HHS Public Access

Author manuscript

ACS Chem Biol. Author manuscript; available in PMC 2022 June 15.

Published in final edited form as:

ACS Chem Biol. 2021 October 15; 16(10): 1795–1813. doi:10.1021/acschembio.0c00880.

Strategies and Tactics for the Development of Selective Glycan-Binding Proteins

Elizabeth M. Ward[1,3],

Megan E. Kizer[1,2],

Barbara Imperiali[1,2,*]

[1] Department of Biology, Massachusetts Institute of Technology, 31 Ames St, Cambridge, MA 02142, USA

[2] Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA

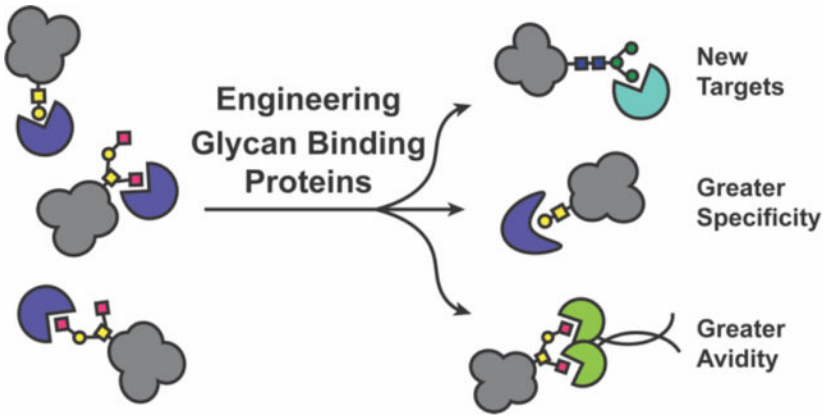
[3] Microbiology Graduate Program, Massachusetts Institute of Technology, 31 Ames St, Cambridge, MA 02142, USA

Abstract

The influences of glycans impinge on all biological processes, disease states and pathogenic interactions. Glycan-binding proteins (GBPs), such as lectins, are decisive tools for interrogating glycan structure and function because of their ease of use and ability to selectively bind defined carbohydrate epitopes and glycosidic linkages. GBP reagents are prominent tools for basic research, clinical diagnostics, therapeutics, and biotechnological applications. However, the study of glycans is hindered by the lack of specific and selective protein reagents to cover the massive diversity of carbohydrate structures existing in nature. In addition, existing GBP reagents often suffer from low affinity or broad specificity, complicating data interpretation. There have been numerous efforts to expand the GBP toolkit beyond those identified from natural sources through protein engineering, in order to improve the properties of existing GBPs or to engineer novel specificities and potential applications. This review details the current scope of proteins that bind carbohydrates and the engineering methods that have been applied to enhance affinity, selectivity, and specificity of binders.

Graphical Abstract

* Corresponding author, imper@mit.edu.



1. INTRODUCTION

1.1 Glycan-Binding Protein Overview

Carbohydrates are ubiquitous molecules requisite for many biological processes, such as mediating interactions between cells, acting as regulatory elements in cellular signaling, and mediating membrane organization.[1] Glycans can exist as glycopolymers, but are often found as glycoconjugates appended to proteins and lipids, influencing the structure and function of these biomolecules. The importance of glycans in biology has long been known, but significant challenges in their manipulation and analysis have hampered their study from being incorporated into general biological research.

The study of glycans is complicated by the amazing diversity of these structures, both at the monosaccharide level and at the glycoside-bond level. Unlike nucleic acids or amino acids, which are linearly polymerized by a single linkage type, glycans can form glycosidic linkages between any of several hydroxyl groups as either the α- or β-anomers. This also allows for branching structures. The diversity of individual monosaccharide building blocks also dwarfs that of nucleic acids and amino acids, with estimates of unique monosaccharides in bacteria being on the order of 800.[2, 3] Glycans are not template-encoded, and their synthesis depends on the sequential action of multiple glycosyltransferase enzymes, making complete structures impossible to predict based only on genetic information. Furthermore, the presence of highly related stereoisomers can confound detection and analysis.

Diverse glycan-recognition approaches have been the focus of many research groups, including mass spectrometry, nucleic acid aptamers, boronolectins, pyrrole receptors and oligomeric aromatic molecules. These methods each have their respective disadvantages, broadly including time-consuming enzymatic digestion, large sample requirements, expensive and specialized equipment, the need for highly trained personnel for operation or synthesis, and potential degradation during analysis. Such methods are outside the scope of this review and the authors direct the reader to other reviews that cover these methods.[4–8]

Carbohydrate-binding reagents are important tools for the study and detection of glycans. Unlike many of the techniques mentioned above, these reagents do not require specialized equipment and can be readily utilized by the wider biological research community. Ideally,

these reagents bind selectively to specific glycan epitopes, allowing for qualitative structure analysis without extensive sample preprocessing. Carbohydrate-binding reagents fall into several categories: glycan-binding proteins (GBPs),[9] nucleic acid aptamers,[7, 8] and small-molecule lectin mimetics.[6, 10, 11] GBPs are the most commonly used carbohydrate-binding reagent, with many commercially available.

There are a number of different types of proteins that recognize carbohydrates. These glycan-binding proteins can be categorized into three major groups: lectins, carbohydrate-binding modules (CBMs), and adaptive immune proteins (antibodies and variable lymphocyte receptors). GBPs are routinely utilized in many different ways (Figure 1). They enable a fundamental understanding of carbohydrate-protein interactions, and are used as tools to isolate or identify specific glycans or glycan-modified biomolecules.[12] In biotechnology, GBPs are important domains of biomass degrading enzymes, and are frequently used to purify or immobilize glycosylated targets.[13] Because characteristic glycans are found on cell surfaces, GBPs are also used in clinical settings for diagnostics, including histology, blood typing, and microorganism detection.[14–17] GBPs are important cancer diagnostics as well, as the aberrant glycosylation patterns found on malignant tissues can act as disease biomarkers.[17] Therapeutic application of GBPs is also an active area of research, with GBPs being explored for cellular targeting of therapeutic molecules.[18–20] Certain GBPs have also been identified with direct anti-microbial, anti-viral and anti-cancer activity.[21–24]

Although many GBPs have been identified and characterized, they fail to cover the immense diversity of glycans present in nature, and often lack the affinity and specificity needed to be useful as reagents. For this reason, GBP engineering has been a growing area of research paced to match major unmet needs. Engineering efforts on the GBP groups highlighted here have been employed to increase affinity and/or specificity for a carbohydrate target, decrease affinity for an off-target carbohydrate, develop novel specificity or develop protein scaffolds for glycan-binding engineering. In this review, the current state of protein binders that recognize carbohydrates and the engineering methods to develop superior GBPs will be presented. To understand the strategies utilized for GBP engineering, it is imperative to first discuss general protein engineering approaches, screening methods, and the molecular determinants for protein-carbohydrate interactions.

1.2 Engineering and Screening Methods

GBP engineering focuses on the modification of existing protein scaffolds to exhibit more desirable properties, such as increased specificity or higher affinity towards the carbohydrate target. This can be accomplished by rational design, directed evolution or a combination of the two. In brief, rational design begins with a protein scaffold of interest, which is manipulated in a low-throughput manner by targeting specific residues to change.[25] Directed evolution involves mutagenesis of the protein scaffold to generate a degenerate protein library, followed by iterative rounds of selection for the desired properties, amplification of the selected variants, and further mutagenesis (Figure 2A).[26, 27] Generating library diversity can be executed in two ways: site-directed mutagenesis and random mutagenesis. The former is a low-throughput approach where defined residues are changed to specific

or randomized amino acids. This usually requires some prerequisite knowledge of the protein structure or, even better, the protein-glycan interaction. The latter approach involves mutation of a portion or entirety of the encoding gene randomly. Since there is no bias from structural information, this method can reveal changes distal to the interaction surface that may have otherwise been overlooked.

Appropriate display and screening methods are paramount to obtaining a GBP with the desired properties. Large protein libraries (diversity ca. 10^7-10^{14}) are displayed using a method that links the phenotype of a selected protein to its genotype (Figure 2B). This underscores the ability to evolve a population of desired GBP variants *in vitro*. Surface-based methods compartmentalize the genetic material inside of a cell (mammalian,[28] bacterial,[29] or yeast[30]) or phage particle,[31] which encodes for the protein variant fused to a cell-surface protein. Conjugation-based methods have a physical linkage between the protein variant and the genetic material, and includes techniques such as ribosome display[32, 33] and mRNA display.[34] To effectively isolate the displayed proteins, high-throughput screening methods are routinely employed. Sorting-based methods include magnetic-activated cell sorting (MACS) and fluorescence-activated cell sorting (FACS), where a target glycan is conjugated to a magnetic bead or fluorophore, allowing for target GBPs to be isolated by the respective glycan-conjugated handle. These methods are desirable for initial screens of protein libraries, where a simple reduction in diversity (*e.g.* MACS) or detection and collection of small populations of binders (*e.g.* FACS) can be achieved. Other high-throughput methods, usually utilized later in the GBP engineering work flow, include plate-based screens such as ELISA with colorimetric or fluorescent readouts and carbohydrate microarrays (Figure 2C).[35, 36]

Microarray technologies have revolutionized the way in which GBP screening is performed.[37–39] They have been employed in many studies involving lectins, viral protein interactions, immunological investigations, anti-glycan antibodies, and vaccine development.[35, 40–44] Production involves the modification of synthetically or chemoenzymatically-prepared glycans (sequence-defined), and glycans isolated from natural sources (shotgun) with a reactive handle and subsequent immobilization on functionalized glass surfaces.[40, 41, 43, 44] The chemical combinations for immobilization are diverse and have been summarized in recent reviews.[40] Glycan immobilization is performed by printing micrometer-sized spots, reducing both carbohydrate and GBP sample requirements. Such a miniaturization into a microarray format not only reduces sample amount, but also allows for multiple glycan epitopes to be screened at once, ultimately minimizing the time and effort to obtain glycan-binding information.

The predominant glycan microarrays available include those developed by the Consortium for Functional Glycomics (CFG) and the National Center for Functional Glycomics (NCFG). The CFG has a mammalian glycan array (in its fifth generation) which has about 609 unique glycans as well as a pathogen array which contains over 300 glycans from various pathogens. The NCFG also has mannose-6-phosphate, modified sialic acid (Sia, N-acetylneuraminicacid/Neu5Ac), schistosome glycan, Tn antigen glycopeptide, and soybean agglutinin (SBA) glycan arrays for more targeted investigations. Another array developed in the Seeberger group utilized a combination of automated glycan assembly, solution-phase

glycan synthesis, chemoenzymatic synthesis, and biological isolation was employed to generate a library with over 300 mono- to eicosaccharides.[45] This array, called the Max Planck Society (MPS) glycan array, has a microbial focus, but also includes mammalian and plant glycan fragments and epitopes. Many shotgun microarrays are available, including the schistosome egg glycan and human milk oligosaccharide (HMO) arrays from the NCFG, as well as many lab-specific developed shotgun glycan arrays.[46, 47] Finally, a neoglycolipid (NGL)-based microarray system is available from the Glycosciences Laboratory (https://glycosciences.med.ic.ac.uk/glycanLibraryIndex.html) and contains about 800 glycan targets including N-glycans, O-glycans, glycolipids, glycosaminoglycans, blood group glycans, Lewis antigen glycans, neutral and sialylated glycans, as well as oligosaccharides from plant and microbial sources.[42]

Because microarrays are so useful for GBP analysis and development, improvements in microarray technology are of great importance and include advances in the chemical conjugation and the breadth of glycans sampled. One microarray development employs a chemoenzymatic approach to access N-linked glycosyl-asparagines that better mimic N-linked glycan epitopes.[48] Another conjugates glycans to polymers and peptides for clustered glycan presentation. Progress in multiplexing with Luminex® beads,[49] DNA-based arrays,[50, 51] and cell-based arrays have also been investigated.[52, 53]

A major limitation of the current array landscape is the inaccessibility of glycans that accurately represent the glycan diversity of both mammals and microbes. It has been estimated that there are 3,000 unique glycan species on glycoproteins and glycolipids in mammals,[54] meanwhile the bacterial glycome is even more diverse.[55] Yet many of these glycans are difficult to obtain in large quantities for biochemical assays. Ultimately, the current glycan array space does not begin to comprehensively cover carbohydrate complexity and diversity, and many biologically-isolated glycans are not sequence defined. Efforts to overcome these limitations are in progress. Despite these limitations, the microarray remains an incredibly powerful tool for glycan-protein interaction analysis and GBP screening.

1.3 Carbohydrate-Protein Interactions

Understanding the interaction between a protein and its carbohydrate target provides a great opportunity for site-directed mutagenesis in GBP engineering. Analysis of residues proximal to bound carbohydrates in structures deposited in the Protein Data Bank reveals several trends.[56] The first is a striking enrichment of the aromatic amino acids Trp, Tyr, and His in carbohydrate binding sites. This seems to contradict the highly hydrophilic nature of carbohydrates, but aromatic amino acids are one of the most important drivers of carbohydrate-protein interactions. The polarized C-H bonds of carbohydrates are able to form CH-π interactions with the electron rich π-systems of aromatic amino acids.[57] Trp in particular is highly abundant in carbohydrate-binding sites. This can be explained by the electronics of the aromatic system. The Trp indole system is more electron-rich than other aromatic amino acids, even when the additional surface area is factored in.[56] In addition to aromatic residues, the polar residues Asp, Asn, Gln, and Arg show some enrichment in carbohydrate-binding sites. Many glycans differ only in the positioning of

certain hydroxyl groups, therefore these polar residues are positioned to make direct contacts with these distinguishing hydroxyl groups. An example of this is shown in galactose- or mannose-binding lectins, which discriminate at the C4 hydroxyl group.[58] These selective interactions can also be promoted by a coordinated Ca^{2+} ion instead of an amino acid, as is the case for C-type lectins.[59] Aliphatic amino acids are particularly disfavored in carbohydrate-binding sites, indicating that the hydrophobic effect does not play a major role in binding.

Carbohydrate-protein interactions are often low affinity, with a single carbohydrate-binding site interacting with a monovalent ligand having K_D values in the μM-mM range. However, GBPs are often oligomeric or contain multiple carbohydrate-binding sites per protein and thus exploit multivalent interactions, thereby increasing the overall strength of the interaction.[60] This increased strength, or avidity, is presented as an apparent K_D and can be orders of magnitude stronger than a monovalent interaction. There also exists multivalency on the glycan side, either manifested as repeating units in one polysaccharide or as multiple proximal carbohydrates, for example on the dense glycan-covered cell surface. The gain in avidity from interaction of GBPs with multivalent carbohydrate ligands is known as the "cluster glycoside effect".[61] A summary of reports of GBPs with multivalent saccharide ligands shows that in all cases a multivalent ligand will show enhanced avidity, though the magnitude of this enhancement is highly variable.[61] As such, incorporating multivalency is frequently employed in GBP engineering.

2. LECTINS

Lectins are sugar-binding proteins without enzymatic activity. First discovered in plants, they have since been identified throughout the natural world in animals, fungi, bacteria, and viruses. There are several ways lectins are classified: by the natural source from which they are identified, by the types of carbohydrates they bind, by amino acid sequence, or by structure. When grouped by three-dimensional structures, lectins are categorized into 48 different families.[62] Many lectins are rich in β-structure, adopting folds such as β-sandwich, β-trefoil and β-propeller, although β-structure is not a necessity.[62] This breadth of structural features provides lectins with the ability to recognize a panoply of carbohydrate targets with varying monosaccharide units, types of linkages and number of branches. Many lectins have shallow and solvent-exposed binding sites, and as mentioned previously often have low affinity for their carbohydrate targets, with apparent K_D values in the high μM to mM range. Because of this, lectins are frequently oligomeric, thereby increasing avidity. However, such requisite oligomerization only complicates the *in vitro* production of this type of GBP.[63, 64] Lectins also suffer from broad specificity toward carbohydrate epitopes, making data interpretation difficult. Regardless, many lectins have been identified and their binding characteristics defined.[65]

The glycobiology field over the past 30 years has embraced natural lectin engineering to overcome the aforementioned limitations and maintain these GBPs as useful tools for glycobiological research. Increased specificities, higher-affinity interactions, and novel carbohydrate-recognition properties have been attained through lectin engineering efforts. Common methods used for lectin engineering are described and specific examples for each

are highlighted. This does not include the computer-assisted directed evolution of inactivated carbohydrate-processing enzymes, called Lectenz®.[66] For other lectin engineering reviews, we direct the reader to Hu et al. and Hirabayashi and Arai[67, 68]

2.1 Site-Directed Mutagenesis

Site-directed mutagenesis is the most frequently applied approach in lectin engineering. This method can be as simple as mutagenesis of a single residue to one or several other amino acids, or as complex as generating randomized mutations of many selected residues to create large libraries.

Creation of chimeric lectins is a rational, site-directed design technique that was performed early on in lectin engineering efforts. This is done by mutating the binding-site residues of one lectin to those of another lectin in order to influence the binding specificity. This method requires knowledge of the residues responsible for carbohydrate interactions, and two lectins that are predicted to have highly similar folds. Structural data is useful for this method, but not essential, and biochemical or sequencing information alone can be used for identification of binding-site residues in certain families of lectins. This has been explored thoroughly in mannose (Man)- and galactose (Gal)-binding C-type lectins, a family of lectins that utilize a coordinated Ca^{2+} for protein stability and enhancing carbohydrate binding.[69] Man-binding lectins have a conserved Glu-Pro-Asn motif in the carbohydrate binding site, while Gal-binding lectins instead have a Gln-Pro-Asp motif. Galactose binding was engineered in C-type Man-binding lectin (MBL) by mutagenesis of Glu185 and Asn187 create the Gln-Pro-Asp motif of C-type Gal-binding lectins.[70] Although Gal binding was achieved, the affinity was much lower than that of the native Gal-binding lectins, indicating other residues outside of the Gln-Pro-Asp motif are involved. To strengthen this interaction, additional binding-site residues of MBL were mutated. The C-type Gal-binding lectin asialoglycoprotein receptor (ASGPR) from rat liver contains a Trp residue and a five-residue glycine-rich insertion following the Gln-Pro-Asp motif that is not present in MBL (Figure 3A).[71] Addition of these residues into the Gln-Pro-Asp mutant of MBL led to increased affinity and selectivity of the mutant MBL for galactose to near ASGPR levels. Manipulation of this region has also been applied to create Man recognition in the Gal/N-acetylgalactosamine (GalNAc) specific C-type lectin *Bahinia purpurea* agglutinin (BPA). BPA was engineered to recognize Man by replacement of nine amino acids in the metal ion binding region with homologous residues found in the legume Man-binding *Lens culinaris* agglutinin (LCA).[72] The creation of chimeric lectins by swapping out one lectin binding site for another has shown some utility, but is only applicable to certain lectin families containing lectins of differing specificity with similar folds, and highly-characterized binding sites.

A lectin with high affinity for a tumor-associated epitope was the subject of site-directed engineering by mutagenesis of a single residue.[73] Peanut agglutinin (PNA) is a legume lectin with clinical importance due to its affinity to the tumor-associated O-glycan Galβ1–3GalNAc, known as the Thomsen-Friedenreich (TF) antigen. The residue Asn41 is critical for TF antigen specificity as it makes a water-mediated hydrogen-bonding contact with the GalNAc portion of the sugar.[74] Replacement of this residue with several different amino acids revealed that the Asn41Gln mutant shows enhanced affinity for the TF antigen, likely

due to a direct hydrogen-bonding interaction with GalNAc, made possible by the increased side-chain length of Gln compared to Asn.[73] Saturation mutagenesis, or the replacement of a residue with all other amino acids, was used successfully on the *Agrocybe cylindracea* galectin (ACG) to improve its selectivity. The WT ACG exhibits broad specificity as it recognizes β-galactosides like *N*-acetyllactosamine (LacNAc) and the TF antigen, as well as α2–3Sia-containing glycans such as 3'-sialyllactose (3'-SL). Sialic acid recognition is mediated by several key interactions with residues Ser44, Arg77, and Trp83 (Figure 3B).[75] In order to improve the specificity for α2–3Sia, residue Glu86, which makes important contacts with the Gal moiety of β-galactosides, was mutated to all other residues.[76] The resulting proteins were assessed by surface plasmon resonance (SPR) with immobilized multivalent glycopolymers; a Glu86Asp variant retained binding to 3'-SL, but lost affinity for the β-galactosides LacNAc and the TF antigen. It was confirmed that this mutant preferentially binds α2–3Sia-containing N-glycans over asialo N-glycans using frontal affinity chromatography, a biophysical technique which is used to determine the affinity of immobilized lectins to various glycans in a flow-based system.[77] A simple method to analyze the contributions of several residues to protein function is through the sequential mutation of residues to alanine, called alanine-scanning mutagenesis. This method was applied to the Gal binding site residues of the same ACG and produced proteins with high specificity for certain β-galactosides. Mutant Asn46Ala showed enhanced affinity for glycans terminating in GalNAcα1–3Galβ such as blood group A tetrasaccharide, and lost affinity to all other β-galactosides, as well as sialyl- and asialo-N-glycans. This specificity change can be explained by a *cis/trans* interconversion of Pro45 when the Asn46Ala mutant is made (Figure 3B).[78, 79] Another mutant, Glu86Ala, lost binding affinity for all glycans tested except those containing a 3'-sulfo-Galβ1–4GlcNAc structure.[80]

A powerful method for protein engineering is creation of combinatorial libraries by random mutagenesis of selected residues. This allows for the rapid production of many protein variants that can then be screened for functional binding sequences. A combinatorial library was made with the Gal/GalNAc-specific lectin BPA, mentioned previously for development of a Man-binding chimera, and phage display was utilized for selection of variants with affinity for Man (Figure 3A).[81] The nine amino-acid binding site of BPA was randomized, holding important Ca^{2+}-binding residues and a conserved Trp constant. Affinity panning using Man-bovine serum albumin (BSA) coated plates isolated phage clones with a strong preference for Man-BSA compared to fucose (Fuc)-, Gal-, GalNAc-, or GlcNAc-BSA. Surprisingly, none of the clones examined contained the sequence of the BPA/LCA chimeric Man-binding lectin, underscoring the utility of unbiased, randomized mutagenesis of a lectin binding site.[72] Phage display has also been used to pan a library of the α2–3Sia-specific plant lectin *Maackia amurensis* hemagglutinin (MAH) with human erythrocytes.[82] Although multiple binding-site residues were randomly mutated, selected clones varied at only two residues. These clones showed wild-type-like affinity to α2–3Sia-containing glycans, but some had also gained a novel affinity to α2–6Sia. In another example, mammalian cell display (Figure 2B) was used for engineering α2–6Sia specificity into the Gal-specific peanut agglutinin (PNA).[28] Mutagenesis of carbohydrate-interacting loops was performed, both to randomly change the amino acid sequence and to vary the length of each loop. An isolated clone gained the ability to bind α2–6Sia-containing glycans after positive selections

with Siaα2–6(Galβ1–3)GalNAc, though it still possessed some recognition of a terminal Gal.

2.2 Random Mutagenesis

Random mutagenesis over entire proteins is much less frequently used for lectin engineering, but this method has been successful for engineering of the Gal-binding earth worm lectin EW29. The C-terminal domain of this lectin, referred to as EW29Ch, underwent error-prone PCR to generate a mutant library, which was then used to isolate Sia and 6'-sulfo-Gal binders.[83, 84] Sialic acid selections were carried out using the high throughput method known as ribosome display (Figure 2B).[33, 85] Agarose beads modified with the glycoprotein Fetuin were used as the selection probe as this protein contains α2–6-sialylated triantennary N-glycan as a major glycan determinant. Analysis of individual clones identified an α2–6Sia-binding mutant containing six amino acid changes. Structural studies of this variant show that one of two Gal-binding subdomains was modified by a dramatic flip in a loop region, allowing for formation of a hydrogen bond between Sia and mutated residue Gly239Ser (Figure 3C). Such a change would be difficult to engineer by site-directed mutagenesis, highlighting the utility of the random mutagenesis method. Frontal affinity chromatography shows this new lectin, named SRC, maintains some Gal recognition of the parental EW29Ch, but specificity shifts significantly towards α2–6-sialylated N-glycans. This new specificity does come with some loss of affinity, as EW29Ch has a K_D value with Gal of about 10 μM, and SRC of about 100 μM. EW29Ch has also undergone selection for 6'-sulfo-Gal, again using ribosome display, but with agarose beads conjugated to biotinylated polyacrylamide polymers bearing 6'-sulfo-LacNAc. Clones bearing the Glu20Lys mutation showed 6'-sulfo-LacNAc binding, with K_D values determined using frontal affinity chromatography of ~3–4 μM, but maintaining a wild-type level of binding to Gal-terminated glycans with a K_D of 20 μM. The mutation is located close to the 6'-hydroxyl group and is expected to make favorable electrostatic interactions with the sulfate group.

2.3 Engineered Multivalency

Because lectins often show low affinity for their targets, efforts have been made to boost avidity not by engineering of the binding site itself, but by engineering multivalency. The previously described C-terminal domain of the earthworm lectin EW29 engineered for Sia binding (SRC) had affinity too low to be useful as a reagent in a monovalent state. The parent lectin, EW29, is a tandem repeat-like lectin with two homologous Gal-binding domains separated by a short linker.[86] The authors genetically fused two SRC domains together using a modified linker and found a 10-fold increase in affinity toward α2–6-sialylated glycans compared to the monovalent SRC (Figure 3C).[87] This affinity is comparable to that of the commercially available α2–6-Sia specific lectin from *Sambucus sieboldiana* (SSA), and was successfully used as a reagent for flow cytometry, fluorescence microscopy, lectin chromatography, and lectin blotting. A similar strategy was used to increase the affinity of a bacterial F-type lectin with Fuc specificity.[88] Partial duplication of the binding site to mimic eukaryotic Fuc-binding proteins led to a 12-fold greater binding affinity than the wild-type lectin to multivalent fucosylated glycans. Although only one Fuc binding site was functional, the protein gained avidity by oligomerizing into higher-order

structures. Another method of engineering lectin multivalency is to fuse the carbohydrate binding domain to a dimeric protein. To this end, a modified high-Man specific lectin from actinomycete (actinohivin) was fused to the fragment crystallizable (Fc) domain of human immunoglobulin G 1 (IgG1).[89] This dimeric "lectibody" showed a 10-fold improvement in binding to high-Man type glycans and maintained the wild-type specificity as assessed by glycan microarray (Figure 2C).

3. CARBOHYDRATE BINDING MODULES

Carbohydrate-binding modules (CBMs) are distinct from lectins as they are sugar-binding domains of a larger sugar-processing enzyme involved in the synthesis, transport, or metabolism of carbohydrates and glycan polymers. CBMs act to increase the catalytic efficiency of the enzyme by either targeting the enzyme to specific regions of the carbohydrate substrate, by increasing the concentration of the enzyme in the vicinity of the substrate, or by disrupting the polysaccharide structure to allow for easier enzyme access.[90–93] CBMs fall into three types.[94] Type A CBMs bind to the surface of crystalline polysaccharides, and have a planar binding face rich in aromatic residues. Type B CBMs are endo-type CBMs that recognize internal glycan chains with a binding cleft or groove that can accommodate multiple monosaccharides. Type C CBMs bind glycan termini with a small binding pocket that can only accommodate 1–3 monosaccharide units. Because of this, Type C CBMs are said to be more "lectin-like". According to the Carbohydrate-Active Enzymes Database (CAZy), these three CBM types are grouped into 87 families based on amino acid sequence.[95] They have also been grouped into seven families based on structure.[94] Like lectins, the most common fold is the β-sandwich, but β-trefoil, cysteine knot, oligonucleotide/oligosaccharide-binding (OB), and hevein folds are also found.[94]

CBMs often recognize plant and fungal cell wall polysaccharides like cellulose, xylose, and chitin, but glycoside hydrolase enzymes with CBM domains are also secreted by pathogenic microorganisms that act on human glycans such as hyaluronan.[96] CBMs, like lectins, often have low affinity for small oligosaccharides, but compensate with either multiple binding sites per domain or multiple domains per enzyme for increased avidity.[94] CBM engineering has mostly been applied with respect to industrially-important polysaccharides to increase enzyme action toward biomass, but may also be a good starting point for engineering of biotechnologically relevant CBMs. Some examples of CBM engineering and the methods used are described here. For other CBM engineering reviews, please see Armenta et al.[97]

3.1 Site-Directed Mutagenesis

Like lectin engineering, site-directed mutagenesis has been the most frequently applied approach for CBM engineering. The demand for cost-effective biofuel production has prompted the engineering of high-affinity binders of cellulose, as efficient cellulolytic enzymes are needed. Toward this goal, the cellulose-binding CBM1 of *Trichoderma reesei* cellobiohydrolase Cel7a has undergone site-directed mutagenesis to increase cellulose affinity.[98] This CBM is a small cysteine knot protein with a flat binding face rich in aromatic amino acids. A homologous cellulose-binding domain from an endoglucanase found in the same organism differs by only nine amino acids, but has higher affinity for cellulose. Four

variants were constructed by mutagenesis of seven different positions to the corresponding residues of the endoglucanase CBM to make two single and two triple mutants. Only the single Tyr5Trp mutant had increased affinity for cellulose, likely due to the differences in the π-systems in these aromatic amino acids (Figure 4A). This same CBM has also undergone site-directed mutagenesis to increase the specificity for cellulose over lignin, the major non-carbohydrate component of cellulosic biomass that can inhibit enzymatic hydrolysis.[99] Four residues were mutated to several other amino acids to vary the charge and polarity at these positions. The protein variants were produced as cellobiohydrolase-CBM complexes and their affinity for microcrystalline cellulose and lignin was quantified using partition coefficients between adsorbed protein to solid or soluble polysaccharide substrate. Substitution of amino acids Val and Pro with the negatively charged Glu at positions 27 and 30 both shifted specificity to cellulose over lignin (Figure 4A). These residues, when paired with several mutations in the linker between the CBM and cellobiohydrolase enzyme, has 2.5-fold reduction in lignin affinity and had no lignin inhibition when assayed for cellulose degradation.

The CBM4–2 of the *Rhodothermus marinus* xylanase, Xyn10A, has been the subject of many CBM engineering efforts. CBM4–2 is a type B CBM with affinity for xylans, β-glucans, and amorphous cellulose. This broad binding specificity, paired with the high thermostability and ease of production in *E. coli*, made this module an attractive candidate for protein engineering. Twelve residues around the binding site were selected for limited substitution of related residues in order to not destabilize the structure.[100] Phage display was used to select binders to the carbohydrate polymers birchwood xylan, Avicel (a microcrystalline cellulose) and ivory nut mannan, as well as the human glycoprotein IgG4. The clone selected for xylan, referred to as X-2, is highly specific for xylan, having lost the wild-type CBM4–2 affinity for glucan-containing polysaccharides like β-glucans and xyloglucans.[101] This clone is unique as one of the two binding site aromatic residues (Phe110) was mutated to the aliphatic residue Leu, leaving only one π-stacking interaction with the xylan chain. Clone X-2 with position 110 mutated back to the wild-type Phe is once again able to bind carbohydrates with glucose-based backbones, regaining the broad specificity of the wild-type CBM4–2 (Figure 4B).[102] The human IgG4-binding clone underwent a dramatic change in specificity as it was shown to bind to the protein itself, not the attached glycans.[103] This phage library has also been screened for binders to xyloglucan, a plant cell wall polysaccharide which lacks appropriate reagents for its study.[104] Xylan was used as a soluble competitor to remove phage displayed variants with wild-type specificity toward xylan. This competition strategy was successful, as two selected clones showed remarkable specificity toward xyloglucan over xylan, Avicel, arabinoxylan, and β-glucans. Structural studies of one clone, XG-34, shows the binding cleft is more narrow than that of the wild-type protein, with Trp69Tyr and Tyr110His moving closer by 5.5 Å (Figure 4C).[105]

3.2 Random Mutagenesis

Random mutagenesis has been performed successfully to change CBM specificity. The family 11 CBM of the *Ruminoclostridium thermocellum* enzyme CelH, RtCBM11, has undergone randomized mutagenesis to create a combinatorial phage library.[106] The wild-type RtCBM11 binds to linear polysaccharides such as glucans and Avicel with high

affinity, but binds branched polysaccharides like xyloglucan with low affinity. Structural studies of the wild-type RtCBM11 complexed with mixed-linked β-glucans show that CH-π stacking with Tyr residues and a hydrogen-bonding network with multiple residues are responsible for ligand binding.[107] Phage selection using xyloglucan revealed a double mutant binder with His102Leu and Tyr152Phe mutations. This variant has about 22-fold higher affinity for xyloglucan compared to the wild-type protein and has reduced affinity for β-glucan. Molecular dynamics simulations were employed to rationalize this enhanced affinity toward the branched xyloglucan and reveal the creation of a xylosyl binding cleft at His102Leu and modified hydrogen-bonding network. Importantly, fusion of the mutant CBM to a xylanase enzyme increased catalytic efficiency of xyloglucan hydrolysis by 38%. Randomized mutagenesis can greatly improve the binding characteristics of a protein that has already been selected from a combinatorial library. This process, referred to as affinity maturation, was performed on the xyloglucan-binding clone XG-34 isolated from the CBM4–2 combinatorial library described previously.[104] Error-prone PCR introduced random mutations throughout the length of the gene to generate a new phage display library.[108] Selections for tight xyloglucan binders produced several clones with higher affinity for xyloglucan than the parent protein. These clones shared a single mutation in close proximity to the carbohydrate-binding site that reverts Asp112 back to the wild-type residue Glu112 that is speculated to directly interact with the bound ligand (Figure 4C). These clones showed specificity for galactose-decorated xyloglucans, with no affinity for fucosylated xyloglucans. The evolved proteins were fluorescently labeled and subsequently used to visualize the non-fucosylated xyloglucan found in tamarind seed, performing better than the parent protein.

4. ADAPTIVE IMMUNE PROTEINS

4.1 Antibodies (Abs)

Unlike lectins and CBMs, antibodies (Abs) are part of the mammalian immune system and form the foundation for the adaptive response to foreign antigens. Ab proteins contain a constant (Fc) region and two variable (Fv) domains composed of heavy (H) and light (L) chains (Figure 5A). The Fv domains provide clonal diversity for specific antigen recognition; the Ab repertoire includes more than 10^{12} different variants in humans.[109] Such diversity typically generates Abs with diverse structural features, which exhibit high specificity and affinity to their target.

While Abs are exceptional for detecting protein antigens, they are not as robust in detecting carbohydrates for many reasons.[110, 111] Anti-glycan immune responses are less T-cell dependent, as polysaccharides do not always recruit T-cell assistance for B-cell maturation.[112, 113] As a result, the immunoglobulin G (IgG) and M (IgM) Ab isotypes elicit the greatest response to glycan antigens, and their somatic mutation to mature Abs is decreased compared to other antigenic groups. The IgG isotype is bivalent, with binding regions approximately 50–100 Å apart (Figure 5A). The IgM isotype is a pentavalent arrangement of the Ab scaffold, where the individual binding domains exhibit lower affinities than IgG, but the Ab-glycan avidity is increased through ten possible binding sites.[114] Abs also preferentially target larger, more complex antigens, and therefore tend to

recognize the non-glycan (e.g. peptidic or lipidic) portion of a biomolecule antigen rather than the carbohydrate. Carbohydrates are not very immunogenic in traditional monoclonal Ab (mAb) production hosts due to the similarity of glycans from mammalian sources and many bacteria employ antigenic mimicry to evade mammalian host immune systems.[113] Accessibility of pure glycans further hinders mAb development. Glycans may not be easily synthesized or isolated from the host organism, rendering many interesting glycans from pathogens and symbionts inaccessible.

The generation of Ab-based GBP reagents is a long and often unrewarding process. Despite various limitations, anti-glycan mAbs have been developed against a number of different targets, including cancer epitopes,[20, 115–117] glycosaminoglycans,[118] human blood group oligosaccharides,[119] viral envelope proteins,[120–122] bacterial cell wall components,[123–125] and many more. Many studies exploring anti-glycan Abs, either generated as monoclonals or isolated directly from host-derived samples, have been recently reviewed by Haji-Ghassemi et al., focusing on their carbohydrate-antibody recognition mechanisms.[113] Much of the current anti-glycan Ab landscape has been collated in the Database of Anti-Glycan Reagents (DARG).[110] The common engineering approaches used to improve anti-glycan Abs are summarized. Despite their prevalence in the glycobiology field, anti-glycan Abs still suffer from the specificity, affinity and development issues described above and, as a result, could benefit greatly from novel or orthogonal engineering tactics.

4.1.1 Fragmentation and Phage Display—Manipulations of the prototypic immunoglobulin form are prominent in Ab engineering. Removal of the crystallizable fragment (Fc) results in the antigen binding fragment (Fab), which is frequently employed for the development of Abs as reagents and for structure determination (Figure 5B). The HIV-1 Ab PGT128 is one example of an anti-glycan Fab which recognizes the high-Man region and a short β -strand segment of the gp120 envelope protein.[126] Further fragmentation to the single-chain variable fragments (scFv) (Figure 5B) generates another major scaffold in Ab engineering as these variants may exhibit superior pharmacokinetic properties, are smaller and more amenable to biochemical manipulations, and are generally easier to produce *in vitro* compared to mAbs.[127] Due to their small size, they also present ideal scaffolds for a phage display library. However, scFvs also exhibit prominent disadvantages including decreased stability, aggregation and misfolding when recombinantly produced in *E. coli*.[128, 129] Therefore, it is common to conjugate scFvs with a crystallizable fragment (Fc) to recapitulate some of the IgG structure (scFv-Fc) (Figure 5B) and regain stability.[130] Nevertheless, the development of anti-glycan scFvs has been commonly explored through the general procedure of immunization, scFv gene isolation, phage expression system incorporation and phage display enrichment. This is exemplified by an scFv that recognizes Man-6-phosphate (M6P), a glycan determinant required for transport of lysosomal hydrolases.[131] Rabbits immunized with a pentamannose phosphate afforded an Ab library that was subsequently enriched by phage display to provide a single scFv variant that displayed an apparent K_D of 28 μM for the M6P substrate and does not exhibit affinity for Man or Glc-6-phosphate. After interrogation with a phosphorylated glycan array, it was found that this scFv Ab binds specifically mono- and diphosphorylated Man_6 glycans and diphosphorylated Man_7 glycans, all containing the M6P determinant. Structural studies of

the scFv:M6P complex reveal six hydrogen-bonding contacts with the Man ring and two salt bridge contacts with the phosphate group imparting specificity.[132]

In another example, Kubota et al. utilized this method to target an α-linked GalNAc, also known as the Tn antigen, a prominent cancer epitope.[133] The phage library was developed using extracted RNA from mouse spleen cells immunized with the Tn antigen. The resulting scFv library was then incubated with a biotin-conjugated Tn antigen; clones were isolated using streptavidin-coated magnetic beads and binding was verified by ELISA. Selected variants were then genetically fused to an Fc domain to promote Ab-dependent cellular cytotoxicity. Two scFv-Fc variants, 3–9 and 3–18, were obtained with high specificity for the Tn antigen. The marked specificity was presumed to result from the mouse pre-immunization and a negative-panning step against blood group A (BGA) trisaccharide (another prevalent antigen with the same terminal a-GalNAc determinant). Although binding affinities were not determined, the two variants obtained from this study were confirmed to selectively recognize the Tn antigen-expressing cell lines Jurkat and CHO-Lec8.

This approach was routinely utilized in many studies to develop anti-glycan antibodies against the GAGs chondroitin sulfate (CS), dermatan sulfate (DS), heparan sulfate (HS) and heparin epitopes.[134] Since GAGs are non-immunogenic in traditional mAb hosts, a semisynthetic human library with 50 VH genes conjugated to the same VL gene then fused to a phage coat protein was utilized for the scFv library. Binders to the GAG epitopes were selected in an iterative manner using glass-surface immobilized GAGs. After multiple rounds of scanning, the resulting scFvs were then subcloned into bacteria for soluble expression and analysis. One particular study that followed this procedure to isolate scFvs to HS afforded three variants, one of which, HS3G8, was obtained after negative-selection screens against CS and DS.[135] An apparent binding affinity of HS3G8 to HS was determined to be 0.15 μM by ELISA. The HS specificity was characterized by immunofluorescence after incubation with GAG-digested and undigested tissue sections. Immunostains of these tissues with HS3G8 showed affinity for the HS target while showing no affinity for other GAGs and polyanionic species, including CS, DS and DNA. Interestingly, the use of this scFv Ab was able to discern heterogeneity of HS in different tissues. These examples highlight the use of Ab scaffold fragments, namely scFvs, and a phage display system to provide Ab-based GBPs. The procedures further emphasize that orthogonal counter selections are key for imparting greater specificity for the target glycan. However, these studies did not incorporate affinity maturation steps in between selection rounds, which could potentially instill greater affinity for the glycan.

4.1.2 Mutagenesis and Engineered Multivalency—

Anti-glycan Abs from an immunized host exhibit reduced mutagenesis, as the immune response to glycan targets presents as a lack of rearranged variable-region genes.[136] This results in essentially the germline sequence, providing a dearth of clonal diversity and generally low affinity, broadly-specific Abs. Therefore, *in vitro* mutagenesis techniques, as discussed for lectins and CBMs, should effectively re-install desired binding characteristics to existing anti-glycan Ab scaffolds. Such mutagenesis has afforded varying success, usually resulting in anti-glycan Abs with increased affinity but decreased specificity.[137] A random codon-based mutagenesis strategy was applied to the anti-Lewis Y (Fucα1–2Galβ1–4(Fucα1–3)GlcNAc; Le^Y) Ab,

BR96 (Figure 5C).[138] Originally expressed as an Fab, BR96 was first fragmented to an scFv for improved expression. Mutagenesis of the three most exposed heavy chain loops, H1, H2 and H3, within the complementarity-determining regions (CDRs) afforded variant M1; the single mutation of Asp97Ala in the H3 region, provided 5 to 10-times greater binding than parent BR96. A double mutant was then developed by site-directed mutagenesis to introduce Asp in the H2 region (Gly53Asp) and bound even better than M1. Affinity maturation on the H1 region of M1 was performed by repeating the codon-based mutagenesis; one triple mutant, M4, exhibited greater binding affinity to Le^Y-coated ELISA plates but diminished binding to Le^Y-positive cells. Although this study was able to produce some Abs with higher binding properties relative to the parent Ab, the specificities decreased (Figure 5C). An scFv was also derived from the anti-BGA trisaccharide IgM, AC1001, and a combination of rational design and phage display was utilized to explore increasing binding affinity.[119] Site-directed mutants of the Leu at position H103 to Ile and Val led to 15- to 30-fold increased binding to BGA compared to the parent scFv. This mutation, though it had a dramatic effect on affinity, did not directly contact the antigen, rather stabilized a nearby loop to reduce entropic contributions to binding. The single mutant scFVs did not show binding to blood group B (BGB) trisaccharide, maintaining some specificity, although a thorough binding analysis with other targets was not performed.

As highlighted previously, protein-glycan binding interactions are characteristically weak and rely on multivalency to increase binding avidity. Exploiting protein oligomerization for multivalent interactions in Ab engineering is usually seen as a genetic fusion of an Ab fragment back with its Fc domain (*e.g.* scFv-Fc). One study used a ligase chain reaction mutagenesis strategy to develop scFvs against the O-polysaccharide of *Salmonella*.[139] Phage panning provided mutants rich in substitutions that increased hydrogen-bonding contacts to the polysaccharide and substitutions that removed steric clashes in distal loop regions. Interestingly, this library was predominantly enriched with dimeric and higher order oligomeric scFvs with apparent K_Ds in the low nanomolar region, confirming the importance of multivalency for greatest avidity.

One unique case of multivalency is exemplified by 2G12, another anti-HIV Ab which recognizes the high-Man glycan of gp120.[140] In 2G12, a phenomenon called domain swapping allows for an Ab conformation with four, proximal binding sites as opposed to two, distal ones (Figure 5D). The two variable heavy chains, VH and VH', cross to the opposite binding region, which brings all four variable sites in close proximity for multivalent binding (Figures 5D). The Lai group developed a phage display approach for engineering glycan-binding Abs utilizing this domain-swapped chain arrangement.[141] The phage library utilized a chimeric construct comprising the stalk, hinge, and variable regions; the stalk and hinge regions were optimized to bring the variable chains into the proper orientation and proximity for domain swapping. A survey of existing anti-glycan Abs and the amino acids that contact the glycan with a van der Waals distance <2.5 Å, revealed that Tyr, Ser and Asp are prevalent in all anti-glycan Abs – in contrast, Arg, Asn and His prevalence are antigen-dependent. Two phage libraries were then developed with random mutagenesis at the *NMC* codon (where N and M are any nucleic acid), which can encode Tyr, Asp, Ser, His, Asn, Thr, Ala and Pro. Although this codon represents most of the high-frequency residues identified in glycan-Ab interactions, it does not encode for Trp,

the amino acid involved in prominent CH-π interactions in GBP binding (See Section 1.3 and Figure 5D). This approach, although conceptually sound and able to provide domain swap-engineered Abs against gp120, the EC_{50} values paled in comparison to the wild-type 2G12.

In summary, many factors contribute to relative lack of engineered Abs compared to lectin and CBM GBPs. There has been relatively limited success in anti-glycan Ab development to highly-selective and tight-binding proteins, likely due to the restricted number of genes that comprise the variable region and the lack of T-cell helper functions for affinity maturation.[113] It is also challenging to develop ideal glycan antigens for mAb generation, especially for bacterial glycans, which are more difficult to isolate or synthesize, and can mimic the mammalian host glycans. Therefore, although there have been attempts at further engineering antibodies by fragmentation, display libraries for evolution, and mutagenesis, these efforts have not resulted in significant improvements in glycan-binding properties.

4.2 Lamprey Variable Lymphocyte Receptors (VLRBs)

Lamprey, a jawless vertebrate, also exhibits an adaptive immune system, in this case *via* variable lymphocyte receptors (VLRs).[142] Recent investigations into these variable antigen receptors have facilitated the development of novel GBPs which exhibit strong binding characteristics to carbohydrate antigens. VLRs gain specificity against foreign antigens through somatic recombination of leucine-rich repeat (LRR) modules (Figure 6A), providing a diversity of 10^{14} different proteins.[143] Lamprey have two VLR genes, *VLRA* and *VLRB*, which are expressed by mutually exclusive lymphocyte populations. VLRB-related proteins have been shown to recognize various foreign antigens and are more widely utilized in the development of binders as they are the prevalent component of the humoral response of the VLR-based immune system.[144]

Carbohydrates are highly immunogenic in lampreys;[145] a few examples in recent years show VLRBs with high affinity and specificity to glycan targets. In 2008, Alder et al. showed that lamprey lymphocytes responded to carbohydrate and protein determinants on bacterial or mammalian cells.[144] Lamprey immunized with human O erythrocytes produced VLRB antibodies that recognized Chinese hamster ovary (CHO) cells expressing the blood group H (BGH) trisaccharide (Fucα1–2Galβ1–4GlcNAc), the key antigen on O erythrocytes. This study also established that VLRBs are composed of multiple VLRB monomers linked by disulfide bonds to form large oligomeric macromolecules; this is the nascent form in the sea lamprey, leading to the high-binding avidities for glycans as they are naturally multivalent immune proteins. In addition, the study showed VLRBs are unresponsive to the soluble proteins BSA or keyhole limpet hemocyanin, two proteins commonly used as conjugates for vaccine development. The utility of the lamprey VLRB system is further emphasized by a comparison of mouse and lamprey glycosyltransferases, which suggests that the lamprey glycan-related genome and glycome are distinct from those of humans and mice.[146] This is promising for the application of glycans and glycoconjugates in lamprey immunization for anti-glycan VLRB development.

VLRBs are excellent candidates as next generation GBPs for many reasons. The lamprey VLRB and mouse immunoglobulin responses to influenza A virus are extremely similar,[142]

yet advantageous structural features provide the VLRB with a greater ability for glycan binding. A structural analysis exemplified this by showing VLRB antigen contact area (~1500 Å^2) is similar to that of Igs (1400–2300 Å^2), but results in a deeper binding pocket that promotes greater binding affinity. Furthermore, the concave structure contains a “thumb” from the C-terminal LRR capping region (LRRCT), which can effectively clamp down onto the carbohydrate antigen (black arrows in Figures 6A).[147] The VLRBs utilize the same non-covalent forces for binding (salt bridges, hydrogen bonds, and van der Waals forces) as other GBPs, and sequence analysis also reveals an enrichment of aromatic amino acids. Indeed, the variable positions on the concave surface, which contact the carbohydrate antigen, are highly biased towards Tyr, Trp, Asn and Asp.[145] The structural “thumb” specifically contains a conserved Trp residue that utilizes CH-π interactions at the carbohydrate binding site. Understanding the molecular determinants for VLRB-carbohydrate binding will inform mutagenesis efforts to engineer VLRBs with greater specificity and/or affinity.

4.2.1 Yeast Surface Display—VLRBs have been shown to be amenable to engineering tactics to generate GBP reagents with desirable characteristics. Yeast-surface display (YSD) has emerged as a predominant VLRB engineering technique (Figure 6E). Tasumi et al. produced VLRBs in immunized lamprey and consequently developed a YSD vector for screening of the VLRB library.[143] Two libraries were developed: a hen egg lysozyme (HEL) library from HEL-immunized lamprey, and a composite library from lamprey immunized with various antigens including β-gal and sheep erythrocytes. The libraries were incorporated into a YSD system and screened for binding to targets including multivalent proteins as well as BGA and BGB trisaccharides. Seven VLRB binders to the trisaccharides were identified, and 6 of the 7 exhibited 1.6- to 4.3-fold higher affinity for BGB compared to BGA, with apparent K_Ds in the 10–900 nM range for all clones. Further, the anti-trisaccharide VLRBs showed no inhibition of binding in the presence of BGH, indicating specificity among the blood group antigens. This study also employed error-prone PCR and FACS enrichment to increase affinity of the isolated protein VLRB.HEL.2D; although this was not a glycan-binding protein, it still exemplifies the possibility of utilizing diversity-generating methods for engineering anti-glycan VLRBs.

In 2013, Hong et al. utilized a YSD library expressing VLRs in a screen against biomedically-relevant glycotopes including the Tn and TF antigens, Lewis A and Lewis X, *N*-glycolylneuraminic acid, poly-Man9, HIV gp120, and glycoproteins asialo-ovine submaxillary mucin (aOSM) and asialo-human glycophorin A (aGPA).[148] Initial screens were performed with a monomeric VLRB library, and the selected VLRBs were then fused with an Fc domain to provide a dimeric VLRB-Fc protein. One of the isolated proteins, VLRB.aGPA.23, was shown to be selective for the BGH trisaccharide, aGPA, and TF by flow cytometry and microarray analysis. Using SPR with its fusion construct, the apparent binding affinities to each glycan target were determined to be in the low nM range. Isothermal calorimetry (ITC) performed on monomeric proteins provided a K_D of 0.221 mM, suggesting that the previously determined nanomolar apparent affinity resulted from multivalent interactions.[149] Tissue microarrays with this VLRB selectively detected cancer-associated carbohydrate antigens in 14 different cancers. Later, the crystal structure of the

VLRB.aGPA.23-TF complex was determined at 2.2 Å resolution.[149] The structural analysis reveals the basis of specificity for the tumor-associated antigen. Key hydrogen-bonding contacts are made between residues Asn86, Asp134 and Ser87 and the TF disaccharide hydroxyls. Trp residues 62, 84, 156 and 187 also contribute to specificity binding through the aforemenmtioned CH-π interactions. Meanwhile, significantly fewer hydrogen-bonding, van der Waals, and CH-π interactions are observed for the VLRB.aGPA.23-BGH. This structure also corroborated the general molecular architecture for VLRB-based glycan binding. The TF disaccharide is sandwiched between the LRRCT domain molecular thumb and the concave surface formed by the short β-strands of the LRR and CP modules (Figures 6A). This is atypical compared to lectins and Abs and allows for greater contacts along the β-strands by larger oligosaccharide targets, potentially accommodating up to 6 monosaccharides.

4.2.2 Mutagenesis and Microarray Enrichment—Specificity for VLRB-based antibodies, like other GBPs, is predominantly determined through the use of glycan microarrays. Lampreys immunized with O erythrocytes generated VLRBs that recognize the BGH trisaccharide.[145] YSD on the immunized library followed by MACS and FACS generated specific VLRBs. The isolated VLRBs were then conjugated to an Fc domain (Figure 6C) to generate the dimeric form for glycan microarray analysis, which revealed a greater specificity for the BGH trisaccharide compared to plant lectin UEA-1. Purified monomeric VLRBs were utilized for isothermal calorimetry (ITC) studies, which found that one VLRB, O-13, exhibited a K_D of 2.6 μM for the BGH trisaccharide. O-13 did not bind 2'-fucosyl-Lac or Le^Y antigens and bound lacto-n-neotetraose (LNnT) with low affinity (K_D 160 μM). A combination of affinity measurements, microarray analysis and crystal structure comparison revealed a greater specificity of the O-13 VLRB for the BGH trisaccharide compared to other selected VLRBs. Mutational studies of the VLRB O-13 then followed; mutants with further enhanced specificity for H-trisaccharide were developed based on site-directed mutagenesis to eliminate cross-reactivity with LNnT, while retaining high affinity interactions with the BGH trisaccharide. One double mutant (Asn81His-Asn82Glu) decreased the size of O-13 binding pocket and eliminated a water-mediated hydrogen bond between position 81, position 82 and C2 hydroxyl group of the internal Gal residue of LNnT (Figure 6D). ITC showed that this mutant binds the BGH trisaccharide with the same affinity as WT O-13, but does not bind LNnT.

Finally, a combination of direct lamprey immunization, YSD, immunoglobulin Fc domain conjugation and microarray analysis afforded a robust platform for generating libraries of VLRB-based anti-glycan reagents (Figure 6E).[150] Identification of glycan-specific VLRBs after immunization with whole fixed cells, tissue homogenates, and human milk was achieved with this platform. The cDNA from lamprey lymphocytes was cloned into a YSD system for VLRB enrichment. VLRB-Fc chimeras (smart anti-glycan reagents, SAGRs) were constructed and specificity was determined by microarray analysis and immunohistochemistry. Fifteen VLRBs were discovered that discriminated between various glycosidic linkages, functional groups and unique presentation of terminal glycan motifs, providing a high throughput method for obtaining binders to a variety of carbohydrate

antigens. The VLRB/YSD system to recognize and enrich glycan binders is a nascent yet promising tool for advancing the scope of GBPs.

5. CONCLUSION AND FUTURE OUTLOOK

Glycan binding proteins are critical reagents for biotechnology and biomedical research and the engineering of novel GBPs is necessary to create reagents with properties superior to those of naturally-existing GBPs, and to address the large diversity of carbohydrate structures found in nature. The lectins, carbohydrate binding modules, antibodies, and lamprey variable lymphocyte receptors (VLRBs) described within this review have expanded the landscape of GBPs well beyond the scope of natural proteins as the foundation for reagents with greater specificity and avidity for previously unaddressed glycan targets.

The utility of lectins in biomedical science has long been known. Many lectins, generally of plant and fungal origin, are well characterized and commercially available, and many efforts to modify these proteins have been published. The predominant engineering approaches for lectins are based on structure-guided site-directed mutagenesis. These include approaches where the diversity of protein variants explored is low, such as the creation of chimeras by swapping of binding site residues from one lectin to another, and mutation of one or several residues to a subset of other residues. Site-directed mutagenesis has also been used to make high-diversity combinatorial libraries by creating variation at targeted residues at or near to the sugar binding site. There are, however, fewer reports of library generation using extensive randomized mutagenesis, but this technique has been successful for lectin EW29. Creation of synthetic multimers by duplication or fusion to a multimerization domain is another strategic approach for boosting lectin affinity and does not require modification of the lectin sequence itself.

There are fewer reports of CBM engineering than lectin engineering, however, like lectins, structure-guided site-directed mutagenesis is the most commonly used method. CBM4–2 of *R. marinus* has been the object of many CBM engineering efforts. A combinatorial library with variable residues proximal to the binding cleft of the protein has been used to isolate binders of many structurally-distinct polysaccharides. Although most CBM engineering studies have targeted plant polysaccharides, the development of CBMs that recognize other biotechnologically-relevant polysaccharides is possible. For example, hyaluronan is structurally similar to other glycosaminoglycans (GAGs) and the CBMs that recognize it can serve as a starting point for the directed evolution of CBMs that recognize heparan sulfate and other GAGs.

Adaptive immune proteins are also exciting starting points for the development of novel GBP reagents. To date, antibody engineering has employed fragmentation techniques, mutagenesis and phage display and multivalency is incorporated by recapitulating an Ab fragment with an Fc domain. Additionally, a recent method of incorporating multivalency involved the development of a domain-swapped Ab phage library. However, even after exploiting such tactics, engineered Abs may be limited by poor affinity and/or specificity. On the other hand, lamprey adaptive immune proteins, VLRBs, have recently provided highly-specific and tight-binding GBPs. The VLRB development strategy includes lamprey

immunization with a target glycan followed by yeast surface display and site directed mutagenesis. With this strategy anti-BGH, anti-Tn and anti-TF VLRBs have been obtained, suggesting that 1) the VLRBs represent a robust GBP generation platform and 2) implementation of tactics, such as affinity maturation, could further improve the affinity or specificity of these reagents.

Alternative scaffolds can be also explored for GBP development. Single-domain Abs (sdAbs) can be obtained by isolating the variable heavy chain (V_HH) of animals such as camelids and cartilaginous fishes, or that of the mammalian IgG Ab. Such sdAbs have been utilized in various studies for antigen recognition,[151] although they have not yet been employed for anti-glycan development. These small Abs, also called “nanobodies,” are good scaffolds for phage[152] and yeast display,[153] and the single domain can be engineered into a multivalent display for greater avidity. Designed ankyrin repeat proteins (DARPins), which are structurally similar to VLRBs, represent another potential scaffold for GBP development.[154]

For all GBPs, the ability to obtain pure glycans is critical to all aspects of GBP engineering. Whether through chemical/chemoenzymatic synthesis or isolation from nature, it is essential that continued efforts to increase glycan accessibility are pursued. For immune-based GBPs, developments in the synthesis of glycoconjugates are also important. Robust chemistries for glycoconjugate generation exist, although creative methods for stimulating immune responses could enhance many efforts.[155, 156] For example, glycan antigens have recently been appended to virus-like particles to bring B- and T-cells in close proximity and increase helper T-cell function, and therefore Ab maturation, during the immune response.[112]

Together, the progress to date clearly advises that investigating alternate protein scaffolds, improving glycan availability and glycoconjugation methods, and advancing microarray technologies will greatly promote the success of future GBP engineering efforts leading to broadly-available reagents for fundamental and applied research and interrogation of the “glyco-universe”.

ACKNOWLEDGEMENTS

We thank Hannah Bernstein for valuable input on this manuscript. We also gratefully acknowledge support from the NIH Biotechnology Training Program T32-GM008334 (EMW), the NIH T32 Postdoctoral Fellowship program F32-GM137477 (MK) and U01 CA231079 (BI).

REFERENCES

1. Varki A (2017) Biological roles of glycans, Glycobiology 27, 3–49. [PubMed: 27558841]
2. Toukach PV, and Egorova KS (2015) Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts, Nucleic Acids Res. 44, D1229–D1236. [PubMed: 26286194]
3. Imperiali B (2019) Bacterial carbohydrate diversity — a Brave New World, Curr. Opin. Chem. Biol. 53, 1–8. [PubMed: 31176085]
4. Han L, and Costello CE (2013) Mass spectrometry of glycans, Biochemistry. Biokhimiia 78, 710–720. [PubMed: 24010834]
5. Leymarie N, and Zaia J (2012) Effective use of mass spectrometry for glycan and glycopeptide structural analysis, Anal. Chem. 84, 3040–3048. [PubMed: 22360375]

6. Jin S, Cheng Y, Reid S, Li M, and Wang B (2010) Carbohydrate recognition by boronolectins, small molecules, and lectins, Med. Res. Rev. 30, 171–257. [PubMed: 19291708]
7. Sprinzl M, Milovnikova M, and Voertler CS (2006) RNA Aptamers Directed Against Oligosaccharides, In RNA Towards Medicine (Erdmann V, Barciszewski J, and Brosius J, Eds.), pp 327–340, Springer Berlin Heidelberg, Berlin, Heidelberg.
8. Wei S, Lupei D, and Minyong L (2010) Aptamer-Based Carbohydrate Recognition, Curr. Pharm. Des. 16, 2269–2278. [PubMed: 20482502]
9. Taylor ME, Drickamer K, Schnaar RL, Etzler ME, and Varki A (2015-2017) Discovery and Classification of Glycan-Binding Proteins, In Essentials of Glycobiology (Varki A, Cummings RD, and Esko JD, Eds.) 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. Matsumoto A, and Miyahara Y (2017) 'Borono-lectin' based engineering as a versatile platform for biomedical applications, Sci. Technol. Adv. Mater. 19, 18–30. [PubMed: 29296128]
11. Mazik M, Cavga H, and Jones PG (2005) Molecular Recognition of Carbohydrates with Artificial Receptors: Mimicking the Binding Motifs Found in the Crystal Structures of Protein−Carbohydrate Complexes, J. Am. Chem. Soc. 127, 9045–9052. [PubMed: 15969582]
12. Wu AM, Lisowska E, Duk M, and Yang Z (2008) Lectins as tools in glycoconjugate research, Glycoconj. J. 26, 899.
13. Shoseyov O, Shani Z, and Levy I (2006) Carbohydrate binding modules: biochemical properties and novel applications, Microbiol. Mol. Biol. Rev. 70, 283–295. [PubMed: 16760304]
14. Brooks SA (2017) Lectin Histochemistry: Historical Perspectives, State of the Art, and the Future, In Histochemistry of Single Molecules: Methods and Protocols (Pellicciari C, and Biggiogera M, Eds.), pp 93–107, Springer New York, New York, NY.
15. Mujahid A, and Dickert LF (2016) Blood Group Typing: From Classical Strategies to the Application of Synthetic Antibodies Generated by Molecular Imprinting, Sensors 16.
16. Slifkin M, and Doyle RJ (1990) Lectins and their application to clinical microbiology, Clin. Microbiol. Rev. 3, 197–218. [PubMed: 2200603]
17. Hashim OH, Jayapalan JJ, and Lee C-S (2017) Lectins: an effective tool for screening of potential cancer biomarkers, PeerJ 5, e3784–e3784. [PubMed: 28894650]
18. Bies C, Lehr C-M, and Woodley JF (2004) Lectin-mediated drug targeting: history and applications, Adv. Drug Del. Rev. 56, 425–435.
19. Sedlik C, Heitzmann A, Viel S, Ait Sarkouh R, Batisse C, Schmidt F, De La Rochere P, Amzallag N, Osinaga E, Oppezzo P, Pritsch O, Sastre-Garau X, Hubert P, Amigorena S, and Piaggio E (2016) Effective antitumor therapy based on a novel antibody-drug conjugate targeting the Tn carbohydrate antigen, Oncoimmunology 5, e1171434. [PubMed: 27622021]
20. Prendergast JM, Galvao da Silva AP, Eavarone DA, Ghaderi D, Zhang M, Brady D, Wicks J, DeSander J, Behrens J, and Rueda BR (2017) Novel anti-Sialyl-Tn monoclonal antibodies and antibody-drug conjugates demonstrate tumor specificity and anti-tumor activity, MAbs 9, 615–627. [PubMed: 28281872]
21. Coelho LCBB, Silva P. M. d. S., Lima V. L. d. M., Pontual EV, Paiva PMG, Napoleão TH, and Correia M. T. d. S. (2017) Lectins, Interconnecting Proteins with Biotechnological/ Pharmacological and Therapeutic Applications, Evid. Based Complement. Alternat. Med. 2017, 1594074. [PubMed: 28367220]
22. Bah CSF, Fang EF, and Ng TB (2013) Medicinal Applications of Plant Lectins, In Antitumor Potential and other Emerging Medicinal Properties of Natural Compounds (Fang EF, and Ng TB, Eds.), pp 55–74, Springer Netherlands, Dordrecht.
23. Breitenbach Barroso Coelho LC, Marcelino dos Santos Silva P, Felix de Oliveira W, de Moura MC, Viana Pontual E, Soares Gomes F, Guedes Paiva PM, Napoleão TH, and dos Santos Correia MT (2018) Lectins as antimicrobial agents, J. Appl. Microbiol. 125, 1238–1252. [PubMed: 30053345]
24. Mitchell CA, Ramessar K, and O'Keefe BR (2017) Antiviral lectins: Selective inhibitors of viral entry, Antiviral Res. 142, 37–54. [PubMed: 28322922]
25. Marshall SA, Lazar GA, Chirino AJ, and Desjarlais JR (2003) Rational design and engineering of therapeutic proteins, Drug Discov. Today 8, 212–221. [PubMed: 12634013]
26. Arnold FH (1998) Design by Directed Evolution, Acc. Chem. Res. 31, 125–131.

27. Sen S, Venkata Dasu V, and Mandal B (2007) Developments in Directed Evolution for Improving Enzyme Functions, Appl. Biochem. Biotechnol. 143, 212–223. [PubMed: 18057449]
28. Soga K, Abo H, Qin S-Y, Kyoutou T, Hiemori K, Tateno H, Matsumoto N, Hirabayashi J, and Yamamoto K (2015) Mammalian Cell Surface Display as a Novel Method for Developing Engineered Lectins with Novel Characteristics, Biomolecules 5, 1540–1562. [PubMed: 26287256]
29. Navaratna T, Atangcho L, Mahajan M, Subramanian V, Case M, Min A, Tresnak D, and Thurber GM (2020) Directed Evolution Using Stabilized Bacterial Peptide Display, J. Am. Chem. Soc. 142, 1882–1894. [PubMed: 31880439]
30. Boder ET, and Wittrup KD (2000) Yeast surface display for directed evolution of protein expression, affinity, and stability, Methods Enzymol. 328, 430–444. [PubMed: 11075358]
31. Fernandez-Gacio A, Uguen M, and Fastrez J (2003) Phage display as a tool for the directed evolution of enzymes, Trends Biotechnol. 21, 408–414. [PubMed: 12948674]
32. Yan X, and Xu Z (2006) Ribosome-display technology: applications for directed evolution of functional proteins, Drug Discov. Today 11, 911–916. [PubMed: 16997141]
33. Plückthun A (2012) Ribosome Display: A Perspective, In Ribosome Display and Related Technologies: Methods and Protocols (Douthwaite JA, and Jackson RH, Eds.), pp 3–28, Springer New York, New York, NY.
34. Xu L, Aha P, Gu K, Kuimelis RG, Kurz M, Lam T, Lim AC, Liu H, Lohse PA, Sun L, Weng S, Wagner RW, and Lipovsek D (2002) Directed Evolution of High-Affinity Antibody Mimics Using mRNA Display, Chem. Biol. 9, 933–942. [PubMed: 12204693]
35. Aleksei T, Olga S, Guzel F, Nikolay K, and Alla R (2020) Glycan-specific antibodies as potential cancer biomarkers: a focus on microarray applications, Clin. Chem. Lab. Med, . [PubMed: 20191161]
36. Feizi T, Fazio F, Chai W, and Wong C-H (2003) Carbohydrate microarrays — a new set of technologies at the frontiers of glycomics, Curr. Opin. Struct. Biol. 13, 637–645. [PubMed: 14568620]
37. Fukui S, Feizi T, Galustian C, Lawson AM, and Chai W (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions, Nat. Biotechnol. 20, 1011–1017. [PubMed: 12219077]
38. Wang D, Liu S, Trummer BJ, Deng C, and Wang A (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells, Nat. Biotechnol. 20, 275–281. [PubMed: 11875429]
39. Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, Razi N, Stevens DJ, Skehel JJ, van Die I, Burton DR, Wilson IA, Cummings R, Bovin N, Wong C-H, and Paulson JC (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins, Proc. Natl. Acad. Sci. U. S. A. 101, 17033. [PubMed: 15563589]
40. Gao C, Wei M, McKitrick TR, McQuillan AM, Heimburg-Molinaro J, and Cummings RD (2019) Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins, Front. Chem. 7.
41. Heimburg-Molinaro J, Song X, Smith DF, and Cummings RD (2011) Preparation and Analysis of Glycan Microarrays, Curr. Protoc. Protein Sci. 64, 12.10.11–12.10.29.
42. Li Z, and Feizi T (2018) The neoglycolipid (NGL) technology-based microarrays and future prospects, FEBS Lett. 592, 3976–3991. [PubMed: 30074246]
43. Rillahan CD, and Paulson JC (2011) Glycan microarrays for decoding the glycome, Annu. Rev. Biochem. 80, 797–823. [PubMed: 21469953]
44. Smith DF, Song X, and Cummings RD (2010) Chapter Nineteen - Use of Glycan Microarrays to Explore Specificity of Glycan-Binding Proteins, In Methods Enzymol (Fukuda M, Ed.), pp 417–444, Academic Press.
45. Geissner A, Reinhardt A, Rademacher C, Johannssen T, Monteiro J, Lepenies B, Thépaut M, Fieschi F, Mrázková J, Wimmerova M, Schuhmacher F, Götze S, Grünstein D, Guo X, Hahm HS, Kandasamy J, Leonori D, Martin CE, Parameswarappa SG, Pasari S, Schlegel MK, Tanaka H, Xiao G, Yang Y, Pereira CL, Anish C, and Seeberger PH (2019) Microbe-focused glycan array screening platform, Proc. Natl. Acad. Sci. USA 116, 1958. [PubMed: 30670663]

46. Byrd-Leotis L, Liu R, Bradley KC, Lasanajak Y, Cummings SF, Song X, Heimburg-Molinaro J, Galloway SE, Culhane MR, Smith DF, Steinhauer DA, and Cummings RD (2014) Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses, Proc. Natl. Acad. Sci. USA 111, E2241. [PubMed: 24843157]
47. Song X, Lasanajak Y, Xia B, Heimburg-Molinaro J, Rhea JM, Ju H, Zhao C, Molinaro RJ, Cummings RD, and Smith DF (2011) Shotgun glycomics: a microarray strategy for functional glycomics, Nat. Methods 8, 85–90. [PubMed: 21131969]
48. Gao C, Hanes MS, Byrd-Leotis LA, Wei M, Jia N, Kardish RJ, McKitrick TR, Steinhauer DA, and Cummings RD (2019) Unique Binding Specificities of Proteins toward Isomeric Asparagine-Linked Glycans, Cell Chem. Biol. 26, 535–547.e534. [PubMed: 30745240]
49. Purohit S, Li T, Guan W, Song X, Song J, Tian Y, Li L, Sharma A, Dun B, Mysona D, Ghamande S, Rungruang B, Cummings RD, Wang PG, and She J-X (2018) Multiplex glycan bead array for high throughput and high content analyses of glycan binding proteins, Nat. Commun. 9, 258. [PubMed: 29343722]
50. Chevolot Y, Laurenceau E, Phaner-Goutorbe M, Monnier V, Souteyrand E, Meyer A, Géhin T, Vasseur J-J, and Morvan F (2014) DNA directed immobilization glycocluster array: applications and perspectives, Curr. Opin. Chem. Biol. 18, 46–54. [PubMed: 24440778]
51. Yan M, Zhu Y, Liu X, Lasanajak Y, Xiong J, Lu J, Lin X, Ashline D, Reinhold V, Smith DF, and Song X (2019) Next-Generation Glycan Microarray Enabled by DNA-Coded Glycan Library and Next-Generation Sequencing Technology, Anal. Chem. 91, 9221–9228. [PubMed: 31187982]
52. Chen Y-H, Narimatsu Y, Clausen TM, Gomes C, Karlsson R, Steentoft C, Spliid CB, Gustavsson T, Salanti A, Persson A, Malmström A, Willén D, Ellervik U, Bennett EP, Mao Y, Clausen H, and Yang Z (2018) The GAGOme: a cell-based library of displayed glycosaminoglycans, Nat. Methods 15, 881–888. [PubMed: 30104636]
53. Briard JG, Jiang H, Moremen KW, Macauley MS, and Wu P (2018) Cell-based glycan arrays for probing glycan–glycan binding protein interactions, Nat. Commun. 9, 880. [PubMed: 29491407]
54. Cummings RD (2009) The repertoire of glycan determinants in the human glycome, Mol. Biosyst. 5, 1087–1104. [PubMed: 19756298]
55. Imperiali B (2019) Bacterial carbohydrate diversity - a Brave New World, Curr. Opin. Chem. Biol. 53, 1–8. [PubMed: 31176085]
56. Hudson KL, Bartlett GJ, Diehl RC, Agirre J, Gallagher T, Kiessling LL, and Woolfson DN (2015) Carbohydrate–Aromatic Interactions in Proteins, J. Am. Chem. Soc. 137, 15152–15160. [PubMed: 26561965]
57. Asensio JL, Ardá A, Cañada FJ, and Jiménez-Barbero J (2013) Carbohydrate–Aromatic Interactions, Acc. Chem. Res. 46, 946–954. [PubMed: 22704792]
58. Drickamer K (1997) Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins, Structure 5, 465–468. [PubMed: 9115445]
59. Taylor ME, and Drickamer K (2014) Convergent and divergent mechanisms of sugar recognition across kingdoms, Curr. Opin. Struct. Biol. 28, 14–22. [PubMed: 25102772]
60. Monsigny M, Mayer R, Roche AC (1991) Sugar-lectin interactions: Sugar cluster, lectin multivalency and avidity, Carbohydr. Lett. 4, 35–52.
61. Lundquist JJ, and Toone EJ (2002) The Cluster Glycoside Effect, Chem. Rev. 102, 555–578. [PubMed: 11841254]
62. Fujimoto Z, Tateno H, and Hirabayashi J (2014) Lectin Structures: Classification Based on the 3-D Structures, In Lectins: Methods and Protocols (Hirabayashi J, Ed.), pp 579–606, Springer New York, New York, NY.
63. Teschke CM, and King J (1992) Folding and assembly of oligomeric proteins in Escherichia coli, Curr. Opin. Biotechnol. 3, 468–473. [PubMed: 1368931]
64. Sørensen HP, and Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli, Microb. Cell Fact. 4, 1. [PubMed: 15629064]
65. Kobayashi Y, Tateno H, Ogawa H, Yamamoto K, and Hirabayashi J (2014) Comprehensive List of Lectins: Origins, Natures, and Carbohydrate Specificities, In Lectins: Methods and Protocols (Hirabayashi J, Ed.), pp 555–577, Springer New York, New York, NY.
66. Woods RJ, and Yang L (2018) Glycan-specific analytical tools.

67. Hu D, Tateno H, and Hirabayashi J (2015) Lectin Engineering, a Molecular Evolutionary Approach to Expanding the Lectin Utilities, Molecules 20.
68. Hirabayashi J, and Arai R (2019) Lectin engineering: the possible and the actual, Interface Focus 9, 20180068. [PubMed: 30842871]
69. Cummings RD, and McEver RP (2015) C-Type Lectins, In Essentials of Glycobiology (rd, Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, and Seeberger PH, Eds.), pp 435–452, Cold Spring Harbor (NY).
70. Drickamer K (1992) Engineering galactose-binding activity into a C-type mannose-binding protein, Nature 360, 183–186. [PubMed: 1279438]
71. Iobst ST, and Drickamer K (1994) Binding of sugar ligands to Ca(2+)-dependent animal lectins. II. Generation of high-affinity galactose binding by site-directed mutagenesis, J. Biol. Chem. 269, 15512–15519. [PubMed: 8195195]
72. Yamamoto K, Konami Y, Osawa T, and Irimura T (1992) Alteration of the Carbohydrate-Binding Specificity of the Bauhinia purpurea Lectin through the Preparation of a Chimeric Lectin, J. Biochem. 111, 87–90. [PubMed: 1607368]
73. Adhikari P, Bachhawat-Sikder K, Thomas CJ, Ravishankar R, Jeyaprakash AA, Sharma V, Vijayan M, and Surolia A (2001) Mutational analysis at Asn-41 in peanut agglutinin. A residue critical for the binding of the tumor-associated Thomsen-Friedenreich antigen, J. Biol. Chem. 276, 40734–40739. [PubMed: 11447220]
74. Ravishankar R, Ravindran M, Suguna K, Surolia A, and Vijayan M (1997) Crystal structure of the peanut lectin – T-antigen complex. Carbohydrate specificity generated by water bridges, Curr. Sci. 72, 855–861.
75. Ban M, Yoon H-J, Demirkan E, Utsumi S, Mikami B, and Yagi F (2005) Structural Basis of a Fungal Galectin from Agrocybe cylindracea for Recognizing Sialoconjugate, J. Mol. Biol. 351, 695–706. [PubMed: 16051274]
76. Imamura K, Takeuchi H, Yabe R, Tateno H, and Hirabayashi J (2011) Engineering of the glycan-binding specificity of Agrocybe cylindracea galectin towards α(2,3)-linked sialic acid by saturation mutagenesis, J. Biochem. 150, 545–552. [PubMed: 21813503]
77. Kasai K (2014) Frontal affinity chromatography: a unique research tool for biospecific interaction that promotes glycobiology, Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci. 90, 215–234.
78. Kuwabara N, Hu D, Tateno H, Makyio H, Hirabayashi J, and Kato R (2013) Conformational change of a unique sequence in a fungal galectin from Agrocybe cylindracea controls glycan ligand-binding specificity, FEBS Lett. 587, 3620–3625. [PubMed: 24036446]
79. Hu D, Tateno H, Sato T, Narimatsu H, and Hirabayashi J (2013) Tailoring GalNAcα1–3Galβ-specific lectins from a multi-specific fungal galectin: dramatic change of carbohydrate specificity by a single amino-acid substitution, Biochem. J. 453, 261–270. [PubMed: 23611418]
80. Hu D, Huang H, Tateno H, Nakakita S. i., Sato T, Narimatsu H, Yao X, and Hirabayashi J (2015) Engineering of a 3′-sulpho-Galβ1–4GlcNAc-specific probe by a single amino acid substitution of a fungal galectin, J. Biochem. 157, 197–200. [PubMed: 25762728]
81. Yamamoto K, Maruyama IN, and Osawa T (2000) Cyborg Lectins: Novel Leguminous Lectins with Unique Specificities1, J. Biochem. 127, 137–142. [PubMed: 10731676]
82. Maenuma K, Yim M, Komatsu K, Hoshino M, Tachiki-Fujioka A, Takahashi K, Hiki Y, Bovin N, and Irimura T (2009) A Library of Mutated Maackia amurensis Hemagglutinin Distinguishes Putative Glycoforms of Immunoglobulin A1 from IgA Nephropathy Patients, J. Proteome Res. 8, 3617–3624. [PubMed: 19368344]
83. Yabe R, Suzuki R, Kuno A, Fujimoto Z, Jigami Y, and Hirabayashi J (2007) Tailoring a Novel Sialic Acid-Binding Lectin from a Ricin-B Chain-like Galactose-Binding Protein by Natural Evolution-Mimicry, J. Biochem. 141, 389–399. [PubMed: 17234683]
84. Hu D, Tateno H, Kuno A, Yabe R, and Hirabayashi J (2012) Directed Evolution of Lectins with Sugar-binding Specificity for 6-Sulfo-galactose, J. Biol. Chem. 287, 20313–20320. [PubMed: 22493425]
85. Hu D, Tateno H, and Hirabayashi J (2014) Directed Evolution of Lectins by an Improved Error-Prone PCR and Ribosome Display Method, In Lectins: Methods and Protocols (Hirabayashi J, Ed.), pp 527–538, Springer New York, New York, NY.

86. Hirabayashi J, Dutta SK, and Kasai K. i. (1998) Novel Galactose-binding Proteins in Annelida: Characterization of 29-kDa Tandem Repeat-type Lectis from the Earthworm Lumbricus terrestris, J. Biol. Chem. 273, 14450–14460. [PubMed: 9603958]
87. Yabe R, Itakura Y, Nakamura-Tsuruta S, Iwaki J, Kuno A, and Hirabayashi J (2009) Engineering a versatile tandem repeat-type α2–6sialic acid-binding lectin, Biochem. Biophys. Res. Commun. 384, 204–209. [PubMed: 19393618]
88. Mahajan S, and Ramya TNC (2018) Nature-inspired engineering of an F-type lectin for increased binding strength, Glycobiology 28, 933–948. [PubMed: 30202877]
89. Hamorsky KT, Kouokam JC, Dent MW, Grooms TN, Husk AS, Hume SD, Rogers KA, Villinger F, Morris MK, Hanson CV, and Matoba N (2019) Engineering of a Lectibody Targeting High-Mannose-Type Glycans of the HIV Envelope, Mol. Ther. 27, 2038–2052. [PubMed: 31471224]
90. Hervé C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, and Knox JP (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects, Proc. Natl. Acad. Sci. USA 107, 15293–15298. [PubMed: 20696902]
91. Din N, Damude HG, Gilkes NR, Miller RC, Warren RA, and Kilburn DG (1994) C1-Cx revisited: intramolecular synergism in a cellulase, Proc. Natl. Acad. Sci. USA 91, 11383–11387. [PubMed: 7972069]
92. Southall SM, Simpson PJ, Gilbert HJ, Williamson G, and Williamson MP (1999) The starch-binding domain from glucoamylase disrupts the structure of starch, FEBS Lett. 447, 58–60. [PubMed: 10218582]
93. Gourlay K, Arantes V, and Saddler JN (2012) Use of substructure-specific carbohydrate binding modules to track changes in cellulose accessibility and surface morphology during the amorphogenesis step of enzymatic hydrolysis, Biotechnology for Biofuels 5, 51. [PubMed: 22828270]
94. Boraston Alisdair B., Bolam David N., Gilbert Harry J., and Davies Gideon J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, Biochem. J. 382, 769–781. [PubMed: 15214846]
95. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, and Henrissat B (2013) The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42, D490–D495. [PubMed: 24270786]
96. Suits MDL, Pluvinage B, Law A, Liu Y, Palma AS, Chai W, Feizi T, and Boraston AB (2014) Conformational analysis of the Streptococcus pneumoniae hyaluronate lyase and characterization of its hyaluronan-specific carbohydrate-binding module, J. Biol. Chem. 289, 27264–27277. [PubMed: 25100731]
97. Armenta S, Moreno-Mendieta S, Sánchez-Cuapio Z, Sánchez S, and Rodríguez-Sanoja R (2017) Advances in molecular engineering of carbohydrate-binding modules, Proteins: Struct., Funct., Bioinf. 85, 1602–1617.
98. Linder M, Lindeberg G, Reinikainen T, Teeri TT, and Pettersson G (1995) The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution, FEBS Lett. 372, 96–98. [PubMed: 7556652]
99. Strobel KL, Pfeiffer KA, Blanch HW, and Clark DS (2016) Engineering Cel7A carbohydrate binding module and linker for reduced lignin inhibition, Biotechnol. Bioeng. 113, 1369–1374. [PubMed: 26616493]
100. Cicortas Gunnarsson L, Nordberg Karlsson E, Albrekt AS, Andersson M, Holst O, and Ohlin M (2004) A carbohydrate binding module as a diversity-carrying scaffold, Protein Eng. Des. Sel. 17, 213–221. [PubMed: 15082834]
101. Cicortas Gunnarsson L, Montanier C, Tunnicliffe RB, Williamson MP, Gilbert HJ, Nordberg Karlsson E, and Ohlin M (2007) Novel xylan-binding properties of an engineered family 4 carbohydrate-binding module, Biochem. J. 406, 209–214. [PubMed: 17506724]
102. von Schantz L, Håkansson M, Logan DT, Walse B, Österlin J, Nordberg-Karlsson E, and Ohlin M (2012) Structural basis for carbohydrate-binding specificity—A comparative assessment of two engineered carbohydrate-binding modules, Glycobiology 22, 948–961. [PubMed: 22434778]

103. Gunnarsson LC, Dexlin L, Karlsson EN, Holst O, and Ohlin M (2006) Evolution of a carbohydrate binding module into a protein-specific binder, Biomol. Eng. 23, 111–117. [PubMed: 16427804]
104. Gunnarsson LC, Zhou Q, Montanier C, Karlsson EN, Brumer H III, and Ohlin M (2006) Engineered xyloglucan specificity in a carbohydrate-binding module, Glycobiology 16, 1171–1180. [PubMed: 16902199]
105. Gullfot F, Tan T-C, von Schantz L, Karlsson EN, Ohlin M, Brumer H, and Divne C (2010) The crystal structure of XG-34, an evolved xyloglucan-specific carbohydrate-binding module, Proteins: Struct., Funct., Bioinf 78, 785–789.
106. Furtado GP, Lourenzoni MR, Fuzo CA, Fonseca-Maldonado R, Guazzaroni M-E, Ribeiro LF, and Ward RJ (2018) Engineering the affinity of a family 11 carbohydrate binding module to improve binding of branched over unbranched polysaccharides, Int. J. Biol. Macromol. 120, 2509–2516. [PubMed: 30195003]
107. Ribeiro DO, Viegas A, Pires VMR, Medeiros-Silva J, Bule P, Chai W, Marcelo F, Fontes CMGA, Cabrita EJ, Palma AS, and Carvalho AL (2020) Molecular basis for the preferential recognition of β1,3–1,4-glucans by the family 11 carbohydrate-binding module from Clostridium thermocellum, FEBS J. 287, 2723–2743. [PubMed: 31794092]
108. von Schantz L, Gullfot F, Scheer S, Filonova L, Cicortas Gunnarsson L, Flint JE, Daniel G, Nordberg-Karlsson E, Brumer H, and Ohlin M (2009) Affinity maturation generates greatly improved xyloglucan-specific carbohydrate binding modules, BMC Biotechnol. 9, 92. [PubMed: 19878581]
109. Alberts B, Johnson A, Lewis J, and et al.. (2002) The Generation of Antibody Diversity, In Mol. Biol. Cell 4th ed., Garland Science, New York, NY.
110. Sterner E, Flanagan N, and Gildersleeve JC (2016) Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database, ACS Chem. Biol. 11, 1773–1783. [PubMed: 27220698]
111. Tommasone S, Allabush F, Tagger YK, Norman J, Köpf M, Tucker JHR, and Mendes PM (2019) The challenges of glycan recognition with natural and artificial receptors, Chem. Soc. Rev. 48, 5488–5505. [PubMed: 31552920]
112. Polonskaya Z, Deng S, Sarkar A, Kain L, Comellas-Aragones M, McKay CS, Kaczanowska K, Holt M, McBride R, Palomo V, Self KM, Taylor S, Irimia A, Mehta SR, Dan JM, Brigger M, Crotty S, Schoenberger SP, Paulson JC, Wilson IA, Savage PB, Finn MG, and Teyton L (2017) T cells control the generation of nanomolar-affinity anti-glycan antibodies, J. Clin. Investig. 127, 1491–1504. [PubMed: 28287405]
113. Haji-Ghassemi O, Blackler RJ, Martin Young N, and Evans SV (2015) Antibody recognition of carbohydrate epitopes†, Glycobiology 25, 920–952. [PubMed: 26033938]
114. Janeway CAJ, Travers P, and Walport M (2001) Immunobiology: The Immune System in Health and Disease, 5th Edition ed., Garland Science, New York.
115. Jeffrey PD, Bajorath J, Chang CY, Yelton D, Hellström I, Hellström KE, and Sheriff S (1995) The X-ray structure of an anti-tumour antibody in complex with antigen, Nat. Struct. Biol. 2, 466–471. [PubMed: 7664109]
116. van Roon A-MM, Pannu NS, de Vrind JPM, van der Marel GA, van Boom JH, Hokke CH, Deelder AM, and Abrahams JP (2004) Structure of an Anti-Lewis X Fab Fragment in Complex with Its Lewis X Antigen, Structure 12, 1227–1236. [PubMed: 15242599]
117. Vazquez AM, Alfonso M, Lanne B, Karlsson KA, Carr A, Barroso O, Fernandez LE, Rengifo E, Lanio ME, Alvarez C, Zeuthen J, and Perez R (1995) Generation of a Murine Monoclonal Antibody Specific for N-Glycolylneuraminic Acid-Containing Gangliosides That Also Recognizes Sulfated Glycolipids, Hybridoma 14, 551–556. [PubMed: 8770642]
118. Conrad GW, Ager-Johnson P, and Woo ML (1982) Antibodies against the predominant glycosaminoglycan of the mammalian cornea, keratan sulfate-I, J. Biol. Chem. 257, 464–471. [PubMed: 6171570]
119. Thomas R, Patenaude SI, MacKenzie CR, To R, Hirama T, Young NM, and Evans SV (2002) Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity, J. Biol. Chem. 277, 2059–2064. [PubMed: 11679577]

120. Mouquet H, Scharf L, Euler Z, Liu Y, Eden C, Scheid JF, Halper-Stromberg A, Gnanapragasam PNP, Spencer DIR, Seaman MS, Schuitemaker H, Feizi T, Nussenzweig MC, and Bjorkman PJ (2012) Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies, Proc. Natl. Acad. Sci. USA 109, E3268. [PubMed: 23115339]

121. McLellan JS, Pancera M, Carrico C, Gorman J, Julien J-P, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang G-Y, Diwanji D, Georgiev I, Do Kwon Y, Lee D, Louder MK, Moquin S, Schmidt SD, Yang Z-Y, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang L-X, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, and Kwong PD (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9, Nature 480, 336–343. [PubMed: 22113616]

122. Calarese DA, Scanlan CN, Zwick MB, Deechongkit S, Mimura Y, Kunert R, Zhu P, Wormald MR, Stanfield RL, Roux KH, Kelly JW, Rudd PM, Dwek RA, Katinger H, Burton DR, and Wilson IA (2003) Antibody Domain Exchange Is an Immunological Solution to Carbohydrate Cluster Recognition, Science 300, 2065. [PubMed: 12829775]

123. Vulliez-Le Normand B, Saul FA, Phalipon A, Bélot F, Guerreiro C, Mulard LA, and Bentley GA (2008) Structures of synthetic O-antigen fragments from serotype 2a Shigella flexneri in complex with a protective monoclonal antibody, Proc. Natl. Acad. Sci. U. S. A. 105, 9976–9981. [PubMed: 18621718]

124. Brooks CL, Müller-Loennies S, Borisova SN, Brade L, Kosma P, Hirama T, Mackenzie CR, Brade H, and Evans SV (2010) Antibodies raised against chlamydial lipopolysaccharide antigens reveal convergence in germline gene usage and differential epitope recognition, Biochemistry 49, 570–581. [PubMed: 20000757]

125. Murase T, Zheng RB, Joe M, Bai Y, Marcus SL, Lowary TL, and Ng KK (2009) Structural insights into antibody recognition of mycobacterial polysaccharides, J. Mol. Biol. 392, 381–392. [PubMed: 19577573]

126. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang PS, Wang SK, Stanfield RL, Julien JP, Ramos A, Crispin M, Depetris R, Katpally U, Marozsan A, Cupo A, Maloveste S, Liu Y, McBride R, Ito Y, Sanders RW, Ogohara C, Paulson JC, Feizi T, Scanlan CN, Wong CH, Moore JP, Olson WC, Ward AB, Poignard P, Schief WR, Burton DR, and Wilson IA (2011) A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield, Science 334, 1097–1103. [PubMed: 21998254]

127. Monnier PP, Vigouroux JR, and Tassew GN (2013) In Vivo Applications of Single Chain Fv (Variable Domain) (scFv) Fragments, Antibodies 2.

128. Markiv A, Anani B, Durvasula RV, and Kang AS (2011) Module based antibody engineering: A novel synthetic REDantibody, J. Immunol. Methods 364, 40–49. [PubMed: 21055406]

129. Birnboim-Perach R, Grinberg Y, Vaks L, Nahary L, and Benhar I (2019) Production of Stabilized Antibody Fragments in the E. coli Bacterial Cytoplasm and in Transiently Transfected Mammalian Cells, Methods Mol. Biol. 1904, 455–480. [PubMed: 30539486]

130. Bujak E, Matasci M, Neri D, and Wulhfard S (2014) Reformatting of scFv Antibodies into the scFv-Fc Format and Their Downstream Purification, In Monoclonal Antibodies: Methods and Protocols (Ossipow V, and Fischer N, Eds.), pp 315–334, Humana Press, Totowa, NJ.

131. Müller-Loennies S, Galliciotti G, Kollmann K, Glatzel M, and Braulke T (2010) A novel single-chain antibody fragment for detection of mannose 6-phosphate-containing proteins: application in mucolipidosis type II patients and mice, Am. J. Pathol. 177, 240–247. [PubMed: 20472886]

132. Blackler RJ, Evans DW, Smith DF, Cummings RD, Brooks CL, Braulke T, Liu X, Evans SV, and Müller-Loennies S (2016) Single-chain antibody-fragment M6P-1 possesses a mannose 6-phosphate monosaccharide-specific binding pocket that distinguishes N-glycan phosphorylation in a branch-specific manner†, Glycobiology 26, 181–192. [PubMed: 26503547]

133. Kubota T, Matsushita T, Niwa R, Kumagai I, and Nakamura K (2010) Novel Anti-Tn Single-Chain Fv-Fc Fusion Proteins Derived from Immunized Phage Library and Antibody Fc Domain, Anticancer Res. 30, 3397–3405. [PubMed: 20944114]

134. Smits NC, Lensen JFM, Wijnhoven TJM, ten Dam GB, Jenniskens GJ, and van Kuppevelt TH (2006) Phage Display-Derived Human Antibodies Against Specific Glycosaminoglycan Epitopes, In Methods Enzymol, pp 61–87, Academic Press.
135. van Kuppevelt TH, Dennissen MA, van Venrooij WJ, Hoet RM, and Veerkamp JH (1998) Generation and application of type-specific anti-heparan sulfate antibodies using phage display technology. Further evidence for heparan sulfate heterogeneity in the kidney, J. Biol. Chem. 273, 12960–12966. [PubMed: 9582329]
136. Bona C (1993) Molecular characteristics of anti-polysaccharide antibodies, Springer Semin. Immunopathol. 15, 103–118. [PubMed: 8256193]
137. Brorson K, Thompson C, Wei G, Krasnokutsky M, and Stein KE (1999) Mutational Analysis of Avidity and Fine Specificity of Anti-Levan Antibodies, J. Immunol. 163, 6694. [PubMed: 10586066]
138. Yelton DE, Rosok MJ, Cruz G, Cosand WL, Bajorath J, Hellström I, Hellström KE, Huse WD, and Glaser SM (1995) Affinity maturation of the BR96 anti-carcinoma antibody by codon-based mutagenesis, J. Immunol. 155, 1994. [PubMed: 7636250]
139. Deng SJ, MacKenzie CR, Hirama T, Brousseau R, Lowary TL, Young NM, Bundle DR, and Narang SA (1995) Basis for selection of improved carbohydrate-binding single-chain antibodies from synthetic gene libraries, Proc. Natl. Acad. Sci. U. S. A. 92, 4992–4996. [PubMed: 7539139]
140. Wu Y, West AP Jr., Kim HJ, Thornton ME, Ward AB, and Bjorkman PJ (2013) Structural basis for enhanced HIV-1 neutralization by a dimeric immunoglobulin G form of the glycan-recognizing antibody 2G12, Cell Rep. 5, 1443–1455. [PubMed: 24316082]
141. Lin T-Y, and Lai JR (2017) Interrogation of side chain biases for oligomannose recognition by antibody 2G12 via structure-guided phage display libraries, Biorg. Med. Chem. 25, 5790–5798.
142. Altman MO, Bennink JR, Yewdell JW, and Herrin BR (2015) Lamprey VLRB response to influenza virus supports universal rules of immunogenicity and antigenicity, eLife 4, e07467.
143. Tasumi S, Velikovsky CA, Xu G, Gai SA, Wittrup KD, Flajnik MF, Mariuzza RA, and Pancer Z (2009) High-affinity lamprey VLRA and VLRB monoclonal antibodies, Proc. Natl. Acad. Sci. USA 106, 12891. [PubMed: 19625627]
144. Alder MN, Herrin BR, Sadlonova A, Stockard CR, Grizzle WE, Gartland LA, Gartland GL, Boydston JA, Turnbough CL, and Cooper MD (2008) Antibody responses of variable lymphocyte receptors in the lamprey, Nat. Immunol. 9, 319–327. [PubMed: 18246071]
145. Collins BC, Gunn RJ, McKitrick TR, Cummings RD, Cooper MD, Herrin BR, and Wilson IA (2017) Structural Insights into VLR Fine Specificity for Blood Group Carbohydrates, Structure (London, England : 1993) 25, 1667–1678.e1664.
146. McKitrick TR, Eris D, Mondal N, Aryal RP, McCurley N, Heimburg-Molinaro J, and Cummings RD (2020) Antibodies from Lampreys as Smart Anti-Glycan Reagents (SAGRs): Perspectives on Their Specificity, Structure, and Glyco-genomics, Biochemistry.
147. Han BW, Herrin BR, Cooper MD, and Wilson IA (2008) Antigen Recognition by Variable Lymphocyte Receptors, Science 321, 1834. [PubMed: 18818359]
148. Hong X, Ma MZ, Gildersleeve JC, Chowdhury S, Barchi JJ, Mariuzza RA, Murphy MB, Mao L, and Pancer Z (2013) Sugar-Binding Proteins from Fish: Selection of High Affinity “Lambodies” That Recognize Biomedically Relevant Glycans, ACS Chem. Biol. 8, 152–160. [PubMed: 23030719]
149. Luo M, Velikovsky CA, Yang X, Siddiqui MA, Hong X, Barchi JJ Jr., Gildersleeve JC, Pancer Z, and Mariuzza RA (2013) Recognition of the Thomsen-Friedenreich pancarcinoma carbohydrate antigen by a lamprey variable lymphocyte receptor, J. Biol. Chem. 288, 23597–23606. [PubMed: 23782692]
150. McKitrick TR, Goth CK, Rosenberg CS, Nakahara H, Heimburg-Molinaro J, McQuillan AM, Falco R, Rivers NJ, Herrin BR, Cooper MD, and Cummings RD (2020) Development of smart anti-glycan reagents using immunized lampreys, Commun. Biol 3, 91. [PubMed: 32111965]
151. Wilton EE, Opyr MP, Kailasam S, Kothe RF, and Wieden H-J (2018) sdAb-DB: The Single Domain Antibody Database, ACS Synth. Biol. 7, 2480–2484. [PubMed: 30441908]

152. English H, Hong J, and Ho M (2020) Ancient species offers contemporary therapeutics: an update on shark VNAR single domain antibody sequences, phage libraries and potential clinical applications, Antibody Therapeutics 3, 1–9. [PubMed: 32118195]

153. McMahon C, Baier AS, Pascolutti R, Wegrecki M, Zheng S, Ong JX, Erlandson SC, Hilger D, Rasmussen SGF, Ring AM, Manglik A, and Kruse AC (2018) Yeast surface display platform for rapid discovery of conformationally selective nanobodies, Nat. Struct. Mol. Biol. 25, 289–296. [PubMed: 29434346]

154. Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grütter MG, and Plückthun A (2004) High-affinity binders selected from designed ankyrin repeat protein libraries, Nat. Biotechnol. 22, 575–582. [PubMed: 15097997]

155. Colombo C, Pitirollo O, and Lay L (2018) Recent Advances in the Synthesis of Glycoconjugates for Vaccine Development, Molecules (Basel, Switzerland) 23, 1712.

156. Mettu R, Chen C-Y, and Wu C-Y (2020) Synthetic carbohydrate-based vaccines: challenges and opportunities, J. Biomed. Sci. 27, 9. [PubMed: 31900143]

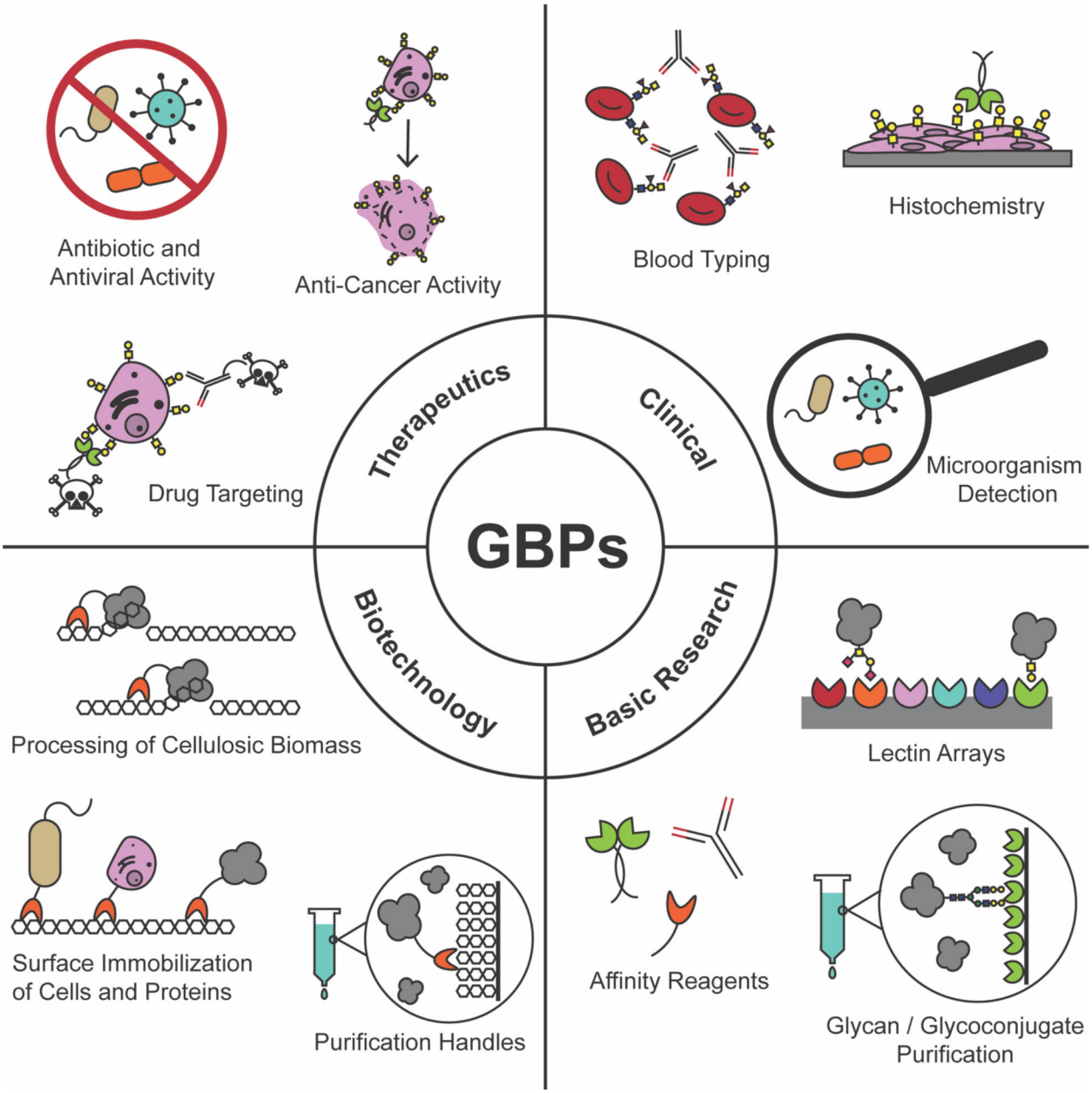


Figure 1.

Applications of glycan-binding proteins (GBPs). GBPs find many uses in therapeutic, clinical, biotechnological and basic research applications. Some of the prominent methods in which they have been utilized are depicted for each application area.

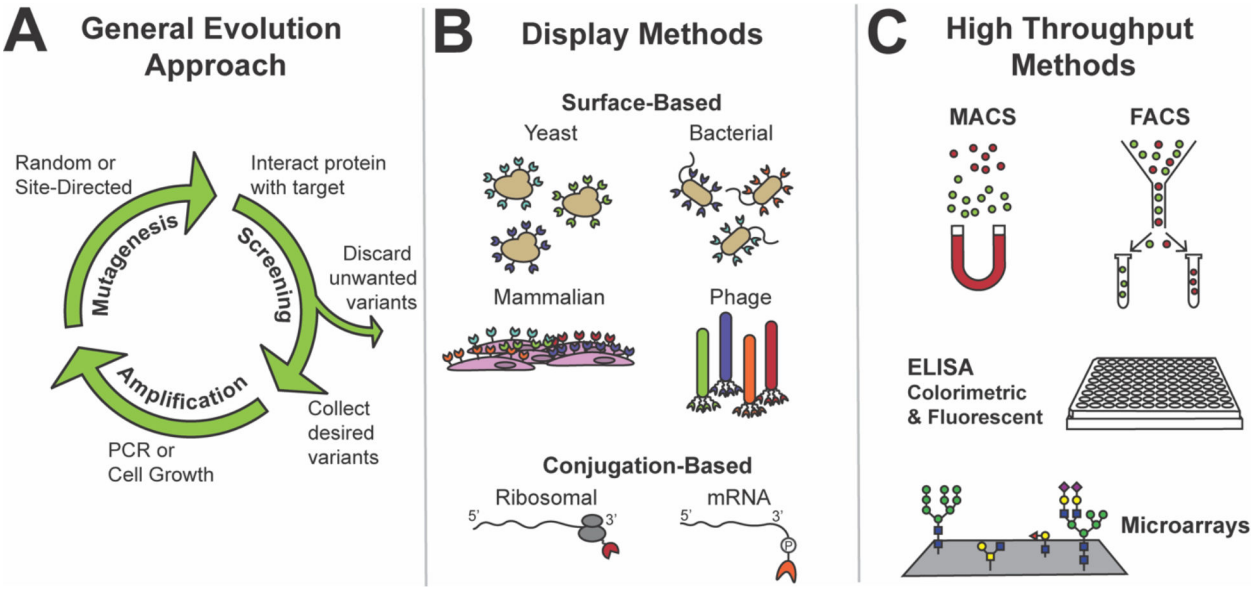


Figure 2.

Evolutionary and screening methods. (A) The directed-evolution process includes an iteration of three main stages: mutagenesis, screening, and amplification. (B) Many display methods have been developed for linking phenotype to genotype in directed evolution experiments. (C) High throughput screening enhances the feasibility of directed evolution by efficient separation or detection of binders to intended targets.

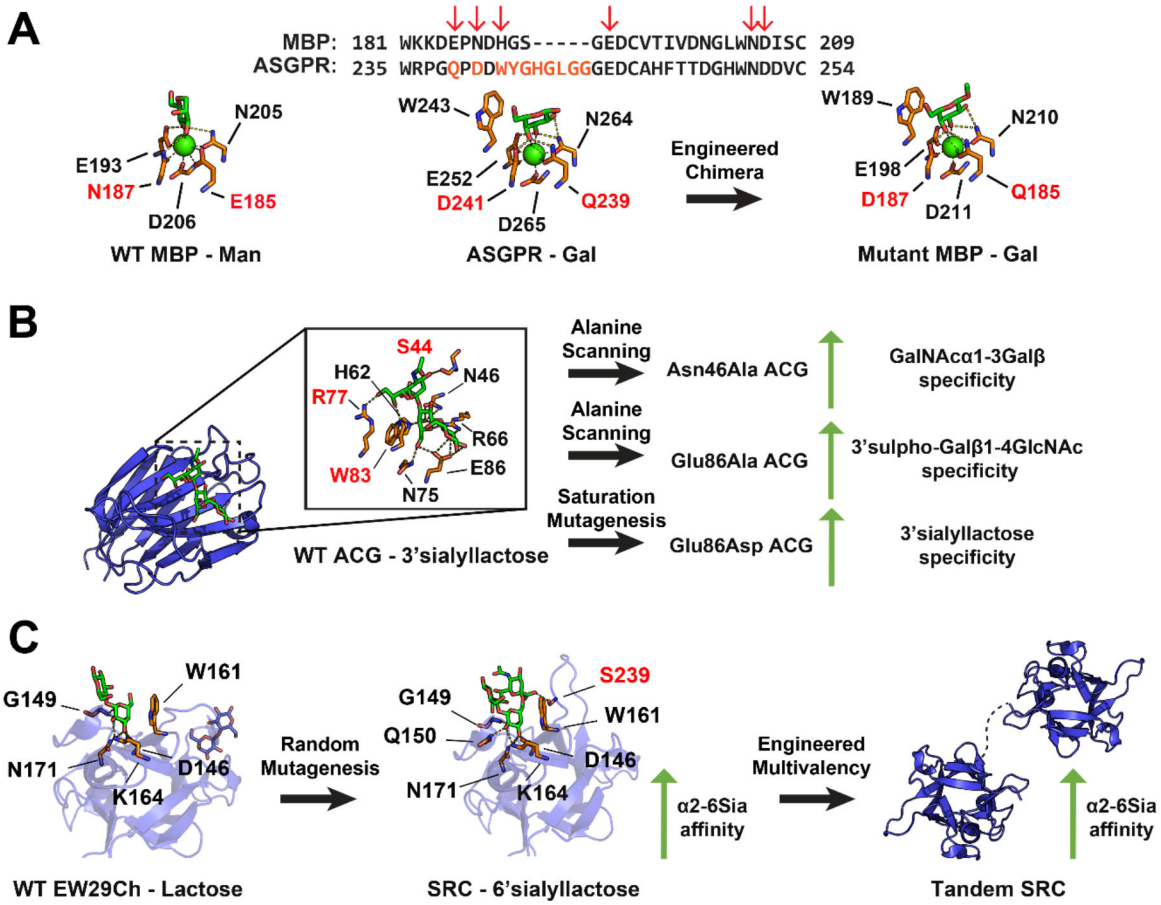


Figure 3.

Lectin engineering. (A) Sequence alignment of rat mannose binding protein (MBP) and rat asialoglycoprotein receptor (ASGPR) binding sites. Orange residues indicate mutations made in mutant MBP, and red arrows indicate residues shown in structures. Wild-type (WT) MBP structure with bound mannose (green) and interacting residues (orange) (PDB 1KX1). Red indicates Gln-Pro-Asp motif residues. Human ASGPR with bound galactose (green) and interacting residues (orange) (PDB 5JPV). Red indicates Glu-Pro-Asn motif residues. Chimeric MBP structure with bound Gal showing interacting residues in orange (PDB 1AFA). (B) WT *A. cylindracea* galectin (ACG) with bound 3'sialyllactose (green) showing interacting residues (orange) (PDB 1WW4). Residues in red show interactions with sialic acid. Alanine scanning found Asn46Ala mutant and Glu86Ala mutant with GalNAcα1–3Galβ specificity and 3'sulpho-Galβ1–4GlcNAc specificity. Saturation mutagenesis found Glu86Asp with 3'sialyllactose specificity. (C) Left: WT Earth worm galectin C-terminal domain (EW29Ch) with bound lactose (green) and interacting residues (orange) (PDB 2ZQN). Center: Mutant EW29Ch (SRC) with bound 6'sialyllactose (green), with interacting residues (orange), mutant residue Ser239 shown in red (PDB 2DS0). Right: Multivalent tandem SRC with increased affinity.

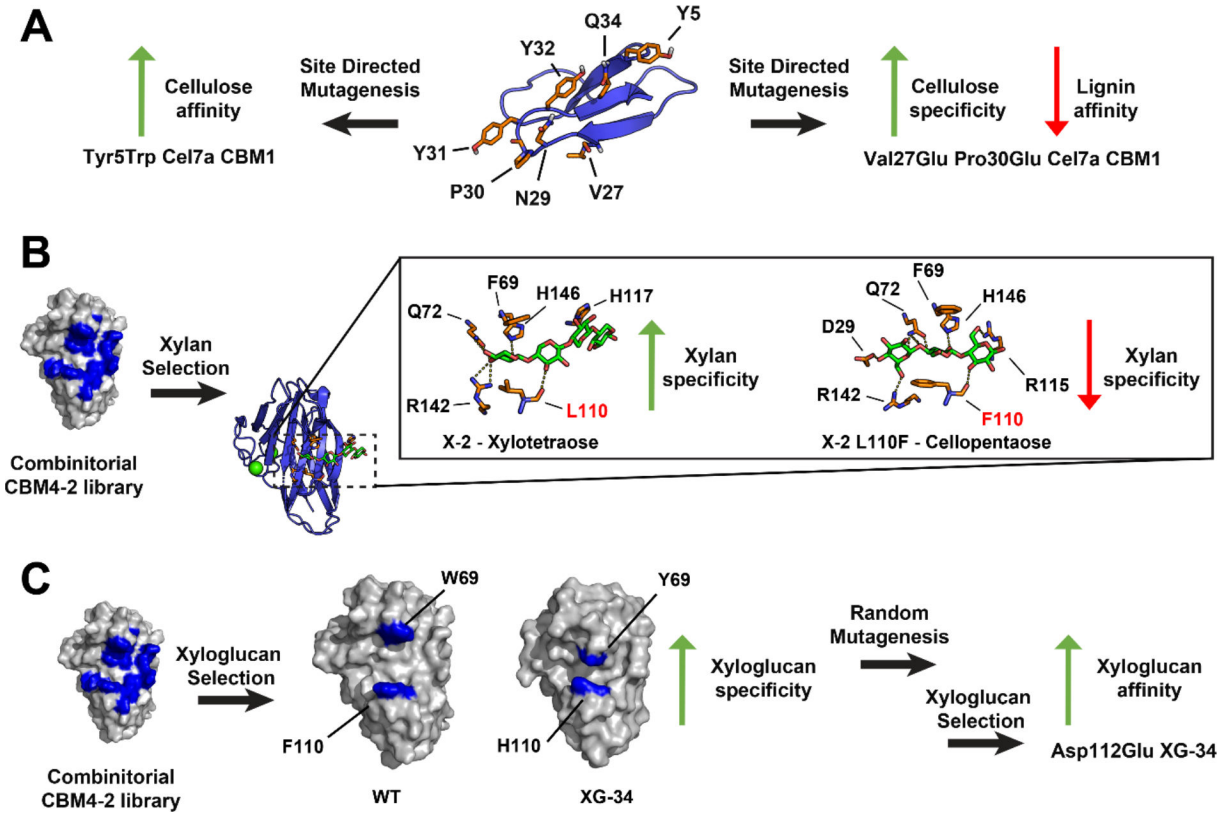


Figure 4.

CBM engineering. (A) Cel7a CBM1 engineering. Site directed mutagenesis used to increase affinity for cellulose, and to decrease affinity for lignin. Structure shows residues involved in cellulose binding, and residues involved in lignin binding (PDB 1CBH). (B) Wild-type CBM4–2 showing the locations of residues mutated to create a combinatorial phage display library (PDB 1K45). Xylan selections isolated clone X-2, shown with bound xylotetraose (green) with interacting residues (orange) (PDB 2Y6K). Green spheres are bound Ca^{2+}. Box: Zoom in of interacting residues of clone X-2 with bound xylotetraose, with key residue L110 shown in red. Clone X-2 with F110L mutation shows decreased xylan specificity. Interacting residues (orange) shown with bound cellopentaose (green), key residue F110 shown in red (PDB 2Y6G). (C) Xyloglucan selection of combinatorial CBM4–2 library produces xyloglucan binding clone XG-34. Structure of wild-type CBM4–2 structure showing important residues W69 and F110 in blue (PDB 1K45) compared to clone XG-34 structure with decreased binding cleft width. Mutations Y69 and H110 shown in blue. After random mutagenesis and xyloglucan selection, E112D XG-34 has increased affinity for xyloglucan.

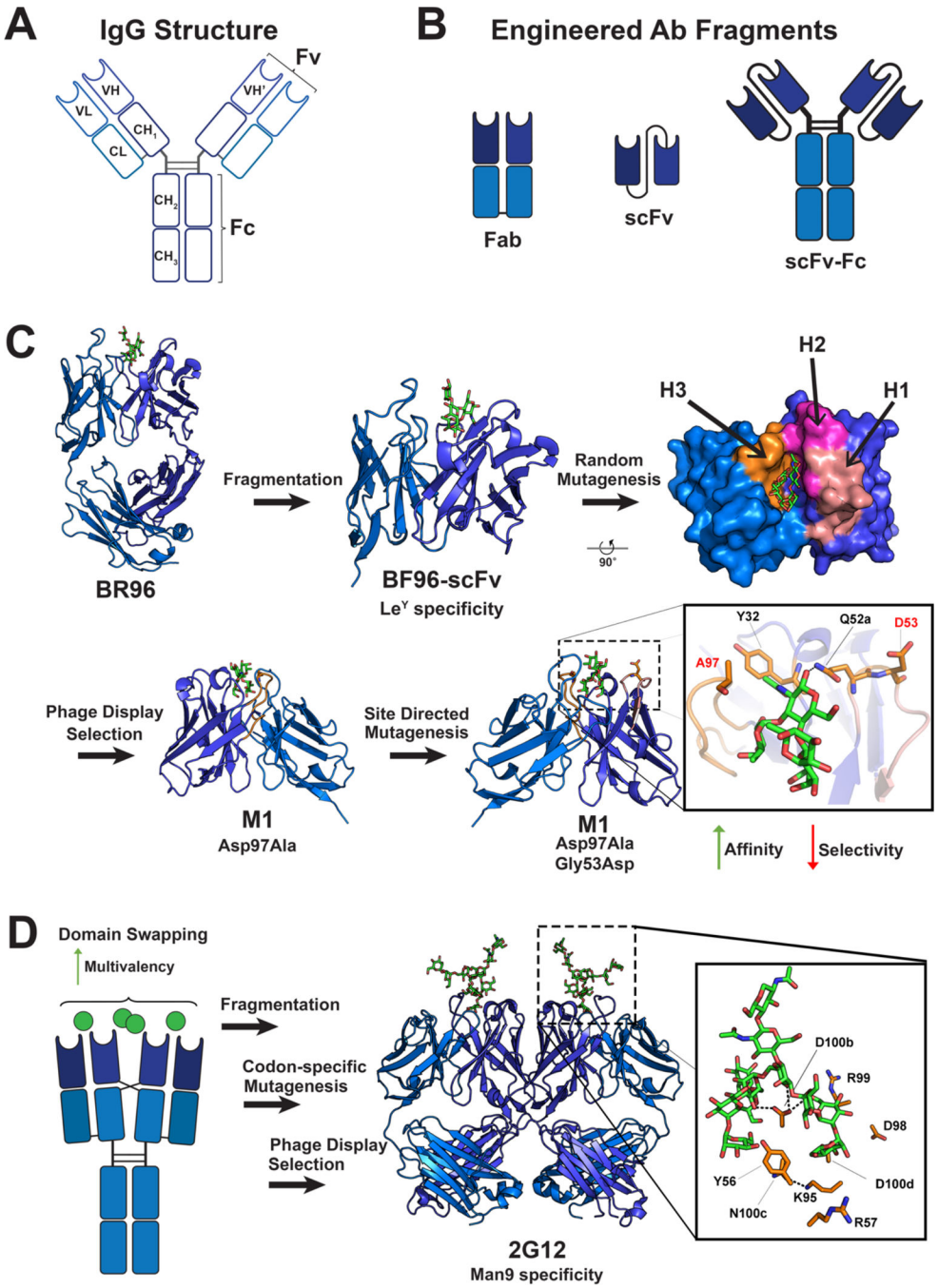


Figure 5.

Antibody (Ab) structure and engineering. (A) Schematic of immunoglobulin G Ab denoting the heavy (H) and light (L) chains, crystallizable fragment (Fc), and variable fragment (Fv). (B) Engineered Ab fragments include the antigen binding fragment (Fab), single-chain variable fragment (scFv), and chimeric scFv with crystallizable fragment (scFv-Fc). (C) Engineering strategy for anti-glycan Ab development. A sequence of fragmentation, random mutagenesis and phage display, and site directed mutagenesis on the anti-Le^Y Ab, BR96 (PDB 1CLY). Regions subjected to mutagenesis are colored orange, magenta and salmon

in the surface representation. Mutated residues denoted in red in the inset. Ultimately, a gain of affinity for the Le^Y substrate was seen at the consequence of loss of specificity. (D) Domain swapping of IgG variable heavy chains (VH and VH') brings binding sites closer together for multivalent glycan binding to the Man-containing glycan epitope (green circles). Detailed structure of the 2G12 Fab in complex with Man9 epitope of the gp120 target (PDB 6N2X). The inset shows the residues mutated during random mutagenesis which contact the Man9 glycan within van der Waals distance. In all 3D structures, the protein backbone is colored tv blue and marine for heavy and light chains respectively. Green sticks represent the glycan ligands and orange sticks represent main chain residues.

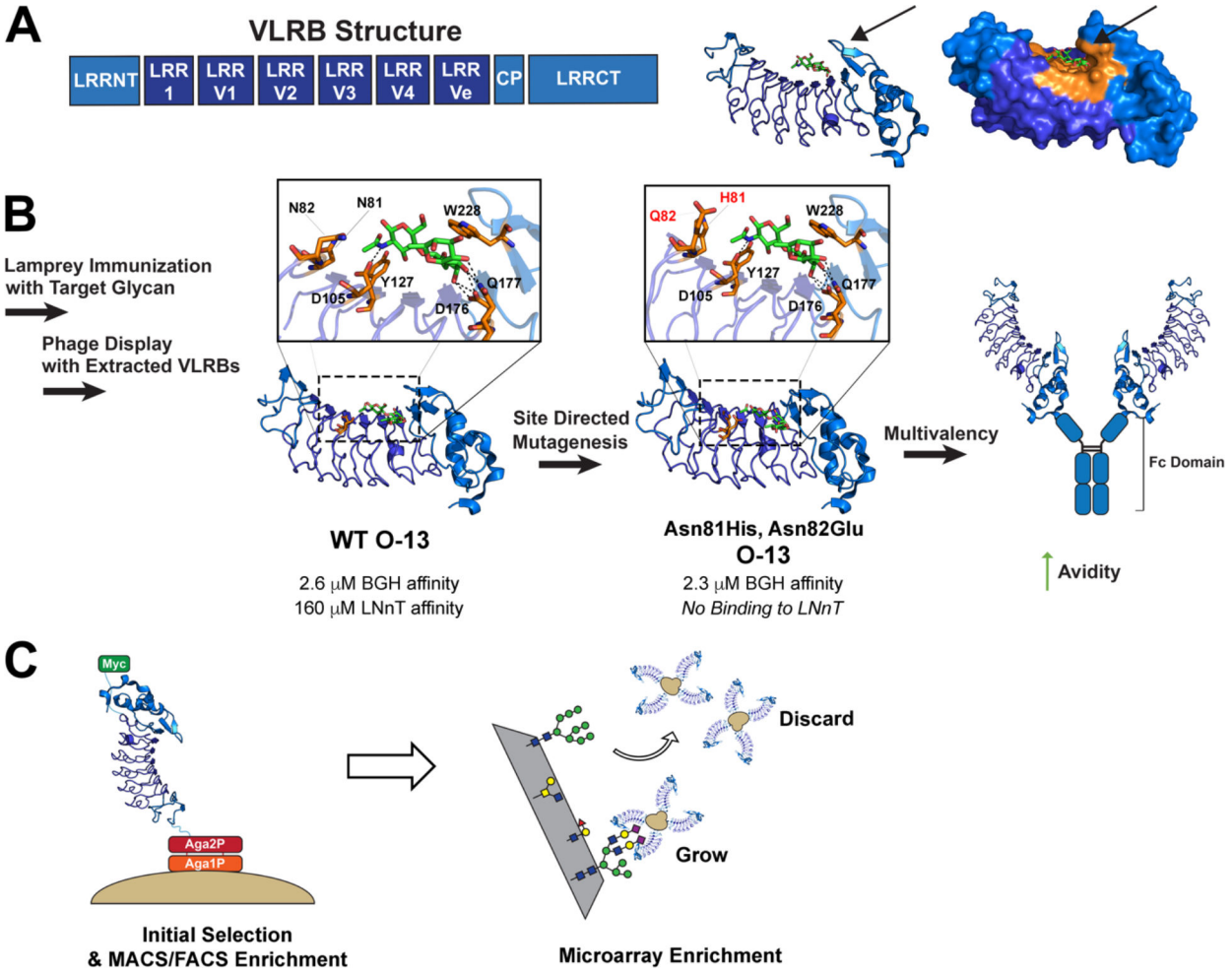


Figure 6.

VLRB structure and engineering. (A) VLRBs are composed of leucine rich repeat (LRR) regions which fold to form a concave variable binding surface. A schematic of the protein sequence, highlighting the LRR motif, and a detailed structure of a monomeric protein in complex with glycan target are shown. Black arrows denote the conserved tryptophan in the LRRCT domain, Trp228, which acts as a molecular “thumb” in anti-glycan recognition. (B) Engineering an anti-blood group H (BGH) trisaccharide VLRB. The VLRB O-13 obtained after lamprey immunization and phage display exhibits affinity to BGH and lacto-n-neotetraose (LNnT). WT O-13 in complex with BGH trisaccharide shows residues involved in binding; mutated residues, denoted in red, maintain BGH binding while preventing LNnT binding. Conjugation of VLRBs to antibody Fc domain generates a dimeric construct utilized in YSD, affinity measurements and specificity determination. (D) An initial YSD selection with VLRBs followed by glycan microarray enrichment can provide VLRBs to any glycan present on the array. For all 3D structures, the protein backbone is tv-blue and marine, contacting residues are colored orange, and glycan substrate is green. PDB 5UF1.