpreting the data acquired using GBPs, and finding reagents to detect candidate structures. We conclude the article with viewpoints and progress on using GBPs to discover glycans associated with cancer and developing biomarker assays from that information.

Ways to use GBPs for probing glycan motifs

The detection of glycan motifs—GBPs have long been used to detect and purify glycans (Sharon, 2007). A feature of GBPs is that they bind glycan motifs. A “motif” refers to a substructure that appears in multiple glycans, and a “glycan” refers to a complete oligosaccharide. For example, the motif N-acetyl-lactosamine (LAcNAc, Gal β 1-4GlcNAc) can occur on N-glycans, O-glycans, glycolipids, each with larger oligosaccharides surrounding the motif. A GBP that recognizes the LAcNAc motif could bind to any glycan that contains the LAcNAc motif. The motifs bound by GBPs are extremely variable; some are simple, such as a single fucose monosaccharide, and some are complex, large, or hard to define. But in any case GBPs can provide information that is hard to uncover using mass spectrometry and chromatography, because such methods do not readily reveal specific linkages, orientations, and presentations. The measurement of motifs is particularly valuable from the viewpoint of function. It is the motifs that a glycan displays, rather than the entire

structure of the glycan, that are important for function. Therefore, while GBPs do not provide details about the overall composition of a glycan, they provide measurements of the motifs that may be most relevant to the function of a glycan.

GBPs are versatile as detection reagents. Below is a survey—not an exhaustive description of the huge volume of research—of the types of experiments that use GBPs to analyze clinical specimens, particularly for cancer research.

Histochemistry—Histochemistry using GBPs has been an effective means of defining the regions in tissue sections and cell types that express various glycan motifs. The studies of cancer tissue have provided fundamental information about the involvement of glycans in disease. An important class of glycans in epithelial cancers is the Lewis blood group structures and their sialylated and sulfated derivatives. Researchers used monoclonal antibodies against variants of sialyl Lewis A to study the expression of the glycans in pancreatic and colon cancer tissue (Osako et al., 1993; Satomura et al., 1991). Kim and coworkers had developed monoclonal antibodies against unsialylated, monosialylated, and disialylated Lewis A, and they used the antibodies to show that the monosialylated version was generally absent in non-neoplastic colonic tissue, in contrast to the unsialylated and disialylated versions (S. H. Itzkowitz et al., 1988). A study of Lewis X and Y related antigens showed that LeX and sLeX were not expressed in normal pancreas but showed up in 50-70% of pancreatic cancers (S. H. Itzkowitz, et al., 1988; Y. S. Kim et al., 1988). Most pancreatic cancers are positive for CA 19-9 at the tissue level (M. Shimizu, Saitoh, Ohyanagi, & Itoh) and for a related glycan called sialyl Lewis C (Suzuki et al.), detected by the DUPAN-2 antibody. Another study found loss of A, B, and H antigens in pancreatic cancer tissue relative to normal tissue (S. H. Itzkowitz et al.).

The above studies used monoclonal antibodies, but other studies of pancreas tissue made use of lectins. The staining of pancreatic cancer and control cells with peanut agglutinin (PNA) that was tagged with peroxidase, followed by development with standard histochemistry protocols, revealed increased PNA binding to mucins secreted from pancreatic cancer cells (Ching et al., 1988). Histochemical studies of the Tn and sialyl-Tn antigens used both lectins and monoclonal antibodies. Lectins that are specific for alpha-linked GalNAc, such as *Vicia villosa* agglutinin (VVA) are useful for detecting Tn, and monoclonal antibodies have been useful for detecting sTn. Studies using such reagents found that most pancreatic cancers express sTn, whereas sTn is not found in the normal pancreas (Ching, Holmes, Holmes, & Long, 1994; S. Itzkowitz et al., 1991; Schuessler et al., 1991). A study of precursors to pancreatic cancer, pancreatic intraepithelial neoplasias (PanINs) showed that sTn expression begins at the PanIN3 stage (G. E. Kim et al.), which is late in precursor development and before the development of invasive cancer. The Tn and sTn antigens also are expressed in another type of precursor, IPMNs (Terada & Nakanuma).

A limitation in the typical use of GBP in histochemistry is the lack of information about the proteins on which the motifs are located; the experiments simply reveal the location of the glycan. David and coworkers developed a method for uncovering the molecular conjugation of a protein and a glycan in a tissue section, thus detecting a particular protein glycoform. The method uses proximity ligation (Weibrecht et al., 2010), which employs nucleic acid

tags on a pair of detection reagents specific for the potentially linked partners. If the partners—for example the protein backbone and the glycan—are in immediate proximity, a ligase enzyme is able to ligate the two nucleic acid tags on the detection reagents. Once the tags are ligated, a DNA polymerase can amplify the sequence to enable detection of the resulting product. Using an antibody against a mucin protein as one reagent, and a lectin as the other, the researchers were able to detect various glycoforms of mucins in tissue sections ([Pinto et al.](#)). The team found that the protein MUC2 is the dominant carrier of the sialyl Tn glycan in gastric cancer ([Conze et al.](#)).

A related mode of using GBPs is to detect glycans on proteins that had been fractionated by electrophoresis or chromatography. For example, researchers used lectins to identify cancer-associated glycan variants on the serum proteins alpha-fetoprotein ([K. Shimizu et al., 1996](#)), haptoglobin ([Okuyama et al., 2006](#); [Thompson, Cantwell, Cornell, & Turner, 1991](#)), a-1-acid glycoprotein (van Dijk, Havenaar, & Brinkman-van der Linden, 1995), and a-1-antitrypsin ([Thompson, Guthrie, & Turner, 1988](#)).

Imaging—An important medical application of the detection of glycans was recently shown in the imaging of glycans in patients with Barrett's esophagus ([Bird-Lieberman et al., 2012](#)). Fitzgerald and coworkers developed a system to spray fluorescein-labeled WGA onto a region of the esophagus and then detect fluorescence using an endoscope. WGA binding to the esophagus was higher in areas with high-grade dysplasia, presumably owing to the overexpression of N-acetylglucosamine in particular presentations, which provided improved detection of high-grade dysplasia relative to standard imaging. This result confirms that glycans are good indicators of progression towards malignancy and demonstrates the use of glycan detection in a clinical setting.

Lectin affinity capture—Researchers seeking to isolate glycoproteins out of a complex mixture have found lectin affinity capture useful. Typically the lectin or antibody is tethered to a bead to allow capture, isolation, and release of the proteins and lipids that display the targeted glycan motif. This type of experiment is particularly useful when coupled to mass spectrometry, as shown in a method to identify N-linked glycoproteins through quantitative mass spectrometry analysis of lectin-captured material ([Kaji et al., 2003](#)). Hancock and coworkers mixed lectins in column chromatography in order to isolate a broader range of glycoproteins than could be isolated using any single lectin ([Yang & Hancock, 2004](#)). In some cases, researchers may be interested in identifying the proteins that carry a particular glycan motif. For that goal, one could perform affinity capture with just one GBP to target the motif of interest and then perform mass spectrometry to identify the captured proteins. Researchers used this approach to identify protein carriers of the sialyl Lewis X ([Cho, Jung, & Regnier, 2008](#)) and sialyl Lewis A glycans ([Yue et al., 2011](#)).

Antibody-lectin sandwich assays—Antibody capture assays are useful because they enable the detection of a glycan motif on a particular protein captured out of a biological sample. An antibody attached to a solid support provides capture and isolation of a specific protein, and a GBP provides a measurement of the glycan motifs on the captured proteins. Thus an antibody-lectin sandwich assay gives information about the glycoforms of a protein, which is useful because the glycosylation state of a protein may be critically important to its

function or its involvement in disease. We previously showed with this method that the abundances of certain proteins do not change between healthy and diseased populations, but the glycosylation states do (S. Chen et al., 2007; Yue et al., 2009). As a result, measuring the glycans on specific proteins provided improved biomarker performance.

Researchers have performed antibody-lectin sandwich assays in microtiter plates to analyze glycoforms of alpha-fetoprotein (Aoyagi et al., 2002; Korekane et al.; J. T. Wu, 1990), haptoglobin (Thompson, Stappenbeck, & Turner, 1989), prostate-specific antigen (Dwek, Jenks, & Leatham; Meany, Zhang, Sokoll, Zhang, & Chan, 2009), carcinoembryonic antigen (Kumada et al.), and human chorionic gonadotropin (Kelly, Kozak, Walker, Pierce, & Puett, 2005), among others. Instead of in microtiter plates, one can run the assays on glass slides in a microarray format, which provides the parallel capture and detection of multiple proteins in a low sample volume (S. Chen, LaRoche, et al., 2007; Y. Li et al., 2011). The microarray strategy was useful for characterizing the variation in glycoforms on mucins in pancreatic cancer patients (Yue, et al., 2009) and in stimulated cell lines (Y. M. Wu, Nowack, Omenn, & Haab, 2009), and for probing the glycoforms of TIMP-1 (D. Li, Chiu, Zhang, & Chan, 2013). Lampe and coworkers used an antibody array containing over 3000 distinct antibodies to identify proteins that may carry Lewis glycotopes. The researchers incubated serum samples on the array and probed the captured proteins with antibodies against sialyl Lewis A and sialyl Lewis X (Rho et al., 2014). An alternate format for multiplexed detection, instead of planar microarrays, is the bead-based approach, which the Chan and Lubman groups developed for the detection of glycoforms of several proteins in single assays (C. Li et al., 2011; D. Li, Chiu, Chen, Zhang, & Chan, 2013) and Yoneyama and coworkers developed to detect glycoforms of PSA (Yoneyama et al., 2014).

Antibody capture assays offer the ability to probe both protein abundance and protein glycosylation—the former by probing the arrays with detection antibodies targeting the core protein, and the latter by probing the arrays with GBPs. This capability is important for determining whether differences between samples in the levels of a protein glycoform are due to differences in protein abundance or differences in protein glycosylation. The probing of both abundance and glycosylation revealed that the protein endorepellin has a nearly equivalent abundance between pre-cancerous cysts and benign cysts of the pancreas, but a glycoform of endorepellin that interacts with WGA is found only in the pre-cancerous cysts (Cao et al., 2013).

A potential difficulty with antibody capture assays is that the antibodies themselves have glycans, so that the GBPs used to probe the sample could bind to the glycans on the antibodies. An approach to eliminating this source of cross-reactivity is to proteolytically fragment the capture antibody to remove the glycosylated domain of the protein, which was used to develop an assay to detect glycoforms of hCG (Kelly, et al., 2005). Single-chain, recombinant antibodies produced in E-coli also lack glycosylation, and researchers who developed recombinant antibodies against carcinoembryonic antigen (CEA) demonstrated lectin detection of captured CEA without cross reactivity to the immobilized antibodies (Kumada, et al., 2012). Another strategy is to chemically derivatize the glycan on the spotted antibodies to reduce GBP binding to the glycans (S. Chen, LaRoche, et al., 2007). In

this strategy, the glycans on the antibodies were oxidized and then reacted with a hydrazide-containing molecule to hinder GBP binding.

Instead of capturing a protein, one can capture a glycan and then probe the glycans in the captured material, forming a GBP-GBP sandwich assay. The standard CA 19-9 assay uses such a format. In the CA 19-9 assay, an immobilized CA 19-9 antibody captures the sLeA glycan motif, along with all the additional glycans, proteins, and lipids to which it is attached, and another CA 19-9 antibody probes the available sLeA in the captured material. If one used a detection antibody that was different from the capture antibody, the assay would detect co-expression of two different glycan motifs. Researchers have not yet widely used such an approach, but the method has potential for uncovering useful information.

Lectin arrays—The lectin microarray made it practically feasible to obtain glycan measurements on a given sample from multiple, different lectins. By incubating samples on an array of lectins and determining the amount of binding to each lectin, a broad profile of the glycans present in the sample can be rapidly obtained with minimal sample consumption. This approach has many advantages over standard methods of glycan analysis, such as reduced cost, time, and sample consumption, with increased reproducibility. An additional advantage is that lectins can provide information about linkages between monosaccharides (for example whether the alpha or beta configuration), which is not discernable using mass-spectrometry analysis.

The major application of lectin microarrays has been to rapidly investigate the glycosylation of purified glycoproteins that were incubated on the arrays. For example, Kuno and coworkers used arrays containing 39 lectins to detect glycosylation differences between various glycoproteins and changes in glycosylation induced by treatment with glycosidases (Kuno et al., 2005). The incubation of purified proteins, as opposed to mixtures of proteins, is important to simplify the interpretation of the data, so that one may know the identity of the protein binding each lectin. However, others have demonstrated the incubation of complex mixtures of proteins onto lectin arrays, thus achieving a summary view of a cell “glycome.” Pilobello and coworkers used a ratiometric approach to examine changes in bacterial cell-surface glycomes (Pilobello, Slawek, & Mahal, 2007). Isolated membrane proteins from two bacterial cultures were differentially labeled with Cy3 and Cy5 fluorescent dyes and co-incubated on arrays containing up to 58 different lectins. The Cy3/Cy5 ratio at each spot provided a sensitive indicator of differences between the cultures and allowed for normalization between arrays. This analysis enabled the observation of glycosylation changes occurring in response to cell differentiation. The evanescent-field fluorescence method was applied to the study of crude glycoproteins extracted from mammalian cells (Ebe et al., 2006). While the approach of incubating multiple proteins on lectin arrays offers a summary view of the glycan structures on a cell, it has the disadvantage of integrating information from all proteins, so that glycan changes that occur only on a subset of proteins may be lost in a background of non-changing proteins.

A view of cell-surface glycosylation has been achieved by incubating live cells on the surfaces of lectin microarrays. The use of whole cells as opposed to cell extracts has an advantage of preserving higher-order structures, which may be biologically significant and

important for lectin binding. Early work by Zheng et al. (Zheng, Peelen, & Smith, 2005) used covalent immobilization of lectins on self-assembled monolayers that were functionalized with NHS. Cultured cells were incubated on the spotted lectins, and the binding of the cells to the lectins was visualized with an inverted microscope. The gold base substrate was thin enough to allow the imaging. The authors showed differences in the glycosylation of the two cell types. In later work by the same group (S. Chen, Zheng, Shortreed, Alexander, & Smith, 2007), the authors used this technology to explore glycan differences between normal and breast cancer cell lines. Significant variation in glycosylation was identified which correlated with metastatic potential as well as metastatic location preference.

Lectin microarrays also were used to examine dynamic changes to *E. coli* bacterial glycosylation (Hsu, Pilobello, & Mahal, 2006). The bacteria were labeled with a dye that binds to DNA to allow detection by fluorescence after incubation on the arrays. The authors could distinguish *E. coli* strains based on glycosylation and could observe growth-dependent variation in glycosylation on particular strains. Lectin arrays employing evanescent-field fluorescence (Kuno, et al., 2005), were used to examine dynamic changes to the cell surfaces glycomes of mammalian cells that had been fluorescently labeled with a DNA-binding dye (Tateno et al., 2007). Alterations in lectin-binding patterns were seen in glycosylation-defective mutants of CHO cells and in splenocytes from mice with a genetic knockout of a glycosyltransferase gene. Changes in cell surface glycosylation associated with erythroblast differentiation also were observed. Another study using arrays of 94 lectins and a similar detection method examined the GBP-binding signatures of 24 different human cell lines and predicted functional phenotypes based on lectin-binding profiles (Tao et al., 2008).

A twist on this strategy is to immobilize the lectins, then incubate a sample on the lectin array, and then probe the captured material with an antibody (Kuno et al., 2009). Each lectin captures a particular glycan motif, along with all proteins carrying that motif, and the detection antibody identifies how much of a particular protein was captured at each lectin.

Defining the fine specificities of GBPs from glycan array data

To effectively use GBPs for probing biological samples, a researcher needs to know, as accurately as possible, what each GBP binds. Typically, researchers are aware of a GBP's specificity in qualitative terms, covering only the primary, simplified binding of the GBP. For example, the specificity of the GBP from *Aleuria aurantia* (AAL) usually is defined as alpha-linked fucose, and the specificity of wheat germ agglutinin is listed as GlcNAc. But the actual specificities are more complicated. AAL indeed binds alpha-linked fucose, but not in every presentation of fucose, and in some presentations much better than others. Such nuances hold true for nearly every GBP. Certain GBPs strongly bind a primary glycan motif but also bind other, related motifs more weakly. For example, the lectin from the snail species *Helix pomatia* binds terminal, alpha-linked N-acetylgalactosamine but also binds terminal, alpha-linked N-acetylglucosamine (Markiv, Peiris, Curley, Odell, & Dwek, 2011); WGA mainly binds GlcNAc but also binds sialic acid and GalNAc in certain presentations;

and concanavalin A mainly binds alpha-linked mannose in certain linkages but also terminal, alpha-linked glucose.

The task of defining the specificities of GBPs has been greatly helped by glycan array technology ([Blixt et al., 2004](#); [Culf, Cuperlovic-Culf, & Ouellette, 2006](#); [Drickamer & Taylor, 2002](#); [Liu, Palma, & Feizi, 2009](#); [Stevens, Blixt, Paulson, & Wilson, 2006](#)). A glycan array is the reverse of a lectin array; it holds immobilized glycans to be probed by GBPs. A glycan array experiment can provide measurements of the binding of a GBP to hundreds of different glycans in a single experiment (Fig. 1A), while using a small amount of each glycan and the sample ([Liu, et al., 2009](#)). With such data, we can compare the levels of binding among the glycans to discern the rules that govern the GBP's binding. Before glycan microarrays, serial testing of GBP binding to each glycan was necessary, requiring so much time and material that testing of large numbers of glycans was impractical. Several laboratories have developed glycan arrays, each with distinct technologies and glycans (reviewed in ([Rillahan & Paulson, 2011](#))).

Developers of glycan arrays are continually increasing the diversity and scope of the arrays. These developments bring the opportunity to better characterize the rules that govern GBP binding, but they also bring the challenge of increased difficulty understanding what the data mean. Because of the structural complexity of some oligosaccharides, and because certain GBPs may have multiple, related specificities, the task of sifting through glycan array data to discern binding specificities can be difficult and time-consuming. A software program for determining binding specificities from glycan array data could ease this task as well as add definable and quantifiable interpretation to the data. In addition, the ability to automate glycan array analysis would enable the cataloging and comparisons of many datasets, which could be used for searching and higher-level analyses.

The above considerations prompted us to develop the Motif Segregation algorithm ([Porter et al., 2010](#)) and the GlycoSearch software ([Kletter, Cao, Bern, & Haab, 2013](#)) for analyzing glycan array data. An additional method, called Outlier Motif Analysis, builds on Motif Segregation to enable increased detail in the identifications of fine specificities ([Maupin, Liden, & Haab, 2011](#)). Our strategy begins with identifying each oligosaccharide in terms of its component motifs (Fig. 1B). We define numerous motifs that frequently appear in various types of glycans and compute the presence or absence of each motif on each glycan on the array. From there we can search for motifs that are always present in the glycans bound by a lectin but never present in the glycans not bound. For a given motif, we segregate the glycans into two groups—those that have the motif and those that do not—and compare the binding of the lectin between the groups using robust Mann-Whitney statistics (Fig. 1C). If the lectin has statistically higher binding in one of the groups, the motif could represent the binding determinant (or at least part of the determinant). For convenient ranking of the motifs, we calculate a “Motif Score” by log-transforming the p value of the segregation and giving it a sign according to the direction of the difference. We repeat this process for each motif. The motifs with the highest scores are the best candidates as the binding determinant of the lectin.

The output of GlycoSearch provides information that helps the user determine how well certain motifs describe the binding activity of the GBP. Each motif tested by GlycoSearch is given a score indicating the preference of the GBP for that motif, and the user can assess whether the top-scoring motifs are present in the glycans bound by the GBP and absent in the glycans not bound. If the top-scoring motifs do not fit the data well, the user can define new or modified motifs, based on inspection of the outlier glycans, and rerun the analysis. The motifs can be as simple or as complicated as necessary. We have found this iterative approach to be effective for identifying the motifs that best fit the data. A recent publication gives detailed instructions on the use of the GlycoSearch analysis program ([Kletter, Cao, et al., 2013](#)).

Other algorithms have subsequently appeared, including methods called Quantitative Structural Activity Relationship (QSAR) ([Xuan, Zhang, Tzeng, Wan, & Luo, 2011](#)) and GlycanMotifMiner ([Cholleti et al., 2012](#)). With these tools we have a way to precisely and numerically define the specificity of a GBP, such that the specificity of a GBP can be defined in terms of a set of motifs with associated motif scores.

Need for the expansion of glycan arrays—In some cases, the glycan array does not tell the whole story; the data are only suggestive of a particular binding preference of a GBP. For example, if a particular motif appeared only on a single glycan on an array, and a GBP showed higher or lower binding to that glycan, one could not make a conclusion based only on that single observation about the preference of the GBP for the motif. To further test the binding to the motif, one would need multiple glycans containing the motif.

A few options are available to meet the need for expanded content of glycan arrays. One option is to use more than one glycan array platform, each platform containing a distinct set of glycans. The combined results from the diverse platforms could give more information than the results from any individual platform. A motif-based analysis would enable cross-platform integration, since it transforms the data from the divergent glycans on each platform into a common set of motifs. Mahal and coworkers used such a strategy to compare GBP binding between six different glycan array platforms, arriving at new insights into GBP binding ([L. Wang et al.](#)). The GlycoSearch software has the ability to handle data from any platform by automatically reading the glycans sequences or IUPAC formulas and determining whether each motif is present or absent in each glycan.

Another option is to synthesize new glycans to meet a particular need. For example, researchers studying the specificities of influenza viruses synthesized glycans to cover a range of sialic acid presentations in order to investigate what presentations of sialic acid were preferred by various virus strains ([Nycholat et al., 2012](#); [Padler-Karavani et al., 2012](#); [Song, Yu, et al., 2011](#)). But the synthesis of new glycans is challenging, time-consuming, and requires much expertise. Boons and coworkers made a significant advance in glycan synthesis methodology, developing a new way to rapidly synthesize glycans with controlled extensions in specified arms of N-glycans ([Z. Wang et al., 2013](#)). The ability to rapidly synthesize new glycans could be particularly valuable to test binding around a known cancer motif. For example, the sialyl Lewis A antigen is strongly overexpressed in pancreatic cancer, and several monoclonal antibodies are available to detect sLeA, but questions

remain about whether the antibodies' affinity to sLeA remains as high when the sLeA motif is slightly modified or presented in differing ways. An analysis of various antibodies (Partyka, Maupin, Brand, & Haab, 2012) showed differences in specificity: some were highly specific—only binding sLeA—and others also bound related motifs.

Regardless of the specific capture and detection method, much important information still remains difficult to ascertain from current arrays. Little is still known about how sulfation, branching, repeating units, or length of the glycan affect binding. Previous studies showed these factors to be important in some cases, so new glycans covering the various glycan motifs would certainly be useful.

Yet another approach for creating new content for glycan arrays is the Shotgun Glycomics method (Song, Lasanajak, et al., 2011). To produce material for shotgun glycan arrays, the glycans from a biological sample are isolated, tagged with a fluorophore, and fractionated by multidimensional chromatography to near purity of each glycan. The resulting Tagged Glycan Library (TGL) is then printed in microarrays. A typical way of using this technology is to probe the arrays with GBPs or glycan-binding antibodies to find glycans of interest, and then to characterize the glycans using mass spectrometry and other methods. The strategy was used to uncover much information about the glycome of human milk (Yu et al., 2012). Cummings and coworkers used monoclonal antibodies that had been generated from inoculations with stem cells, followed by characterizations by MS, to show that human milk contains glycans associated with stem cells. The milk glycome also showed an enrichment of blood group related epitopes as well as ligands for pathogens, which may serve as decoys to prevent infection of the infant gut (Yu et al.).

Advantages of this method are that it facilitates discovery of new glycans, not just those previously described, and that it does not require glycan synthesis. Disadvantages are that it involves significant labor and material to produce the arrays and that it requires characterization of the glycans after purification. In any case, the Shotgun Glycomics method promises to be a good complement to standard glycan arrays and to be useful for characterizing the glycomes of samples.

Higher order influences on GBP binding: density, location, and accessibility

Glycan arrays give information about the primary motif that a GBP binds, but relatively little information—in most of the current glycan arrays—about other factors that may affect binding, such as density, branching, and accessibility. The characterization of these influences is important to fully understand how to use GBPs to probe biological samples and to interpret the resulting data.

Density—Lectin-glycan interactions in biology frequently rely upon multivalent interactions to effect their functions. Some lectins have a weak affinity for their target glycan, typically in the μM to mM range (Collins & Paulson, 2004), at least when used as a reagent outside the biological context. In such systems, the strength of any one protein-glycan interaction is not great, but linking several such interactions together increases the overall interaction strength. The galectin family of lectins uses a well-described system of forming lattices in order to induce downstream effects (Brewer, Miceli, & Baum, 2002).

Galectin-3 can self-associate and potentially form oligomers when induced by a molecule displaying multiple glycan ligands ([Lepur, Salomonsson, Nilsson, & Leffler](#)), such as lipopolysaccharide ([Fermino et al., 2011](#)). Plant lectins also have multivalent forms that have increased binding to cell wall glycans relative to their monovalent forms, and bacterial and viral infections typically rely on multivalent interactions to achieve necessary binding strengths ([Saba, Dutta, Hemenway, & Viner, 2012](#)). In some cases, the clustering of glycan ligands on a cell surface by a bound lectin induces signaling, as in the case of cross-linking cell surface glycans by galectin-1 to induce apoptosis in human cells ([Pace, Lee, Stewart, & Baum](#)).

A study by Dam and Brewer provides insights into the energetics involved in matching lectin orientation to glycan clustering ([Chaffer et al., 2013](#)). In their model, the best improvement of lectin affinity to glycoproteins occurs in face-to-face interactions, in which multiple binding sites of a lectin multimer are aligned in a common direction and bind to glycoproteins with clustered epitopes. Under a contrasting scenario, a tetramer-shaped lectin with binding sites in four directions would have lower binding efficiency to linear glycan epitopes.

Glycan arrays or glycopeptide arrays potentially could give additional information about the effect of multivalent glycan presentations on binding from multimeric lectins. For example, Gildersleeve and coworkers attached various amounts of glycan structures to the lysine residues of BSA to achieve a range of glycan densities, from which they observed preferential binding of certain lectins and antibodies to specific glycan with higher densities ([Oyelaran, Li, Farnsworth, & Gildersleeve, 2009](#)). The same group also varied glycan density by mixing various concentrations of a non-glycosylated protein with a fixed amount of glycosylated protein, which they used to identify multivalent inhibitors of lectin-glycan interactions ([Y. Zhang, Li, Rodriguez, & Gildersleeve, 2010](#)). Bertozzi and Godula varied glycan density using synthetic polymers with diverse spacing between attached N-acetyl-lactosamine groups, which enabled measurements of crosslinking between glycans by lectins ([Godula & Bertozzi, 2012](#)). These types of arrays may reveal unexpected effects. For example, bauphia purpurea lectin (BPL) was found to switch carbohydrate ligands as a function of the density of glycan-modified, self-assembled monolayers ([Horan, Yan, Isobe, Whitesides, & Kahne, 1999](#)). Such a finding highlights the need for in-depth characterization of lectin binding under many conditions.

Location of a motif within a glycan—Recent experiments have shown that the location of a motif within the overall structure of a glycan can affect the amount of lectin binding. In a study of the factors influencing the binding of the hemagglutinin protein from the influenza virus to host glycans, researchers developed microarrays containing glycans with sialic acid in various presentations ([Nycholat, et al., 2012](#)). HA bound to its primary sialic acid target in each glycan, but the level of binding varied according to the location of the sialic acid within the glycan; factors such as amount of extension, number of branches, and type of branching affected binding. A study of the epitopes of three monoclonal antibodies against the exopolysaccharide coat of the *Pseudomonas aeruginosa* pathogen also employed synthetic glycans containing the epitopes in various locations within the glycans. As with

HA, the reactivity of the antibodies was highly dependent upon the location of the primary glycan motif within the glycan ([H. Li et al., 2013](#)).

Another study examined the effects of differing lengths of branches ([Z. Wang, et al., 2013](#)). As referenced above, Boons and coworkers devised a practical and efficient means of producing N-glycans with defined, asymmetric branches. Glycan arrays printed with the defined N-glycans were useful for investigating whether the location of a particular motif in reference to asymmetrical branching affected the binding of particular GBPs. The authors found that the hemagglutinin protein from the influenza virus had differential binding depending on whether its epitope was on a single extension, symmetric branches, or asymmetric branches. Only with glycans having such precisely defined branch structures can researchers more fully explore the importance of asymmetric branching in GBP binding.

In some cases a glycan could contain a motif that normally is recognized by a GBP, but the motif is not accessible to the GBP because of steric hindrance from the environment. A study of glycan arrays containing glycans with variable distances between the binding determinant and the array surface showed that a lack of accessibility can hinder binding ([Leteux et al.](#)). Computational modeling by Woods and coworkers, combined with results from glycan array studies, provided more insights into steric hindrance when a glycan motif is close to a surface ([Grant, Smith, Firsova, Fadda, & Woods, 2014](#)). If the GBP must orient itself in such a way that it collides with the surface, binding to the motif is hindered. Such a relationship might also affect protein-glycan interactions on cell surfaces.

Quantitative interpretation of GBP measurements

In the typical use of GBPs to probe for the presence of a glycan motif in biological samples, the researcher infers the presence or absence of the primary target of the GBP based on the amount of GBP binding. Usually the researcher judges the amount of GBP binding in qualitative terms and does not account for the subtleties of the GBP binding behavior. This approach is subject to much error. In the first place, the specificity of a GBP usually is more complicated than the primary specificity used in the interpretation. In the second place, the lack of quantitation can lead to imprecise interpretation of the amount of glycan present.

A strategy to more accurately interpret GBP data is to use quantified measurements in combination with quantified information about the specificity of each GBP. For example, when interpreting the binding of a GBP to a sample, instead of making a judgment based on experience and personal knowledge, the researcher could use an algorithm to give the probabilities that various glycan motifs are present in the sample. Such quantitative interpretation could more accurately account for the complexities in GBP specificities and would remove the burden from researchers for acquiring a detailed knowledge of the subtleties of each GBP. In addition, quantitative interpretation could enable the use of combined measurements from multiple GBPs to get more information about a sample. An individual GBP can give good information about the presence of a motif, but the use of several GBPs together could account for some of the ambiguities of individual GBPs.

Glycan array analyses make the above strategy possible. We recently demonstrated an algorithm for quantitative interpretation of GBP binding, called Motif Prediction ([McCarter](#)

et al., 2013). In a typical experiment, a researcher would use a series of GBPs to probe for the presence of selected motifs, and the computer algorithm would interpret the data (Fig. 2A). Two pieces of information are available for each GBP measurement: the previously-determined likelihood that the GBP binds each motif (the motif scores from glycan array data), and the amount of binding to the unknown sample. If a GBP has a strong motif score for a particular motif, a high amount of binding predicts the presence of the motif, whereas weak binding predicts the absence of the motif. In contrast, if a GBP has a weak motif score for the motif in question, the amount of binding is not predictive of the presence or absence of the motif.

Therefore, for each GBP measurement and for each motif of interest, the algorithm multiplies the measurement of GBP binding by the motif score for that GBP. To arrive at a final prediction score for each motif, it adds the contributions from each GBP (Fig. 2A). (The contributions from the individual GBPs would be additive, given that each GBP is independent.) The final motif prediction (MP) score for each motif, indicating the likelihood that a motif is present in a glycan, is $MP_{M1} = (I_{L1} \times M_{M1,L1}) + (I_{L2} \times M_{M1,L2}) + (I_{L3} \times M_{M1,L3}) + \dots$ etc. for additional GBPs, where MP_{M1} is the motif prediction score for motif 1, I_{L1} is the intensity (amount of binding) of GBP 1, $M_{M1,L1}$ is the motif score for motif 1 and GBP 1, and so on for additional GBPs. We showed that this method was effective for predicting motifs present on glycan arrays (Fig. 2B)—a system for which we knew the correct answers—and we also demonstrated interpretation of the motifs present on MUC1 produced from various cell lines (McCarter, et al., 2013).

The quantitative interpretation of GBP binding could be useful in several types of studies, especially those involving relationships between glycans and phenotypes in clinical samples. For example, researchers could examine whether unusual protein glycoforms have increased levels in pre-neoplastic tissue; whether the glycoforms are more abundant in individuals with specific genotypes; whether the expression of particular glycosyltransferases is linked to protein glycosylation; if differences in protein glycosylation exist between humans and mice or other model systems; or whether drug or cytokine treatments affect protein glycosylation. Furthermore, manufacturers of protein and antibody drugs are interested in characterizing the effects of protein glycosylation on drug activity and retention (Astronomo & Burton, 2010). With further development of the capabilities suggested here, drug researchers could retrieve a protein from the *in vivo* setting and determine the relative amounts of each glycoform. Such analyses could be effectively performed on GBP arrays (Kuno, et al., 2005; Pilobello, Krishnamoorthy, Slawek, & Mahal, 2005).

Linking with mass spectrometry data—A limitation of GBP-based data is that it does not give information about the overall structure of the glycan; it provides a measurement only of the GBP's binding determinant. An approach to getting complementary information about the glycans bound by a GBP is to link GBP measurements with mass spectrometry (MS) measurements (Yu, et al., 2012). MS analysis provides the monosaccharide composition of glycans and some sequence information, but it leaves ambiguities about sequence and linkage variants (Adamczyk, Tharmalingam, & Rudd, 2011; Pan, Chen, Aebersold, & Brentnall, 2011). A convenient way to resolve the ambiguities is to probe the glycans with GBPs that differentially bind the variants (Haab, 2010; Hirabayashi, 2004; Ito

et al., 2009; Smith & Cummings, 2013; Yu, et al., 2012). Cummings and Smith and coworkers used the combination of lectin measurements, MS analyses, and glycosidase digestions to improve assignments of structure to glycans isolated from human milk (Yu, et al., 2014), in an approach called metadata-assisted glycan sequencing. The group found that the VP4 outer capsid protein from rotavirus bound to many of the glycans. Analysis of the glycans revealed previously unknown specificities of rotavirus, including non-sialylated N-acetyllactosamine.

The combination of MS and GBP data is clearly powerful, yet it is still imprecise in current practice. Because of the complexities of the binding specificities of GBPs, and the multiple glycan structures that might be revealed in an MS experiment, researchers find it difficult, if not impossible, to precisely link information between a GBP experiment and an MS experiment. A solution to this problem could be to use the quantitative interpretation of GBP binding described above. If we could express the GBP and MS data in a common language, we could develop algorithms for quantitative linkage of the data types. We currently are working on developing such a method, which hopefully will provide practical and accurate analyses of protein glycosylation with low sample consumption and in a format that allows translation to a clinical assay.

Finding the right reagent: mining glycan array data, engineering GBPs, and creating antibodies

Mining glycan array data—Many GBPs are commercially available, yet the range of structures covered by the available GBPs is relatively limited. The commonly used GBPs target basic motifs such as sialic acid, fucose, blood-group glycans, and lactosamine. These structures tend to be immunogenic in foreign organisms and likely are particularly amenable to high-affinity binding by proteins. But many other glycan structures appear in biology and may be important to detect. For example, an unusual sialylated structure, the type 1 H antigen sialylated at the 6' carbon (Sia α 2,6(Fuca1,2)Gal β 1,3GlcNAc), was recently found in the pancreas of some individuals with pancreatic cancer (Shida et al., 2010). Such structures likely have physiological significance, but none of the GBPs available through commercial sources detect them.

It may be possible, in many cases, to meet this need by making use of the enormous diversity of GBPs in biology. Lectins are present in every cell of every organism, and GBPs with specificities previously unobserved are being discovered on a regular basis. But making use of this information is a difficult task. This information is often in obscure journals, and the characterizations of specificities are minimal.

An initial step in making a greater number of GBPs available for research will be to assemble detailed information about the specificities of GBPs in a searchable and well-annotated format. The data available from the CFG provided an opportunity to meet that need. The CFG has since 2004 provided high-quality glycan microarray analyses to participating investigators (Blixt, et al., 2004). The investigators can send a protein of interest to the CFG, and the Consortium provides a report of the binding to each of the glycans on the array. The CFG has performed several thousand experiments, covering

hundreds of different glycan binding proteins, and the data are available to the public via a web portal.

Using software to automate the GlycoSearch analysis, we processed the entire repository of CFG data—currently over 3000 data sets—and assembled the information, along with the metadata for each experiment, into a relational database (Fig. 3A) (Kletter, Cao, et al., 2013; Kletter, Singh, Bern, & Haab, 2013). To ensure the usefulness of the “GlycanBinder” database, we checked, corrected, and enhanced the metadata, which includes sample information, experimental conditions, and the purpose of the experiment. The initial data had been freely entered by each participating investigator without strict guidelines and independent confirmation of correct entry, so the entries were often inconsistent and in some cases incomplete. We enhanced the information by adding the searchable fields of “Species”, “Species Class”, and “Protein Family”. The Species field indicates the species from which the GBP was taken (in the case of antibodies, it is the species in which the antibody was generated); the Species Class field indicates the larger class, such as mammalian, viral, or bacterial; and the Protein Family field indicates the larger family to which the GBP belongs, such as galectin, C-type GBP, or adhesin, etc.

The GlycanBinder database has tremendous potential as a resource for finding GBPs to meet particular needs, particularly for researchers who do not have a broad knowledge of GBPs. Acquiring such knowledge and identifying a suitable GBP could be daunting, and even those involved in GBP studies cannot be deeply acquainted with all GBPs for which data are available. But with the GlycanBinder database, the researcher could simply query the database with a motif and inspect the output to find the GBPs with the highest scores for that motif. A graphical representation of the motif scores of the queried motif and related motifs helps to determine the diversity in specificities between the GBPs, as we demonstrated in a search for binders against sialyl Lewis A (Fig. 3B). Some of the binders also recognized sialyl Lewis C, which differs from sialyl Lewis A in the absence of fucose, and sialyl Lewis X, a structural isomer of sialyl Lewis A. Searches like this one, combined with comparative analyses of the identified binders, could uncover promising and unexpected GBPs.

Engineering GBPs—If a GBP or antibody with the desired specificity is not available, it may be possible to modify the sequence of a GBP to achieve alterations in specificity to meet the need. For example, a GBP might bind several related motifs, but the researcher wants to limit the specificity to only a single motif. Or it may be possible in some cases to introduce a new specificity (Arnaud, Audfray, & Imberty, 2013).

Hirabayashi and coworkers used this strategy with the galectin from agrocybe cylindracea (Imamura, Takeuchi, Yabe, Tateno, & Hirabayashi). The native galectin has high affinity toward sialic acid linked α 2,3 to lactosamine as well as galactose in various linkages, and the researchers wanted to remove the binding of terminal galactose. Using information from the crystal structure, they substituted the amino acid that was critical for binding lactosamine with all other possibilities and found a version that retained the desired specificity but eliminated the additional specificities. The same group engineered a multivalent GBP specific for α 2,6-linked sialic acid (Yabe et al.). Leffler et al. mutated the galactose-binding site of human galectin-3 to produce a version with increased specificity for poly-N-

acetyllactosamine and increased ability to activate leukocytes (Salomonsson et al.). Mehta and coworkers sought to reduce the range of binding of the *Aleuria aurantia* GBP (AAL), which normally binds fucose in any linkage (Romano et al.). Specific mutations resulted in differential binding to the various linkages.

Other groups have sought to modify the valency of GBPs through altering the ability of GBP monomers to form higher-order multimers. Lectins with fewer subunits, and correspondingly smaller sizes, potentially could be better suited for accessing constrained sites, and GBPs with more subunits might bind with higher avidity or induce biological signaling in new ways. Imberty and coworkers found that the ability of fucose-binding GBPs and toxins to invaginate cell membranes requires specific distances between adjacent binding sites (Arnaud et al.; Arnaud et al.). Pahlsson and colleagues found that a monomeric form of AAL, produced by removing the ability of the GBP to form dimers as in the native state, had reduced hemagglutinating capability but retained the ability to bind fucosylated oligosaccharides (Olausson, Astrom, Jonsson, Tibell, & Pahlsson).

Raising antibodies to specific glycans—In some cases a monoclonal antibody could be preferable to a GBP because of the more predictable behavior of antibodies in immunoassays and as therapeutics. Researchers have raised antibodies against established cancer markers and many other glyco-epitopes (Kannagi & Hakomori). For example, to develop antibodies against the Tn antigen, Blixt and coworkers immunized mice with Jurkat cells, which display high levels of the Tn antigen on their surfaces (Blixt et al.). This procedure resulted in multiple antibodies recognizing Tn, with high diversity in specificities. Because the immunogenicity of whole cells is variable, other researchers have immunized with synthetic glycans conjugated to a carrier protein such as keyhole limpet hemocyanin (KLH) or bovine serum albumin (BSA). Researchers used BSA coated with the Tn antigen to produce anti-Tn IgM antibodies (Danussi et al.). The procedure requires a large amount of purified glycan, which can be costly and difficult to obtain for unusual oligosaccharides, and the glycans might not be immunogenic. A strategy to boost the effectiveness of immunization was to use mice that were deficient for particular sulfotransferases and then immunize the mice with cells that overexpress the same sulfotransferases (Hirakawa et al.). The generated antibodies bound the predicted 6-sulfated glycan structures and were found to be useful for inhibiting lymphocyte adhesion and homing in mice. An alternative to immunizing with cells or with synthetic glycans is to immunize with whole glycoprotein purified from a natural source, as performed by a group that isolated mucin from the salivary glands of sheep to produce antibodies against the sialosyl-Tn antigen (Kjeldsen et al.).

An alternative to immunizing animals is to use display technologies to screen for binders against glycans. In order to obtain antibodies against linear poly-N-acetyllactosamine, known as the i antigen, Valmu and others at the Red Cross Blood Service in Finland developed a library of IgM sequences from a donor who had high blood titers against the i antigen (Hirvonen et al.). Screening the library against the antigen resulted in the production of recombinant single-chain antibodies. Another group used a human scFv phage display library to find a binder against cell surface chondroitin sulfate proteoglycan 4 (CSPG4), a proteoglycan that is highly expressed on cancer cells but that is difficult to produce antibodies against using standard immunizations (X. Wang et al.).

Discovering glycan motifs using GBPs: application to cancer biomarkers

Most studies using GBPs involve measuring known glycans, rather than discovering new motifs associated with disease, but several studies suggest that GBPs also can be used for discovery. The primary approaches to discover glycans using GBPs are the generation of new antibodies and the screening of candidate GBPs.

Discovery by antibody generation—This approach makes use of the immune system of an animal to create antibodies against abnormal glycans, some of which may be unknown. It involves immunizing a mouse, rabbit, or some other host with cancer material, developing monoclonal antibodies from the animal, and testing which antibodies bind selectively to cancer tissue. Some of the most notable cancer markers were found in this way, including some with unusual glycan epitopes.

A prominent example is the CA 19-9 monoclonal antibody, which forms the basis for the CA 19-9 serum test and which is the current best test for pancreatic cancer. The CA 19-9 antibody was developed by immunizing mice with the colon carcinoma cell line SW 1116 (Herlyn, Steplewski, Herlyn, & Koprowski, 1979; Koprowski et al., 1979). Later studies revealed that the antigen was highly expressed in pancreatic cancer (Herlyn, Sears, Steplewski, & Koprowski, 1982) and that it was a sialylated glycolipid (Magnani et al., 1981). The antigen furthermore was found on mucins (Magnani, Steplewski, Koprowski, & Ginsburg, 1983) and was determined to be the sialyl Lewis A carbohydrate (Magnani et al., 1982). Researchers have thoroughly studied the performance of the CA 19-9 serum test for the diagnosis of pancreatic cancer and other diseases (Goonetilleke & Siriwardena, 2007).

Another antigen associated with pancreatic cancer was discovered by monoclonal antibody production. The Dupan-2 antibody was developed by immunizing mice with the HPAF cell line (Metzgar et al., 1982). It was found to stain pancreatic tumor glands and to be elevated in the serum of some patients who did not have elevations in the CA 19-9 test (Takasaki et al., 1988). Epitope analysis of the Dupan-2 antibody showed that, like the CA 19-9 antibody, it bound a carbohydrate antigen (Kawa et al., 1994). The antigen turned out to be very similar to that of the CA 19-9 antibody, having the sequence Sia α 2,3Gal β 1,3GlcNAc β and termed sialyl Lewis C.

Another example shows the discovery of an unusual structure. In an effort to develop mAbs against metastatic, androgen-independent prostate cancer, researchers immunized mice with the PC3 cell line and developed antibodies from the resulting immune response (Carroll et al., 1984). One of the antibodies, called F77, is specific for prostate cancer cells and suppresses the growth of prostate cancer growth xenografts (G. Zhang et al., 2010). Subsequent analysis of the epitope showed that the antibody binds a blood-group H glycan appearing on branches of glycans (Gao et al., 2014; Nonaka et al., 2014). Such a structure would not be easily found by other means. This result shows the potential value of using the immune system to find glycan biomarkers. As a general strategy, however, it is inconsistent and unpredictable.

Screening candidates—Another approach to discover new carbohydrate antigens is to screen multiple GBPs. Based on hypotheses about the categories of motifs that are

associated with cancer, one could collect an assortment of GBPs covering the hypothetical space of motifs. Through the use of high-throughput methods, a researcher could acquire measurements from many different GBPs and use motif prediction to quantify the likelihood of specific motifs being associated with disease. Furthermore, one could take an iterative approach toward discovery, by using the output from the first round of screening to suggest additional GBPs that more specifically target the predicted motifs. The availability of a searchable database of glycan array analyses facilitates such a strategy, as does high-throughput, low-volume methods (S. Chen & Haab, 2009; Forrester, Kuick, Hung, Kucherlapati, & Haab, 2007; Haab, Partyka, & Cao, 2013) of running antibody-GBP sandwich assays. We used this strategy to find a glycoform of MUC5AC that is prevalent in the cyst fluid of patients with pre-cancerous pancreatic cysts (Cao, et al., 2013) and that is defined by terminal N-acetylglucosamine—an atypical structure that may be elevated in gastric cancers (Karasawa et al., 2012).

Conclusions and prospects

Lectins and glycan-binding antibodies are powerful tools for the detection and discovery of glycan motifs in biological samples. Recent developments are opening up new capabilities, such as increased detail in motif characterization, the ability to analyze many samples using small amounts of each sample, and quantitative interpretation enabling higher accuracy and linkage with diverse data types. Lectin engineering, new antibody development, and the continued discovery and characterization of new GBPs are further expanding the diversity of glycan motifs detectable by GBPs. An exciting possibility is that databases of quantitative information about GBPs, and software to assist with interpreting data from GBPs, will enable researchers who are not experts in the field to effectively use of GBPs, thus expanding the applications of GBPs.

Considering the persistent shortcoming in knowledge about glycans in many areas of biology and pathology, the improved methods could yield important information. If researchers can better characterize the glycans associated with disease—the motifs, the cells where they are expressed, the conditions in which they occur, and their associations with outcomes and other information—we will have a better starting point for studying mechanisms. Such information also is the starting point for developing tools to improve outcomes for patients afflicted with diseases characterized by changes in glycans. In addition, improved studies of glycans in clinical samples could help clarify the size of this class of diseases, which could be larger than currently appreciated.

References

- Adamczyk B, Tharmalingam T, Rudd PM. Glycans as cancer biomarkers. *Biochim Biophys Acta*. 2011; 1820(9):1347–1353. [PubMed: 22178561]
- Aoyagi Y, Mita Y, Suda T, Kawai K, Kuroiwa T, Igarashi M, et al. The fucosylation index of serum alpha-fetoprotein as useful prognostic factor in patients with hepatocellular carcinoma in special reference to chronological changes. *Hepatol Res*. 2002; 23(4):287. [PubMed: 12191676]
- Arnaud J, Audfray A, Imberty A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chemical Society Reviews*. 2013; 42(11):4798–4813. [PubMed: 23353569]

- Arnaud J, Claudinon J, Trondle K, Trovaslet M, Larson G, Thomas A, et al. Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity. *ACS Chem Biol*. 2013; 8(9):1918–1924. [PubMed: 23855446]
- Arnaud J, Trondle K, Claudinon J, Audfray A, Varrot A, Romer W, et al. Membrane Deformation by Neolectins with Engineered Glycolipid Binding Sites. *Angew Chem Int Ed Engl*. 2014 epub ahead of print.
- Astronomo RD, Burton DR. Carbohydrate vaccines: developing sweet solutions to sticky situations? *Nat Rev Drug Discov*. 2010; 9(4):308–324. [PubMed: 20357803]
- Bird-Lieberman EL, Neves AA, Lao-Sirieix P, O'Donovan M, Novelli M, Lovat LB, et al. Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. *Nat Med*. 2012; 18(2):315–321. [PubMed: 22245781]
- Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, et al. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A*. 2004; 101(49):17033–17038. [PubMed: 15563589]
- Blixt O, Lavrova OI, Mazurov DV, Clo E, Kracun SK, Bovin NV, et al. Analysis of Tn antigenicity with a panel of new IgM and IgG1 monoclonal antibodies raised against leukemic cells. *Glycobiology*. 2012; 22(4):529–542. [PubMed: 22143985]
- Brewer CF, Miceli MC, Baum LG. Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. *Curr Opin Struct Biol*. 2002; 12(5):616–623. [PubMed: 12464313]
- Cao Z, Maupin K, Curnutte B, Fallon B, Feasley CL, Brouhard E, et al. Specific glycoforms of MUC5AC and endorepellin accurately distinguish mucinous from non-mucinous pancreatic cysts. *Molecular & Cellular Proteomics*. 2013; 10:2724–2734. [PubMed: 23836919]
- Carroll AM, Zalutsky M, Schatten S, Bhan A, Perry LL, Sobotka C, et al. Monoclonal antibodies to tissue-specific cell surface antigens. I. Characterization of an antibody to a prostate tissue antigen. *Clin Immunol Immunopathol*. 1984; 33(2):268–281. [PubMed: 6488592]
- Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F, et al. Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity. *Cell*. 2013; 154(1):61–74. [PubMed: 23827675]
- Chen S, Haab BB. Analysis of glycans on serum proteins using antibody microarrays. *Methods Mol Biol*. 2009; 520:39–58. [PubMed: 19381946]
- Chen S, LaRoche T, Hamelinck D, Bergsma D, Brenner D, Simeone D, et al. Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nat Methods*. 2007; 4(5):437–444. [PubMed: 17417647]
- Chen S, Zheng T, Shortreed MR, Alexander C, Smith LM. Analysis of cell surface carbohydrate expression patterns in normal and tumorigenic human breast cell lines using lectin arrays. *Anal Chem*. 2007; 79(15):5698–5702. [PubMed: 17580952]
- Chen WC, Completo GC, Sigal DS, Crocker PR, Saven A, Paulson JC. In vivo targeting of B-cell lymphoma with glycan ligands of CD22. *Blood*. 2010; 115(23):4778–4786. [PubMed: 20181615]
- Ching CK, Black R, Helliwell T, Savage A, Barr H, Rhodes JM. Use of lectin histochemistry in pancreatic cancer. *J Clin Pathol*. 1988; 41(3):324–328. [PubMed: 3283174]
- Ching CK, Holmes SW, Holmes GK, Long RG. Blood-group sialyl-Tn antigen is more specific than Tn as a tumor marker in the pancreas. *Pancreas*. 1994; 9(6):698–702. [PubMed: 7846011]
- Cho W, Jung K, Regnier F. Use of Glycan Targeting Antibodies To Identify Cancer-Associated Glycoproteins in Plasma of Breast Cancer Patients. *Anal Chem*. 2008; 80:5286–5292. [PubMed: 18558770]
- Cholleti SR, Agravat S, Morris T, Saltz JH, Song X, Cummings RD, et al. Automated motif discovery from glycan array data. *Omics : a journal of integrative biology*. 2012; 16(10):497–512. [PubMed: 22877213]
- Collins BE, Paulson JC. Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Curr Opin Chem Biol*. 2004; 8(6):617–625. [PubMed: 15556405]
- Conze T, Carvalho AS, Landegren U, Almeida R, Reis CA, David L, et al. MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. *Glycobiology*. 2010; 20(2):199–206. [PubMed: 19815850]

- Culf AS, Cuperlovic-Culf M, Ouellette RJ. Carbohydrate microarrays: survey of fabrication techniques. *Omics*. 2006; 10(3):289–310. [PubMed: 17069509]
- Danussi C, Coslovi A, Campa C, Mucignat MT, Spessotto P, Uggeri F, et al. A newly generated functional antibody identifies Tn antigen as a novel determinant in the cancer cell-lymphatic endothelium interaction. *Glycobiology*. 2009; 19(10):1056–1067. [PubMed: 19528665]
- Drickamer K, Taylor ME. Glycan arrays for functional glycomics. *Genome Biol*. 2002; 3(12):REVIEWS1034. [PubMed: 12537579]
- Dwek MV, Jenks A, Leatham AJ. A sensitive assay to measure biomarker glycosylation demonstrates increased fucosylation of prostate specific antigen (PSA) in patients with prostate cancer compared with benign prostatic hyperplasia. *Clin Chim Acta*. 2010; 411(23-24):1935–1939. [PubMed: 20708609]
- Ebe Y, Kuno A, Uchiyama N, Koseki-Kuno S, Yamada M, Sato T, et al. Application of lectin microarray to crude samples: differential glycan profiling of lec mutants. *J Biochem*. 2006; 139(3):323–327. [PubMed: 16567396]
- Fermino ML, Polli CD, Toledo KA, Liu FT, Hsu DK, Roque-Barreira MC, et al. LPS-induced galectin-3 oligomerization results in enhancement of neutrophil activation. *PLoS ONE*. 2011; 6(10):e26004. [PubMed: 22031821]
- Forrester S, Kuick R, Hung KE, Kucherlapati R, Haab BB. Low-volume, high-throughput sandwich immunoassays for profiling plasma proteins in mice: identification of early-stage systemic inflammation in a mouse model of intestinal cancer. *Molecular Oncology*. 2007; 1:216–225. [PubMed: 19305640]
- Fuster MM, Esko JD. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer*. 2005; 5(7):526–542. [PubMed: 16069816]
- Gao C, Liu Y, Zhang H, Zhang Y, Fukuda MN, Palma AS, et al. Carbohydrate sequence of the prostate cancer-associated antigen F77 assigned by a mucin O-glycome designer array. *J Biol Chem*. 2014 epub ahead of print.
- Godula K, Bertozzi CR. Density variant glycan microarray for evaluating cross-linking of mucin-like glycoconjugates by lectins. *J Am Chem Soc*. 2012; 134(38):15732–15742. [PubMed: 22967056]
- Goonetilleke KS, Siriwardena AK. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur J Surg Oncol*. 2007; 33(3):266–270. [PubMed: 17097848]
- Grant OC, Smith HM, Firsova D, Fadda E, Woods RJ. Presentation, presentation, presentation! Molecular-level insight into linker effects on glycan array screening data. *Glycobiology*. 2014; 24(1):17–25. [PubMed: 24056723]
- Haab BB. Antibody-lectin sandwich arrays for biomarker and glycobiology studies. *Expert Rev Proteomics*. 2010; 7(1):9–11. [PubMed: 20121469]
- Haab BB, Partyka K, Cao Z. Using antibody arrays to measure protein abundance and glycosylation: considerations for optimal performance. *Current Protocols in Protein Science*. 2013; 73:27.26.21–27.26.16.
- Herlyn M, Sears HF, Steplewski Z, Koprowski H. Monoclonal antibody detection of a circulating tumor-associated antigen. I. Presence of antigen in sera of patients with colorectal, gastric, and pancreatic carcinoma. *J Clin Immunol*. 1982; 2(2):135–140. [PubMed: 7068815]
- Herlyn M, Steplewski Z, Herlyn D, Koprowski H. Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proc Natl Acad Sci U S A*. 1979; 76(3):1438–1442. [PubMed: 286328]
- Hirabayashi J. Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconj J*. 2004; 21(1-2):35–40. [PubMed: 15467396]
- Hirakawa J, Tsuboi K, Sato K, Kobayashi M, Watanabe S, Takakura A, et al. Novel anti-carbohydrate antibodies reveal the cooperative function of sulfated N- and O-glycans in lymphocyte homing. *J Biol Chem*. 2010; 285(52):40864–40878. [PubMed: 20929857]
- Hirvonen T, Suila H, Tiitinen S, Natunen S, Laukkanen ML, Kotovuori A, et al. Production of a recombinant antibody specific for i blood group antigen, a mesenchymal stem cell marker. *Biores Open Access*. 2013; 2(5):336–345. [PubMed: 24083089]

- Horan N, Yan L, Isobe H, Whitesides GM, Kahne D. Nonstatistical binding of a protein to clustered carbohydrates. *Proc Natl Acad Sci U S A*. 1999; 96(21):11782–11786. [PubMed: 10518527]
- Hsu KL, Pilobello KT, Mahal LK. Analyzing the dynamic bacterial glycome with a lectin microarray approach. *Nat Chem Biol*. 2006; 2(3):153–157. [PubMed: 16462751]
- Imamura K, Takeuchi H, Yabe R, Tateno H, Hirabayashi J. Engineering of the glycan-binding specificity of *Agrocybe cylindracea* galectin towards alpha(2,3)-linked sialic acid by saturation mutagenesis. *J Biochem*. 2011; 150(5):545–552. [PubMed: 21813503]
- Ito H, Kuno A, Sawaki H, Sogabe M, Ozaki H, Tanaka Y, et al. Strategy for Glycoproteomics: Identification of Glyco-Alteration Using Multiple Glycan Profiling Tools (dagger). *J Proteome Res*. 2009; 8(3):1358–1367. [PubMed: 19178301]
- Itzkowitz S, Kjeldsen T, Frier A, Hakomori S, Yang US, Kim YS. Expression of Tn, sialosyl Tn, and T antigens in human pancreas. *Gastroenterology*. 1991; 100(6):1691–1700. [PubMed: 1850375]
- Itzkowitz SH, Yuan M, Ferrell LD, Ratcliffe RM, Chung YS, Satake K, et al. Cancer-associated alterations of blood group antigen expression in the human pancreas. *J Natl Cancer Inst*. 1987; 79(3):425–434. [PubMed: 2442445]
- Itzkowitz SH, Yuan M, Fukushi Y, Lee H, Shi ZR, Zurawski V Jr. et al. Immunohistochemical comparison of Lea, monosialosyl Lea (CA 19-9), and disialosyl Lea antigens in human colorectal and pancreatic tissues. *Cancer Res*. 1988; 48(13):3834–3842. [PubMed: 3288336]
- Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol*. 2003; 21(6):667–672. [PubMed: 12754521]
- Kannagi R, Hakomori S. A guide to monoclonal antibodies directed to glycotopes. *Adv Exp Med Biol*. 2001; 491:587–630. [PubMed: 14533823]
- Karasawa F, Shiota A, Goso Y, Kobayashi M, Sato Y, Masumoto J, et al. Essential role of gastric gland mucin in preventing gastric cancer in mice. *J Clin Invest*. 2012; 122(3):923–934. [PubMed: 22307328]
- Kawa S, Tokoo M, Oguchi H, Furuta S, Homma T, Hasegawa Y, et al. Epitope analysis of SPan-1 and DUPAN-2 using synthesized glycoconjugates sialyllact-N-fucopentaose II and sialyllact-N-tetraose. *Pancreas*. 1994; 9(6):692–697. [PubMed: 7531332]
- Kelly LS, Kozak M, Walker T, Pierce M, Puett D. Lectin immunoassays using antibody fragments to detect glycoforms of human chorionic gonadotropin secreted by choriocarcinoma cells. *Anal Biochem*. 2005; 338(2):253–262. [PubMed: 15745745]
- Kim GE, Bae HI, Park HU, Kuan SF, Crawley SC, Ho JJ, et al. Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas. *Gastroenterology*. 2002; 123(4):1052–1060. [PubMed: 12360467]
- Kim YS, Itzkowitz SH, Yuan M, Chung Y, Satake K, Umeyama K, et al. Lex and Ley antigen expression in human pancreatic cancer. *Cancer Res*. 1988; 48(2):475–482. [PubMed: 3335015]
- Kjeldsen T, Clausen H, Hirohashi S, Ogawa T, Iijima H, Hakomori S. Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2----6 alpha-N-acetylgalactosaminyl (sialosyl-Tn) epitope. *Cancer Res*. 1988; 48(8):2214–2220. [PubMed: 2450649]
- Kletter D, Cao Z, Bern M, Haab B. Determining Lectin Specificity from Glycan Array Data using Motif Segregation and GlycoSearch Software. *Current Protocols in Chemical Biology*. 2013; 5:1–13. [PubMed: 23788322]
- Kletter D, Singh S, Bern M, Haab BB. Global comparisons of lectin glycan interactions using a database of analyzed glycan array data. *Molecular & Cellular Proteomics*. 2013; 12(4):1026–1035. [PubMed: 23399549]
- Koprowski H, Stepkowski Z, Mitchell K, Herlyn M, Herlyn D, Fuhrer P. Colorectal carcinoma antigens detected by hybridoma antibodies. *Somatic Cell Genet*. 1979; 5(6):957–971. [PubMed: 94699]
- Korekane H, Hasegawa T, Matsumoto A, Kinoshita N, Miyoshi E, Taniguchi N. Development of an antibody-lectin enzyme immunoassay for fucosylated alpha-fetoprotein. *Biochim Biophys Acta*. 2012; 1820(9):1405–1411. [PubMed: 22233759]

- Kumada Y, Ohigashi Y, Emori Y, Imamura K, Omura Y, Kishimoto M. Improved lectin ELISA for glycosylation analysis of biomarkers using PS-tag-fused single-chain Fv. *J Immunol Methods*. 2012; 385(1-2):15–22. [PubMed: 22884622]
- Kuno A, Kato Y, Matsuda A, Kaneko MK, Ito H, Amano K, et al. Focused differential glycan analysis with the platform antibody-assisted lectin profiling for glycan-related biomarker verification. *Mol Cell Proteomics*. 2009; 8(1):99–108. [PubMed: 18697734]
- Kuno A, Uchiyama N, Koseki-Kuno S, Ebe Y, Takashima S, Yamada M, et al. Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat Methods*. 2005; 2(11):851–856. [PubMed: 16278656]
- Lee K, Zhong X, Gu S, Krudel AM, Dorner MB, Perry K, et al. Molecular basis for disruption of E-cadherin adhesion by botulinum neurotoxin A complex. *Science*. 2014; 344(6190):1405–1410. [PubMed: 24948737]
- Lepur A, Salomonsson E, Nilsson UJ, Leffler H. Ligand induced galectin-3 protein self-association. *J Biol Chem*. 2012; 287(26):21751–21756. [PubMed: 22549776]
- Leteux C, Stoll MS, Childs RA, Chai W, Vorozhaikina M, Feizi T. Influence of oligosaccharide presentation on the interactions of carbohydrate sequence-specific antibodies and the selectins. Observations with biotinylated oligosaccharides. *J Immunol Methods*. 1999; 227(1-2):109–119. [PubMed: 10485259]
- Li C, Zolotarevsky E, Thompson I, Anderson MA, Simeone DM, Casper JM, et al. A multiplexed bead assay for profiling glycosylation patterns on serum protein biomarkers of pancreatic cancer. *Electrophoresis*. 2011; 32(15):2028–2035. [PubMed: 21732554]
- Li D, Chiu H, Chen J, Zhang H, Chan DW. Integrated analyses of proteins and their glycans in a magnetic bead-based multiplex assay format. *Clin Chem*. 2013; 59(1):315–324. [PubMed: 23099556]
- Li D, Chiu H, Zhang H, Chan DW. Analysis of serum protein glycosylation by a differential lectin immunosorbant assay (dLISA). *Clin Proteomics*. 2013; 10(1):12. [PubMed: 24015777]
- Li H, Mo KF, Wang Q, Stover CK, DiGiandomenico A, Boons GJ. Epitope mapping of monoclonal antibodies using synthetic oligosaccharides uncovers novel aspects of immune recognition of the Psl exopolysaccharide of *Pseudomonas aeruginosa*. *Chemistry*. 2013; 19(51):17425–17431. [PubMed: 24248772]
- Li Y, Tao SC, Bova GS, Liu AY, Chan DW, Zhu H, et al. Detection and verification of glycosylation patterns of glycoproteins from clinical specimens using lectin microarrays and lectin-based immunosorbent assays. *Anal Chem*. 2011; 83(22):8509–8516. [PubMed: 21975078]
- Liu Y, Palma AS, Feizi T. Carbohydrate microarrays: key developments in glycobiology. *Biol Chem*. 2009; 390(7):647–656. [PubMed: 19426131]
- Magnani JL, Brockhaus M, Smith DF, Ginsburg V, Blaszczyk M, Mitchell KF, et al. A monosialoganglioside is a monoclonal antibody-defined antigen of colon carcinoma. *Science*. 1981; 212(4490):55–56. [PubMed: 7209516]
- Magnani JL, Nilsson B, Brockhaus M, Zopf D, Steplewski Z, Koprowski H, et al. A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-fucopentaose II. *J Biol Chem*. 1982; 257(23):14365–14369. [PubMed: 7142214]
- Magnani JL, Steplewski Z, Koprowski H, Ginsburg V. Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. *Cancer Res*. 1983; 43(11):5489–5492. [PubMed: 6193872]
- Markiv A, Peiris D, Curley GP, Odell M, Dwek MV. Identification, cloning, and characterization of two N-acetylgalactosamine-binding lectins from the albumen gland of *Helix pomatia*. *J Biol Chem*. 2011; 286(23):20260–20266. [PubMed: 21372134]
- Maupin KA, Liden D, Haab BB. The fine specificity of mannose-binding and galactose-binding lectins revealed using outlier-motif analysis of glycan array data. *Glycobiology*. 2011; 22(1):160–169. [PubMed: 21875884]
- McCarter C, Kletter D, Tang H, Partyka K, Ma Y, Singh S, et al. Prediction of glycan motifs using quantitative analysis of multi-lectin binding: Motifs on MUC1 produced by cultured pancreatic cancer cells. *Proteomics Clinical Applications*. 2013; 7:632–641. [PubMed: 23956151]

- Meany DL, Zhang Z, Sokoll LJ, Zhang H, Chan DW. Glycoproteomics for prostate cancer detection: changes in serum PSA glycosylation patterns. *J Proteome Res.* 2009; 8(2):613–619. [PubMed: 19035787]
- Metzgar RS, Gaillard MT, Levine SJ, Tuck FL, Bossen EH, Borowitz MJ. Antigens of human pancreatic adenocarcinoma cells defined by murine monoclonal antibodies. *Cancer Res.* 1982; 42(2):601–608. [PubMed: 7034925]
- Nonaka M, Fukuda MN, Gao C, Li Z, Zhang H, Greene MI, et al. Determination of Carbohydrate Structure Recognized by Prostate-specific F77 Monoclonal Antibody through Expression Analysis of Glycosyltransferase Genes. *J Biol Chem.* 2014; 289(23):16478–16486. [PubMed: 24753248]
- Nycholat CM, McBride R, Ekiert DC, Xu R, Rangarajan J, Peng W, et al. Recognition of Sialylated Poly-N-acetylactosamine Chains on N- and O-Linked Glycans by Human and Avian Influenza A Virus Hemagglutinins. *Angew Chem Int Ed Engl.* 2012; 51:4860–4863. [PubMed: 22505324]
- Okuyama N, Ide Y, Nakano M, Nakagawa T, Yamanaka K, Moriaki K, et al. Fucosylated haptoglobin is a novel marker for pancreatic cancer: A detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int J Cancer.* 2006; 118(11):2803–2808. [PubMed: 16385567]
- Olausson J, Astrom E, Jonsson BH, Tibell LA, Pahlsson P. Production and characterization of a monomeric form and a single-site form of Aleuria aurantia lectin. *Glycobiology.* 2011; 21(1):34–44. [PubMed: 20798114]
- Osako M, Yonezawa S, Siddiki B, Huang J, Ho JJ, Kim YS, et al. Immunohistochemical study of mucin carbohydrates and core proteins in human pancreatic tumors. *Cancer.* 1993; 71(7):2191–2199. [PubMed: 8384065]
- Oyelaran O, Li Q, Farnsworth DW, Gildersleeve JC. Microarrays with Varying Carbohydrate Density Reveal Distinct Subpopulations of Serum Antibodies. *J Proteome Res.* 2009; 8:3529–3538. [PubMed: 19366269]
- Pace KE, Lee C, Stewart PL, Baum LG. Restricted receptor segregation into membrane microdomains occurs on human T cells during apoptosis induced by galectin-1. *J Immunol.* 1999; 163(7):3801–3811. [PubMed: 10490978]
- Padler-Karavani V, Song X, Yu H, Hurtado-Ziola N, Huang S, Muthana S, et al. Cross-comparison of protein recognition of sialic acid diversity on two novel sialoglycan microarrays. *J Biol Chem.* 2012; 287(27):22593–22608. [PubMed: 22549775]
- Pan S, Chen R, Aebersold R, Brentnall TA. Mass spectrometry based glycoproteomics--from a proteomics perspective. *Mol Cell Proteomics.* 2011; 10(1):R110 003251. [PubMed: 20736408]
- Partyka K, Maupin KA, Brand RE, Haab BB. Diverse monoclonal antibodies against the CA 19-9 antigen show variation in binding specificity with consequences for clinical interpretation. *Proteomics.* 2012; 12:2212–2220. [PubMed: 22623153]
- Pilobello KT, Krishnamoorthy L, Slawek D, Mahal LK. Development of a lectin microarray for the rapid analysis of protein glycopatterns. *Chembiochem.* 2005; 6(6):985–989. [PubMed: 15798991]
- Pilobello KT, Slawek DE, Mahal LK. A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome. *Proc Natl Acad Sci U S A.* 2007; 104(28):11534–11539. [PubMed: 17606908]
- Pinto R, Carvalho AS, Conze T, Magalhaes A, Picco G, Burchell JM, et al. Identification of new cancer biomarkers based on aberrant mucin glycoforms by in situ proximity ligation. *J Cell Mol Med.* 2012; 16(7):1474–1484. [PubMed: 21883895]
- Porter A, Yue T, Heeringa L, Day S, Suh E, Haab BB. A motif-based analysis of glycan array data to determine the specificities of glycan-binding proteins. *Glycobiology.* 2010; 20(3):369–380. [PubMed: 19946132]
- Rho JH, Mead JR, Wright WS, Brenner DE, Stave JW, Gildersleeve JC, et al. Discovery of sialyl Lewis A and Lewis X modified protein cancer biomarkers using high density antibody arrays. *J Proteomics.* 2014; 96:291–299. [PubMed: 24185138]
- Rillahan CD, Paulson JC. Glycan microarrays for decoding the glycome. *Annu Rev Biochem.* 2011; 80:797–823. [PubMed: 21469953]

- Romano PR, Mackay A, Vong M, DeSa J, Lamontagne A, Comunale MA, et al. Development of recombinant *Aleuria aurantia* lectins with altered binding specificities to fucosylated glycans. *Biochem Biophys Res Commun*. 2011; 414(1):84–89. [PubMed: 21945439]
- Saba J, Dutta S, Hemenway E, Viner R. Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *International journal of proteomics*. 2012; 2012:560391. [PubMed: 22701174]
- Salomonsson E, Carlsson MC, Osla V, Hendus-Altenburger R, Kahl-Knutson B, Oberg CT, et al. Mutational tuning of galectin-3 specificity and biological function. *J Biol Chem*. 2010; 285(45):35079–35091. [PubMed: 20807768]
- Satomura Y, Sawabu N, Takemori Y, Ohta H, Watanabe H, Okai T, et al. Expression of various sialylated carbohydrate antigens in malignant and nonmalignant pancreatic tissues. *Pancreas*. 1991; 6(4):448–458. [PubMed: 1678888]
- Schuessler MH, Pintado S, Welt S, Real FX, Xu M, Melamed MR, et al. Blood group and blood-group-related antigens in normal pancreas and pancreas cancer: enhanced expression of precursor type 1, Tn and sialyl-Tn in pancreas cancer. *Int J Cancer*. 1991; 47(2):180–187. [PubMed: 1988363]
- Sharon N. Lectins: carbohydrate-specific reagents and biological recognition molecules. *J Biol Chem*. 2007; 282(5):2753–2764. [PubMed: 17145746]
- Shida K, Korekane H, Misonou Y, Noura S, Ohue M, Takahashi H, et al. Novel ganglioside found in adenocarcinoma cells of Lewis-negative patients. *Glycobiology*. 2010; 20(12):1594–1606. [PubMed: 20667987]
- Shimizu K, Katoh H, Yamashita F, Tanaka M, Tanikawa K, Taketa K, et al. Comparison of carbohydrate structures of serum alpha-fetoprotein by sequential glycosidase digestion and lectin affinity electrophoresis. *Clin Chim Acta*. 1996; 254(1):23–40. [PubMed: 8894307]
- Shimizu M, Saitoh Y, Ohyanagi H, Itoh H. Immunohistochemical staining of pancreatic cancer with CA19-9, KM01, unabsorbed CEA, and absorbed CEA. A comparison with normal pancreas and chronic pancreatitis. *Arch Pathol Lab Med*. 1990; 114(2):195–200. [PubMed: 1689141]
- Smith DF, Cummings RD. Application of microarrays to deciphering the structure and function of the human glycome. *Molecular & cellular proteomics*. 2013; 12(4):902–912. [PubMed: 23412570]
- Song X, Lasanajak Y, Xia B, Heimbarg-Molinari J, Rhea JM, Ju H, et al. Shotgun glycomics: a microarray strategy for functional glycomics. *Nat Methods*. 2011; 8(1):85–90. [PubMed: 21131969]
- Song X, Yu H, Chen X, Lasanajak Y, Tappert MM, Air GM, et al. A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses. *J Biol Chem*. 2011; 286:31610–31622. [PubMed: 21757734]
- Stevens J, Blixt O, Paulson JC, Wilson IA. Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat Rev Microbiol*. 2006; 4(11):857–864. [PubMed: 17013397]
- Suzuki Y, Ichihara T, Nakao A, Sakamoto J, Takagi H, Nagura H. High serum levels of DUPAN2 antigen and CA19-9 in pancreatic cancer: correlation with immunocytochemical localization of antigens in cancer cells. *Hepatogastroenterology*. 1988; 35(3):128–135. [PubMed: 3042577]
- Takasaki H, Uchida E, Tempero MA, Burnett DA, Metzgar RS, Pour PM. Correlative study on expression of CA 19-9 and DU-PAN-2 in tumor tissue and in serum of pancreatic cancer patients. *Cancer Res*. 1988; 48(6):1435–1438. [PubMed: 3162196]
- Tao SC, Li Y, Zhou J, Qian J, Schnaar RL, Zhang Y, et al. Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology*. 2008; 18(10):761–769. [PubMed: 18625848]
- Tateno H, Uchiyama N, Kuno A, Togayachi A, Sato T, Narimatsu H, et al. A novel strategy for mammalian cell surface glycome profiling using lectin microarray. *Glycobiology*. 2007; 17(10):1138–1146. [PubMed: 17693441]
- Terada T, Nakanuma Y. Expression of mucin carbohydrate antigens (T, Tn and sialyl Tn) and MUC-1 gene product in intraductal papillary-mucinous neoplasm of the pancreas. *Am J Clin Pathol*. 1996; 105(5):613–620. [PubMed: 8623771]

- Thompson S, Cantwell BM, Cornell C, Turner GA. Abnormally-fucosylated haptoglobin: a cancer marker for tumour burden but not gross liver metastasis. *Br J Cancer*. 1991; 64(2):386–390. [PubMed: 1892771]
- Thompson S, Guthrie D, Turner GA. Fucosylated forms of alpha-1-antitrypsin that predict unresponsiveness to chemotherapy in ovarian cancer. *Br J Cancer*. 1988; 58(5):589–593. [PubMed: 3265332]
- Thompson S, Stappenbeck R, Turner GA. A multiwell lectin-binding assay using lotus tetragonolobus for measuring different glycosylated forms of haptoglobin. *Clin Chim Acta*. 1989; 180(3):277–284. [PubMed: 2743580]
- van Dijk W, Havenaar EC, Brinkman-van der Linden EC. Alpha 1-acid glycoprotein (orosomucoid): pathophysiological changes in glycosylation in relation to its function. *Glycoconj J*. 1995; 12(3):227–233. [PubMed: 7496136]
- Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Stanley, P.; Bertozzi, CR., et al. *Essentials of Glycobiology*. 2nd ed.. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 2009.
- Wang L, Cummings RD, Smith DF, Huflejt M, Campbell CT, Gildersleeve JC, et al. Cross-platform comparison of glycan microarray formats. *Glycobiology*. 2014; 24(6):507–517. [PubMed: 24658466]
- Wang X, Katayama A, Wang Y, Yu L, Favoino E, Sakakura K, et al. Functional characterization of an scFv-Fc antibody that immunotherapeutically targets the common cancer cell surface proteoglycan CSPG4. *Cancer Res*. 2011; 71(24):7410–7422. [PubMed: 22021902]
- Wang Z, Chinoy ZS, Ambre SG, Peng W, McBride R, de Vries RP, et al. A general strategy for the chemoenzymatic synthesis of asymmetrically branched N-glycans. *Science*. 2013; 341(6144):379–383. [PubMed: 23888036]
- Weibrecht I, Leuchowius KJ, Clausson CM, Conze T, Jarvius M, Howell WM, et al. Proximity ligation assays: a recent addition to the proteomics toolbox. *Expert Rev Proteomics*. 2010; 7(3):401–409. [PubMed: 20536310]
- Wu JT. Serum alpha-fetoprotein and its lectin reactivity in liver diseases: a review. *Ann Clin Lab Sci*. 1990; 20(2):98–105. [PubMed: 1691611]
- Wu YM, Nowack DD, Omenn GS, Haab BB. Mucin glycosylation is altered by pro-inflammatory signaling in pancreatic-cancer cells. *J Proteome Res*. 2009; 8(4):1876–1886. [PubMed: 19714813]
- Xuan P, Zhang Y, Tzeng TR, Wan XF, Luo F. A quantitative structure-activity relationship (QSAR) study on glycan array data to determine the specificities of glycan-binding proteins. *Glycobiology*. 2011; 22(4):552–560. [PubMed: 22156918]
- Yabe R, Itakura Y, Nakamura-Tsuruta S, Iwaki J, Kuno A, Hirabayashi J. Engineering a versatile tandem repeat-type alpha2-6sialic acid-binding lectin. *Biochem Biophys Res Commun*. 2009; 384(2):204–209. [PubMed: 19393618]
- Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J Chromatogr A*. 2004; 1053(1-2):79–88. [PubMed: 15543974]
- Yoneyama T, Ohyama C, Hatakeyama S, Narita S, Habuchi T, Koie T, et al. Measurement of aberrant glycosylation of prostate specific antigen can improve specificity in early detection of prostate cancer. *Biochem Biophys Res Commun*. 2014; 448(4):390–396. [PubMed: 24814705]
- Yu Y, Lasanajak Y, Song X, Hu L, Ramani S, Mickum ML, et al. Human milk contains novel glycans that are potential decoy receptors for neonatal rotaviruses. *Mol Cell Proteomics*. 2014 epub ahead of print.
- Yu Y, Mishra S, Song X, Lasanajak Y, Bradley KC, Tappert MM, et al. Functional glycomic analysis of human milk glycans reveals the presence of virus receptors and embryonic stem cell biomarkers. *J Biol Chem*. 2012; 287(53):44784–44799. [PubMed: 23115247]
- Yue T, Goldstein IJ, Hollingsworth MA, Kaul K, Brand RE, Haab BB. The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol Cell Proteomics*. 2009; 8(7):1697–1707. [PubMed: 19377061]

- Yue T, Partyka K, Maupin KA, Hurley M, Andrews P, Kaul K, et al. Identification of blood-protein carriers of the CA 19-9 antigen and characterization of prevalence in pancreatic diseases. *Proteomics*. 2011; 11(18):3665–3674. [PubMed: 21751362]
- Zhang G, Zhang H, Wang Q, Lal P, Carroll AM, de la Llera-Moya M, et al. Suppression of human prostate tumor growth by a unique prostate-specific monoclonal antibody F77 targeting a glycolipid marker. *Proc Natl Acad Sci U S A*. 2010; 107(2):732–737. [PubMed: 20080743]
- Zhang Y, Li Q, Rodriguez LG, Gildersleeve JC. An array-based method to identify multivalent inhibitors. *J Am Chem Soc*. 2010; 132(28):9653–9662. [PubMed: 20583754]
- Zheng T, Peelen D, Smith LM. Lectin arrays for profiling cell surface carbohydrate expression. *J Am Chem Soc*. 2005; 127(28):9982–9983. [PubMed: 16011345]

A. Glycan Array Data

55786	GlcNAc1-3Galβ-Sp8
51494	GlcNAc1-3(Fuca1-2)Galβ-Sp8
48880	GlcNAc1-4GlcNAcβ-Sp0
47230	a-GalNAc-Sp8
47136	β-GalNAc-Sp8
47131	GlcNAc1-4GlcNAcβ-Sp8
45707	GlcNAc1-3Galα1-4Galβ1-4GlcNAcβ-Sp0
12347	Neu5Aca2-3(GlcNAcβ1-4)Galβ1-4Glcβ-Sp0
4229	GlcNAc1-3GlcNAcβ-Sp8
2199	Neu5Aca2-8Neu5Aca2-3(GlcNAcβ1-4)Galβ1-4Glcβ-Sp0
2139	GlcNAc1-4(Fuca1-2)Galβ1-4GlcNAcβ-Sp8
1839	Neu5Aca2-3(GlcNAcβ1-4)Galβ1-4GlcNAcβ-Sp8
1175	Neu5Aca2-3(GlcNAcβ1-4)Galβ1-4GlcNAcβ-Sp0
1092	[3OS03]Galβ1-4[6OS03]GlcNAcβ-Sp8
672	(GlcNAcβ1-4)β-Sp8
570	GlcNAc1-4GlcNAcβ1-4GlcNAcβ-Sp8
410	Manα1-6(Manα1-3)Manα1-6(Manα2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Asn
377	GlcNAc1-3(Fuca1-2)Galβ1-4GlcNAcβ-Sp0
372	GlcNAc1-3Galβ1-4GlcNAcβ-Sp8
362	GlcNAc1-3GlcNAc-Sp8
...	...

B. Motif Classification of Glycans

C. Motif Segregation Method

Figure 1. The Motif Segregation algorithm for analyzing glycan array data

A) Glycan array data consists of measurements of the binding of a GBP to each distinct glycan. The binding usually is measured by fluorescence intensity at each glycan on the array (left panel). The example shows the quantified fluorescence at the top glycans bound by the GBP from *Vicia villosa* (VVL) (right panel). B) For each motif in a predefined set of motifs, the software identifies the glycans that contain the motif. A value of 1 indicates the motif is present, and a 0 indicates absent. C) For each motif, the glycans are separated into two groups: those containing the motif and those not containing it. The fluorescence intensities are compared between the two groups by a statistical test, and a score is given to

indicate the significance of the difference (left panel). The example shows the groupings and calculations for three motifs.

A. Building a Database of Analyzed Glycan Array Data

B. Output from a Query of Binders to Sialyl Lewis A

Motifs Similar to Sialyl Lewis A

Figure 2. Motif prediction using information from multiple GBPs

A) Antibody-lectin sandwich arrays are useful for obtaining measurements of glycans on proteins captured by the antibodies. Arrays can be repeatedly incubated with a sample, and each replicate array is probed with a different lectin. For each motif and each lectin measurement, the signal from the lectin binding is multiplied by the motif score for the given lectin. In the example, $MS_{1,1}$ is the motif score for motif 1 and lectin 1, $MS_{1,2}$ is the motif score for motif 1 and lectin 2, $MS_{2,1}$ is the motif score for motif 2 and lectin 1, and so forth. I_1 is the intensity from lectin 1, I_2 is the intensity from lectin 2, and so forth. For each

motif, the products are summed over all the lectins to arrive at a motif prediction score. B) We tested and validated the method using glycan array data. We calculated motif prediction scores for the motif “terminal, alpha-GlcNAc” for every glycan on the array using data from the lectins GSL-2 (top left) and HAA (top right). The graphs show the scores for the glycans without the motif (n = 602) and the glycans with the motif (n = 9). The size of each circle indicates the number of glycans with values in the region. Both lectins had high scores for some glycans that did not have terminal, alpha-GlcNAc. When we calculated the motif prediction scores using both lectins (bottom, calculated by adding the scores from the two lectins for each glycan), all the glycans with the motif had higher scores than all the glycans without the motif. This result shows that integrating data from more than one lectin can give more accurate prediction.

A. Calculating a Motif Prediction Score

B. Validation Using Glycan Array Data

Figure 3. Building a querying a database of analyzed glycan array data
A) The GlycanBinder database contains the raw data, metadata, motifs, and analyzed data from thousands of glycan array experiments, with all information linked in a relational format, enabling all types of searches and analysis. B) We queried the database for binders with the highest motif scores for the motif sialyl Lewis A. For 14 of the top binders, we examined the motif scores for sialyl Lewis A and for related motifs. The motifs are indicated by the column labels, the binders by the row labels, and the color of each square indicates

the motif score according to the scale on the color bar. All binders have a good score for sialyl Lewis A, but they show diverse binding to other, related motifs.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript